DOI 10.15826/vestnik.2025.24.4.047 Original Paper

УДК 336.76

Выявление социального разгона акций с использованием машинного обучения: эмпирический анализ российского рынка

Г. А. Хазиев ✉

Национальный исследовательский университет «Высшая школа экономики», г. Москва, Россия

✉ *glebhaziev@mail.ru*

Аннотация. В последние годы российский фондовый рынок сталкивается с ростом влияния социальных сетей на поведение частных инвесторов, что создает предпосылки для возникновения краткосрочных ценовых и объемных аномалий. Влияние публикаций в социальных сетях на динамику акций отдельных компаний активно растет с 2019 г., однако до сих пор отсутствуют теоретические и практические работы по его комплексному изучению, что определяет актуальность настоящего исследования. Целью исследования является выявление и количественная оценка социального разгона акций — ситуаций, при которых публикации в социальных сетях предшествуют аномальному росту цен и объемов торгов. В рамках исследования выдвигается гипотеза о существовании причинной связи между социальной активностью и рыночными отклонениями, а также проверяется гипотеза о более высокой эффективности моделей машинного обучения в выявлении социального разгона акций по сравнению с традиционными эконометрическими моделями. Для анализа использовались данные по 104 российским компаниям за период с 2019 по 2025 г., включая рыночные признаки, агрегированные индексы обсуждений и сентимента инвесторов. Методология включала построение модели нормального поведения на основе исторических данных и расчет отклонений в дни с признаками социального разгона. Для классификации использовались четыре модели машинного обучения (CatBoost, SVM, Random Forest, KNN) и модель логистической регрессии. Результаты подтвердили наличие устойчивых аномалий в поведении цен и объемов в дни социального разгона, а также статистически значимое превышение социальных индикаторов. Модели машинного обучения достигли показателя ROC-AUC > 0,9, что превосходит результаты многих ранее опубликованных работ. Полученные выводы имеют теоретическое значение для поведенческих финансов и практическую ценность для регуляторов, трейдеров и аналитиков, которые могут использовать результаты для выявления аномалий в торговой активности и учета фактора социального влияния в торговых стратегиях.

Ключевые слова: фондовый рынок; акции; манипуляции; социальный разгон; машинное обучение; сентимент инвесторов.

1. Введение

За последние несколько лет структура участников фондового рынка существенно изменилась. С 2020 г. начался активный рост доли частных инвесторов, который к 2025 г. привел к их доминирующему положению на рынке: количество физических лиц с брокерскими счетами на Московской бирже

превысило 36,7 млн чел., а доля частных инвесторов в общем обороте торгов за 2021–2025 гг. увеличилась с 40 до 75 % (MOEX[1]).

С изменением структуры торгов трансформируются также формы воздействия на поведение инвесторов. Социальные сети становятся инструментом для формирования мнения частных инвесторов о публичных компаниях и способствуют совершению ими операций по покупке и продаже акций конкретных компаний. Согласно исследованию инвестиционной грамотности ЦБ РФ[2], 37 % частных инвесторов в России используют «Телеграм» и другие социальные платформы как основной источник информации для принятия инвестиционных решений.

Изменение структуры участников фондового рынка и появление новых каналов распространения информации среди инвесторов уже сейчас оказывает существенное влияние на динамику торгов отдельными акциями. В 2021 г. Центральный банк Российской Федерации (ЦБ РФ) выявил случай манипулирования акциями компании «Россети Юг» с использованием схемы «pump and dump» (схемы манипулятивного повышения курса на рынках ценных бумаг, криптовалют или иных подобных активов с последующим их сильным обвалом, в основу которой заложена попытка увеличить стоимость при помощи ложных, ничем не обоснованных рекомендаций), когда после призывов к покупке акций в «Телеграм»-каналах объемы торгов выросли в десятки раз, а цена акции выросла на 5 %. Несмотря на публичность таких случаев, в официальном перечне выявленных случаев манипулирования рынком ЦБ РФ содержатся только два зафиксированных кейса манипуляций с использованием социальных сетей, в то время как активность подобных каналов остается стабильно высокой.

В рамках исследования рассматривается феномен «социального разгона акций», при котором резкое увеличение интенсивности обсуждений и позитивного сентимента вокруг акций конкретной компании в социальных сетях сопровождается краткосрочными аномалиями изменения цены и объема торгов. Под социальным разгоном в работе понимаются не противоправные действия, хотя фактически они и могут быть таковыми (это невозможно проверить без доступа к информации об операциях на фондовом рынке со стороны конкретных лиц), а любые существенные аномальные изменения цен и объемов торгов, которые произошли в результате публикации сообщений в социальных сетях.

Актуальность исследования для рынка России подчеркивает тот факт, что работы по влиянию социальных сетей на аномальное изменение торговой

[1] Частные инвесторы в апреле вложили в акции и облигации на Московской бирже 181,5 млрд рублей / Новости и пресс-релизы // Московская биржа : [офиц. сайт]. URL: https://www.moex.com/n90170 (дата обращения: 06.06.2025).

[2] Исследование инвестиционной грамотности // Центральный банк Российской Федерации : [офиц. сайт]. URL: https://www.cbr.ru/analytics/szpp/fin_literacy/issledovanie-investicionnoy-gramotnosti (дата обращения: 06.06.2025).

активности по отдельным акциям практически отсутствуют. Так, по теме влияния социальных сетей и сентимента инвесторов на доходность российских акций и портфели различных активов исследования ведутся [1, 2], также имеются единичные публикации экономической тематики по выявлению и минимизации манипуляций [3, 4], но по теме выявления социального разгона акций на российском рынке работы полностью отсутствуют, что позволяет нам заполнить существующий пробел в исследованиях.

С развитием методов машинного обучения (Machine learning — ML) и искусственного интеллекта стало появляться все больше исследований [5–8], в которых данные методы используются для выявления социального разгона, а также манипуляций акциями. Это вызвано ростом количества доступных данных, требующих анализа, более надежными результатами ML-моделей по сравнению с традиционными методами эконометрического анализа данных [9–11] и их возможностью самостоятельно определять нелинейные, сложные паттерны манипуляций. Также во многих современных исследованиях, основанных на применении машинного обучения и искусственного интеллекта для выявления манипуляций, в качестве одной из значимых переменных стали включать сентимент частных инвесторов, который определяется путем анализа публикаций и обсуждений конкретных акций в социальных сетях [12–15].

Таким образом, в рамках текущий работы будет не только закрыт существующий исследовательский пробел и рассмотрена проблематика социального разгона акций на российском рынке, но и применены передовые методы (использования ML-моделей и учета фактора сентимента) для выявления фактов социального разгона российских акций.

В работе рассматриваются два *исследовательских вопроса*:

1. Могут ли публикации в социальных сетях выступать причиной аномальных рыночных колебаний цен и объемов торгов акциями на российском фондовом рынке?
2. Способны ли модели машинного обучения эффективно выявлять случаи социального разгона на российском фондовом рынке и делать это лучше традиционных эконометрических моделей?

Целью исследования является выявление и количественная оценка феномена социального разгона акций. Под социальным разгоном акций понимаются ситуации, при которых публикации в социальных сетях предшествуют аномальному росту цен и объемов торгов.

Гипотезы исследования:

H1. Публикации в социальных сетях могут выступать причиной аномального роста цен и объемов торгов акциями российских компаний, приводя к эффекту социального разгона.

В рамках данной гипотезы проверяется причинная связь между всплесками сетевой активности (по времени, интенсивности обсуждений, сентименту инвесторов) и резкими отклонениями рыночных данных торгов от нормы.

Аномалии не должны совпадать с фундаментальными корпоративными и рыночными событиями.

H2. Модели машинного обучения позволяют более точно выявлять случаи социального разгона акций по сравнению с традиционными эконометрическими моделями.

Для проверки гипотезы сравниваются метрики качества ML-моделей (CatBoost, Random Forest, SVM, KNN) и эконометрической модели логистической регрессии, построенной на тех же признаках.

Структура статьи. В разделе 2 представлен обзор предыдущих исследований, в разделе 3 описаны данные и методология. В разделе 4 сначала представлены результаты, полученные для выявления влияния социального разгона на динамику торгов акций, а затем для моделей выявления социального разгона на российском рынке. Раздел 5 посвящен обсуждению полученных результатов. Заключительный раздел содержит ключевые выводы исследования и описание возможностей их практического применения.

2. Обзор литературы

2.1. Социальные сети и сентимент инвесторов как источники рыночной информации

Традиционно в рамках исследований XX в. динамика фондового рынка объяснялась гипотезой эффективного рынка (effective market hypothesis, EMH) и теорией случайных блужданий, описанных Fama [16] и Samuelson [17]. В рамках этой парадигмы поведение инвесторов рассматривалось как рациональное, а нерыночные факторы, включая информационные и новостные потоки, массовую психологию, не признавались значимыми источниками изменения цен на рыночные активы.

Однако последующие эмпирические наблюдения поставили под сомнение универсальность EMH. Shiller [18] показал, что волатильность цен акций на рынке может быть существенно выше, чем волатильность будущих дивидендов, что противоречит гипотезе эффективного рынка. Это стало одним из первых эмпирических подтверждений иррациональности поведения инвесторов.

В ответ на эти несоответствия в XXI веке получила развитие теория поведенческих финансов, утверждающая, что иррациональные действия инвесторов могут приводить к устойчивым отклонениям цен от фундаментальных значений. В частности, Baker & Wurgler [19, 20] одними из первых доказали, что рыночный сентимент — совокупность эмоций и ожиданий частных инвесторов — может оказывать систематическое влияние на волатильность, недооценку или переоценку отдельных акций.

Появление и рост социальных сетей, таких как Twitter, Reddit, «Телеграм» и платформ для инвесторов (например, «Пульс» в России), усилило роль сентимента в рыночных колебаниях. Публикации в соцсетях стали новой формой трансляции мнений и формирования настроений инвесторов, способной

вызывать краткосрочные ценовые импульсы даже при отсутствии фундаментальных причин [21, 22]. Развитие новых каналов информации создало предпосылки для изучения их влияния на рыночные котировки.

Cookson et al. [23] изучили влияние настроений и внимания (частоты обсуждения) компаний на социальных платформах Twitter, StockTwits и Seeking Alpha и пришли к выводу, что сентимент инвесторов лучше предсказывает рост акций, тогда как повышенное внимание — снижение. Также было обнаружено, что влияние на котировки и предсказательная способность зависят от платформы публикации сообщений, что подчеркивает важность анализа не только информационного фона, но и конкретных каналов его распространения.

Chen et al. [15] проанализировали публикации в популярной платформе для инвесторов Seeking & Alpha и показали, что коллективные мнения инвесторов могут опережать движения рынка и использоваться в предсказательных моделях изменения цен акций.

Gupta et al. [14] исследовали, как колебания уровня доверия инвесторов (investor confidence) влияют на предсказуемость рыночной волатильности. Авторы показали, что высокий уровень доверия инвесторов, определяемый на основе анализа текста новостей и социальных медиа, сопровождается более высокой точностью прогнозов рыночных колебаний. Это подчеркивает важность учета поведенческих факторов при построении моделей оценки рыночной доходности и волатильности акций.

Cai et al. [13] исследуют влияние настроений инвесторов в режиме реального времени на высокочастотную доходность акций фондового рынка Китая и приходят к выводу, что публикации частных инвесторов в социальных сетях могут существенно влиять на внутридневную доходность акций. Авторы приходят к выводу, что наибольшее влияние сентимента проявляется в периоды низкой волатильности акций, когда отсутствуют другие рыночные сигналы на покупку или продажу, с ростом волатильности влияние сентимента становится менее заметным.

Теплова и др. [1] и Teplova et al. [2] изучали влияние сентимента частных инвесторов на рыночные характеристики акций российского фондового рынка. На основе данных из социальной сети «Телеграм» и платформы mfd.ru построен авторский индекс сентимента, показавший значимую взаимосвязь с доходностью, волатильностью и ликвидностью акций. Наиболее выраженное влияние на биржевые характеристики было обнаружено при росте количества негативных сообщений. Также доказано, что для частного инвестора следование за рыночными настроениями не позволяет получить высокую доходность.

Большинство работ по изучению влияния социальных сетей и настроения инвесторов на динамику рыночных торгов ориентировано на западные рынки и не рассматривает структуру и особенности социальных сетей, популярных у российских инвесторов. Несмотря на наличие отдельных

исследований по влиянию сентимента инвесторов на динамику российского фондового рынка, эмпирическая база остается ограниченной. В особенности недостаточно изучены публикации в социальной сети «Пульс», активно используемой частными инвесторами в России. Также не сформированы подходы к количественной оценке влияния социальной активности на краткосрочные рыночные аномалии. Указанные пробелы определяют необходимость дальнейшего анализа поведенческих факторов и особенностей принятия решений российскими инвесторами.

2.2. Роль машинного обучения в выявлении рыночных аномалий

С развитием методов машинного обучения и искусственного интеллекта они стали все чаще применяться для исследования причин и закономерностей возникновения аномалий на фондовом рынке. В отличие от традиционных эконометрических подходов, ML-модели способны выявлять скрытые и нелинейные зависимости в поведении цен и объемов торгов акций, что делает их особенно полезными в задачах обнаружения аномалий. Особенный интерес в контексте тематики социального разгона акций представляют исследования, посвященные выявлению рыночных манипуляций.

Islam et al. [24] исследовали возможность выявления случаев инсайдерской торговли с использованием модели машинного обучения LSTM RNN (рекуррентые нейронные сети). На основе данных о подозрительных операциях авторы показали, что поведенческие признаки в сочетании с ML-классификаторами позволяют достичь высокой точности в идентификации незаконных инсайдерских сделок.

Liu et al. [25] использовали реальные кейсы рыночных манипуляций, раскрытые Китайской комиссией по регулированию ценных бумаг. Построенные модели машинного обучения (SVM и модель логистической регрессии) позволили точно классифицировать эпизоды манипуляций на основе рыночных индикаторов, таких как объем и волатильность торгов, а также отклонение от тренда. Особенностью работы является включение в модель выявления манипуляций переменной рыночного сентимента, которая строится на основе анализа рейтинговых отчетов, финансовых новостей, а также сообщений финансового форума Guba.

Chong et al. [26] сравнили эффективность моделей глубинного обучений (deep learning networks) и классических методов прогнозирования цен на примере фондового рынка Южной Кореи. Результаты продемонстрировали явное преимущество моделей глубокого обучения (deep learning) в исследовании сложных рыночных паттернов, особенно на высокочастотных данных. Авторы приходят к выводу, что модели глубинного обучения могут быть эффективно использованы для анализа и прогнозирования фондового рынка.

Li et al. [27] сосредоточились на сравнении различных моделей классификации для автоматического выявления случаев манипулирования ценами акций на фондовом рынке Китая. Исследование показало, что наибольшую

эффективность по всем показателям удалось достичь моделям KNN (метод *k*-ближайших соседей) и DT (дерево решений).

Nam, Skillicorn [28] исследовали выявление манипуляций по схеме «pump and dump» на платформе Reddit. На основе 18 555 постов и 312 578 комментариев с использованием методов машинного обучения (XGBoost, Random Forest, SVM, MLP, CNN, BiLSTM) было показано, что наиболее эффективным является подход на основе сверточной нейронной сети (CNN). При этом автор отмечает, что анализ комментариев без учета согласия или несогласия инвесторов с предлагаемыми инвестиционными идеями может снизить точность выявления манипуляций.

Domashova et al. [29] проанализировали аномальные сделки с акциями на Московской бирже и предложили модели выявления операций, потенциально связанных с манипулированием рынком или использованием инсайдерской информации. В исследовании сформулированы ключевые признаки подозрительной торговой активности, включая высокую концентрацию сделок перед резкими ценовыми изменениями, аномально большую долю инвестора в дневном объеме торгов по конкретной бумаге, а также регулярные однонаправленные операции крупным объемом и др.

Несмотря на высокую точность и широкие возможности применения, рассмотренные исследования преимущественно фокусируются на торговых паттернах и не учитывают влияние социальной активности инвесторов, также в исследованиях редко встречается добавление в модель переменных технического анализа, что может улучшить их предсказательную способность. Кроме того, большинство из них опираются на данные зарубежных рынков и не адаптированы под российскую специфику торгов.

2.3. Исследования эффекта социального разгона акций

В контексте текущего исследования особый интерес представляют работы, посвященные изучению «социального разгона» — ситуаций, когда активность инвесторов в социальных сетях приводит к аномальному росту цен и объемов торгов отдельных акций. Этот эффект часто возникает без фундаментальных причин и может рассматриваться как форма поведенческой аномалии или в крайних случаях рыночной манипуляции.

Münster et al. [30] на основе данных торгов на платформе Robinhood и активности в социальной сети Reddit показывают, что публикации в социальных сетях способны вызывать резкий рост торгов отдельными акциями и приводить к краткосрочному эффекту социального разгона. Также, согласно исследованию, влияние публикаций в социальных сетях способно значительно превышать силу влияния традиционных новостей.

Warkulat, Pelster [31] исследовали влияние публикаций в сообществе WallStreetBets (Reddit) на рискованность торгов и пришли к выводу, что в период высокой активности инвестиционного сообщества на платформе Reddit частные инвесторы склонны принимать решения на основе публикаций

и комментариев в социальной сети, а не на основе аналитики и советов инвестиционных аналитиков. Торговля в соответствии с трендами сообщества WallStreetBets повышает уровень риска портфеля и приводит к отрицательной доходности за период удержания позиции.

Keasey et al. [32] исследовали влияние публикаций популярных аккаунтов в социальной сети Instagram (более 1 млн подписчиков) на динамику торгов акций отдельных компаний. Авторы доказали, что социальная сеть Instagram является неподходящим каналом для манипулирования и социального разгона акций, так как публикации влияют на внимание инвесторов, волатильность и объем торгов, но не оказывают существенного влияния на доходность акций.

Semenova, Winkler [33] доказали существование эффекта «социального заражения» на основе анализа публикаций социальной сети Reddit: волны позитивных публикаций в социальных сетях приводят к синхронизированным покупкам частных инвесторов и краткосрочным всплескам цен акций. Авторы приходят к выводу, что спрос розничных трейдеров следует за обсуждениями WallStreetBets (Reddit) и объясняет резкий рост торговой активности по акциям отдельных компаний.

Таким образом, современная литература подтверждает наличие эффекта социального разгона и подчеркивает значимость социальных сетей как одной из причин краткосрочных рыночных аномалий. Однако большинство исследований ориентированы на зарубежные рынки, прежде всего США и Китай, и редко предлагают универсальные алгоритмы выявления разгона акций, которые могут быть использованы для других рынков. В российском контексте, учитывая специфику торгов и социальные сети, в которых инвесторы обмениваются информацией, проблема социального разгона акций остается слабо изученной.

2.4. Отличительные характеристики исследования

Основные отличия нашего исследования.

1. Формирование уникальной базы данных случаев социального разгона акций, включающей в себя 615 случаев разгона и 6 765 наблюдений: 615 дней социального разгона и 6 150 нормальных торговых дней, что позволяет получить более достоверные результаты и сделать корректные выводы о достоверности моделей для российского рынка. Выборка данных включает в себя больший набор данных по сравнению с работами других исследователей [9, 34].

2. Разработка авторской методики выявления случаев социального разгона акций, основанной на применении 4 моделей машинного обучения и классической модели логистической регрессии.

3. Учет в классификационных моделях по выявлению случаев социального разгона акций авторских индексов интенсивности обсуждения акций в социальных сетях Russian social media intensive index (RSMII) и сентимента частных инвесторов Russian social media sentiment index (RSMSI).

4. По сравнению с другими исследованиями [14, 26], в которых прогнозные модели цен и классификационные модели выявления разгона акций включаются только изменения биржевых показателей акций, для повышения точности моделей нами дополнительно используются показатели технического анализа.

3. Данные и методы

3.1. Данные

В исследовании используются данные по 104 российским публичным компаниям за период с 2019 по 2025 г. Данные по всем компаниям являются полными, выборка сбалансирована. Для сбора информации использовались парсеры данных, которые были написаны на языке Python специально для проведения исследования, а также базы данных finam.ru с торговыми показателями по каждой компании. Исследование производится на дневном интервале (таймфрейме) и включает 6 765 наблюдений.

Для построения индекса сентимента используются сообщения из социальных сетей «Пульс» и «Телеграм». Данные из социальной сети «Пульс» собраны за 2019–2025 гг. и включают в себя 601 297 публикаций, данные из социальной сети «Телеграм» за период 2015–2025 гг. — 234 926 публикаций.

В табл. 1 представлены описательные статистики по выборке исследования.

Таблица 1. **Описательные статистики за период 2019–2025 гг. (число наблюдений = 6 765)**

Table 1. **Descriptive statistics for the period 2019–2025 (number of observations = 6765)**

Фаза	Показатель	Число наблюдений	Среднее значение	Минимальное значение	Максимальное значение	Стандартное отклонение
5 дней до разгона	*OPEN*	3 075	1235	0,105	20300	2108
	HIGH	3 075	1255	0,106	20650	2136
	LOW	3 075	1216	0,104	19800	2081
	CLOSE	3 075	1238	0,105	20150	2112
	VOLUME	3 075	28 678 307	2	7 864 390 000	256 098 936
	High-LowClose	3 075	3,3 %	0,2 %	31,2 %	2,7 %
	PriceChangeDay	3 075	0,2 %	–16,5 %	21,5 %	2,4 %
	PriceChange PreviousDay	3 075	0,4 %	–17,3 %	23,4 %	2,5 %

Окончание табл. 1

Фаза	Показатель	Число наблюдений	Среднее значение	Минимальное значение	Максимальное значение	Стандартное отклонение
	VOLChange	3 075	46,4 %	–99,0 %	8 972,1 %	301,6 %
	IndexIntensive	3 075	27	0	407	47
	IndexSentiment	3 075	1	–50	175	8
День разгона	*OPEN*	615	1 251	0,109	20 100	2 138
	HIGH	615	1 304	0,115	22 650	2 277
	LOW	615	1 230	0,106	19 800	2 107
	CLOSE	615	1 277	0,111	21 800	2 225
	VOLUME	615	72 071 813	154	12 443 370 000	675 525 520
	High-LowClose	615	5,9 %	0,7 %	37,7 %	5,3 %
	PriceChangeDay	615	2,2 %	–9,5 %	41,6 %	4,7 %
	PriceChange PreviousDay	615	2,5 %	–8,2 %	40,4 %	4,9 %
	VOLChange	615	546,3 %	–74,9 %	80 236,4 %	3 618,8 %
	IndexIntensive	615	50	0	709	75
	IndexSentiment	615	6	–68	99	15
5 дней после разгона	*OPEN*	3 070	1 274	0,102	24 500	2 234
	HIGH	3 070	1 297	0,104	25 850	2 300
	LOW	3 070	1 247	0,101	23 300	2 180
	CLOSE	3 070	1 271	0,102	24 600	2236
	VOLUME	3 070	52 206 227	10	17 778 216 000	507 849 993
	High-LowClose	3 070	4,0 %	0,4 %	44,5 %	3,5 %
	PriceChangeDay	3 070	–0,1 %	–20,9 %	38,5 %	3,2 %
	PriceChange PreviousDay	3 070	0,2 %	–22,4 %	39,9 %	3,3 %
	VOLChange	3 070	2,8 %	–97,3 %	1 745,1 %	94,3 %
	IndexIntensive	3 070	35	0	1 238	88
	IndexSentiment	3 070	1	–95	335	12

Источник: рассчитано автором на основании данных торгов с сайта Finam.

3.2. Методология

3.2.1. Методология проверки гипотезы H1

В рамках проверки первой гипотезы исследовалась причинно-следственная связь между показателями интенсивности обсуждений отдельных акций и сентимента частных инвесторов с резким изменением торговых показателей от нормы.

Методология проверки гипотезы *H1* включает семь этапов.

1. Сбор (парсинг) сообщений по популярным среди российских частных инвесторов «Телеграм» каналам финансово-экономической тематики. С помощью python и API «Телеграм» было собрано 234 926 публикаций за период с ноября 2015 г. (дата самой ранней публикации) по март 2025 г.

2. Анализ публикаций и отбор «Телеграм»-каналов, в которых явным образом выделяются торговые идеи: в публикациях торговых идей есть теги «идея», «торговая идея», «оценка» либо фигурируют слова «рекомендация», «рекомендуем», «покупка», «покупать», «инвестидея», «инвест идея». Всего по двум выбранным каналам было проанализировано 83 079 публикаций, из которых 5 927 содержали ключевые слова, указывающие на то, что сообщение является торговой идеей либо направлено на призыв покупать акции конкретной компании.

3. Отбор публикаций с целью выделения только инвестиционных идей, которые могли быть использованы участниками рынка как сигнал на покупку акций: 1) исключили из выборки все публикации, не относящиеся к российским компаниям и не торгующимся на Московской бирже (MOEX); 2) удалили инвестиционные идеи, направленные на short (продажу акций), так как данная торговая стратегия не свойственна широкому кругу российских инвесторов; 3) удалили все дублирующиеся сообщения, повторные публикации одной торговой идеи за короткий срок; 4) удалили все сообщения, в которых упоминались более одной компании, так как такие сигналы не являются явными и могут не восприниматься инвесторами как торговая идея (все публикации в которых речь шла не более чем о двух компаниях были размечены как один конкретный тикер в зависимости от количества ключевых слов по каждой компании, публикации с тремя и более компаниями были отсмотрены и размечены вручную).

Всего в итоговую выборку включено 615 публикаций, которые могли оказать существенное влияние на динамику торгов акциями и являться социальным разгоном.

4. Сбор датасета с торговыми показателями компаний (цены открытия и закрытия торгов, минимальная и максимальная цена за торговый период, объемы торгов) в день социального разгона, а также за 10 дней до и после разгона. Сбор торговых показателей производился с помощью языка программирования Python путем парсинга дневных торговых показателей с сайта finam.ru. Всего была собрана информация по 12 915 торговым дням.

5. Объединение в единый датасет информации о торгах, отметок о днях социального разгона, а также днях до и после разгона, данных об интенсивности обсуждений и сентимента инвесторов в социальных сетях.

Интенсивность обсуждений в социальных сетях и сентимент инвесторов представлены двумя авторскими индексами: индекс интенсивности обсуждений *RSMII* (Russian social media intensive index) и индекс сентимента частных инвесторов *RSMSI* (Russian social media sentiment index).

Индекс интенсивности обсуждений (*TgIndex*) рассчитывается на основании анализа публикаций из социальных сетей «Телеграм» и «Пульс» по следующим формулам:

$$TgIndex = \sum\left(\left(\frac{NViews}{10000}\right) + \frac{NReposts}{NViews \cdot RepostFreq}\right), \quad (1)$$

где Σ — суммирование значений индекса интенсивности обсуждений для каждого сообщения по отдельной компании-эмитенту акций за определенный период (день), *NViews* — количество просмотров сообщения, *NReposts* — количество репостов, *RepostFreq* — средний процент переадресованных сообщений по отношению к просмотрам в данной социальной сети (для выборки исследования = 0,00194).

Индекс сентимента частных инвесторов (*PulseIndex*) рассчитывается:

$$PulseIndex = \Sigma\left(NComments\right), \quad (2)$$

где *NComments* — количество комментариев к сообщению.

Далее каждый из индексов группируется по торговым дням для каждой акции и вычисляется итоговое значение авторского индекса интенсивности обсуждений *RSMII* (Russian social media intensive index):

$$RSMII = TgIndex + PulseIndex. \quad (3)$$

Авторский индекс сентимента настроений инвесторов *RSMSI* (Russian social media sentiment index), представляющий собой сумму двух индексов Telegram Sentiment Index (*TgSIndex*) и Pulse Sentiment Index (*PulseSIndex*), рассчитывается по следующим формулам:

$$TgSIndex = \sum\left(\begin{array}{c} Sentiment\left(\frac{NViews}{1000}\right) \times \\ \times\left(1 + \frac{NLikes}{NViews \cdot EngageLevel} + \frac{NReposts}{NViews \cdot RepostFreq}\right) \end{array}\right), \quad (4)$$

где Σ — суммирование значений сентимента для каждого сообщения по отдельной компании-эмитенту акций за определенный период (ч);

Sentiment — величина, характеризующая тональность сообщения и равная 1, если сообщение имеет положительную окраску и –1, если отрицательную; *NViews* — число просмотров сообщения; *EngageLevel* — средний процент реакций по отношению к просмотрам в данной социальной сети; *NLikes* — разность числа положительных и отрицательных реакций; *NReposts* — число репостов; *RepostFreq* — средний процент переадресованных сообщений по отношению к просмотрам в данной социальной сети.

$$PulseSIndex = \sum \left(Sentiment \cdot NLikes \cdot NComments \right), \quad (5)$$

где *NComments* — число комментариев к сообщению.

Для построения указанных индексов по социальной сети «Пульс» были собраны публикации за период с сентября 2019 по март 2025 г., в которых фигурировало название или тикеры компаний, которые попали в итоговую выборку нашего исследования (104 российских компании). Всего собрано 991 742 публикации. Далее из выборки были исключены все сообщения, для которых тикер и компанию, о которой пишут в публикации, невозможно однозначно определить, а также сообщения, в которых говорится о трех и более компаниях. Всего по целевым компаниям в итоговую выборку попало 601 207 сообщений.

В социальной сети «Телеграм» произведен сбор сообщений из 18 популярных каналов инвестиционной тематики, в которых с различной периодичностью публикуются торговые идеи. Общая аудитория всех каналов составляет более 2 млн чел. Всего собрано 234 926 сообщений за период с ноября 2015 (дата самой ранней публикации) по март 2025 г. С помощью специально подготовленного словаря каждое сообщение было размечено, ему присвоен тикер конкретной компании. В финальную выборку попали только сообщения по компаниям, которые фигурируют в исследовании, всего 77 087 публикаций.

6. Для исключения влияния корпоративных событий, способных объяснить рыночные аномалии независимо от активности в социальных сетях, из выборки удалены все наблюдения, по которым в день потенциального разгона или в течение трех предшествующих дней происходили инсайдерские сделки, выплаты дивидендов или публикация финансовой отчетности. Информация о таких событиях собрана с официального сайта раскрытия информации e-disclosure.ru.

7. Проведение статистических тестов значимости факторов обсуждения компаний в социальных сетях, а также отличий рыночной динамики торгов в дни социального разгона и в период трех дня до разгона. Также мы построили модель предсказания нормального изменения цены и объема торгов на основе периода за 10 дней до дня социального разгона и сравнили отклонение от нормы в дни разгона, дополнительно проведя статистическую проверку значимости результатов.

3.2.2. Методология проверки гипотезы H2

Для проверки второй гипотезы построены шесть моделей выявления случаев социального разгона и произведено сравнение их эффективности.

1. Методология проверки гипотезы H2 включает в себя пять этапов.

Расчет дополнительных торговых индикаторов для последующего добавления в модели выявления социального разгона акций. Всего было рассчитано четыре группы индикаторов: 1) показатели, характеризующие динамику изменения торговых показателей по дням: процентное изменение цены акций за день (разница между ценой открытия и закрытия), отношение разницы максимальной и минимальной цены за день к цене закрытия, процентное изменение цены акций относительно предыдущего торгового дня, процентное изменение объема торгов относительно предыдущего торгового дня, процентное изменения цен открытия, закрытия, максимальной и минимальной цен относительно показателей предыдущего дня; 2) технические индикаторы: процентные изменения экспоненциальных скользящих средних за 12 и 26 дней, быстрая MACD, сигнальная MACD, Stohastic, Momentum, RSI; 3) dummy-переменные (табл. 2), характеризующие динамику изменения торговых показателей по дням с лагом или с накопительным эффектом: нормализованное закрытие цены, нормализованный объем, Z-оценка объема, скользящее среднее объема, импульс объема, корреляция цены и объема, разрыв цен открытия, внутридневной диапазон, дивергенция тренда объема, волатильность, лагированная доходность (1, 2, 3 дня), изменение объема (1, 2, 3 дня), нормализованный объем (альтернативный), нормализованное скользящее среднее объема, нормализованный импульс объема.

2. Анализ корреляционных матриц всех независимых переменных и исключение из модели парных переменных с коэффициентом корреляции выше 0,7 для недопущения мультиколлинеарности в модели.

3. Объединение в единый датасет торговые показателей и индикаторов, сигналы технического анализа, а также индексов обсуждаемости акций, сентимента инвесторов в социальных сетях по каждой компании.

4. Построение классификационных моделей выявления социального разгона акций: логистическая регрессия (Logistic Regression), случайный лес (Random Forest), метод опорных векторов (SVM), k ближайших соседей (KNN), продвинутый градиентный бустинг (CatBoost модель, разработанная Яндексом).

5. Оценка качества надежности моделей по следующим метрикам:

Accuracy — показатель, который описывает общую точность предсказания модели по всем классам. Рассчитывается как отношение количества правильно размеченных наблюдений к их общему количеству.

Recall — указывает на долю правильно найденных положительных объектов (манипуляций) среди всех объектов положительного класса.

Таблица 2. **Dummy-переменные модели, характеризующие динамику изменения торговых показателей по дням**
Table 2. **Dummy variables of the model characterizing the dynamics of changes in trading indicators by day**

Dummy-переменная	Обозначение и формула расчета	Объяснение значения переменной
Нормализованное закрытие	$\text{NORM\_CLOSE} = \frac{\text{CLOSE-MIN}(\text{CLOSE})}{\text{MAX}(\text{CLOSE}) - MIN(CLOSE)}$	Показывает, насколько цена выше минимума в выбранном окне
Нормализованный объем	$\text{NORM\_VOLUME} = \frac{\text{VOLUME-MIN}(\text{VOLUME})}{\text{MAX}(\text{VOLUME}) - MIN(\text{VOLUME})}$	Относительная величина объема по сравнению с экстремумами
Z-оценка объема	$VOLUME\_ZSCORE = \frac{VOLUME - \mu}{\sigma}$	Показывает, насколько объем больше или меньше среднего значения за период
Скользящее среднее объема	$\text{VOLUME\_MA} = \frac{\sum_{i=1}^{n} \text{VOLUME}_{t-i}}{N}$	Среднее значение объема за последние *N* дней
Импульс объема	$VOLUME\_SPIKE = \frac{VOLUME}{VOLUME\_MA}$	Показывает, насколько текущий объем выше/ниже среднего
Корреляция цены и объема	$PRICE\_VOLUME\_CORR = Correlation(CLOSE, VOLUME)$	Степень связи между ценой и объемом за период
Разрыв открытия	$GAP = OPEN_t - CLOSE_{t-1}$	Разница между ценой открытия и вчерашней ценой закрытия
Внутридневной диапазон	$INTRADAY\_RANGE - HIGH - LOW$	Разница максимальной и минимальной цены внутри дня
Дивергенция тренда объема	$VOLUME\_TREND\_DIVERGENSE = VOLUME - VOLUME\_TREND$	Показывает, насколько объем отклоняется от тренда

Окончание табл. 2

Dummy-переменная	Обозначение и формула расчета	Объяснение значения переменной
Волатильность	$VOLATILITY = std\left(\frac{CLOSE_t}{CLOSE_{t-1}}\right) - 1$	Показывает, насколько сильно колебалась цена за последние N дней
Доходность с лагом 1 день	$\text{RETURN\_LAG1} = \frac{\text{CLOSE}_{\text{t-1}} - \text{CLOSE}_{\text{t-2}}}{\text{CLOSE}_{\text{t-2}}}$	Доходность на 1 день назад
Доходность с лагом 2 дня	$\text{RETURN\_LAG2} = \frac{\text{CLOSE}_{\text{t-2}} - \text{CLOSE}_{\text{t-3}}}{\text{CLOSE}_{\text{t-3}}}$	Доходность на 2 дня назад
Доходность с лагом 3 дня	$\text{RETURN\_LAG3} = \frac{\text{CLOSE}_{\text{t-3}} - \text{CLOSE}_{\text{t-4}}}{\text{CLOSE}_{\text{t-4}}}$	Доходность на 3 дня назад
Изменение объема лаг 1	$VOLUME\_CHANGE\_LAG1 = VOLUME_{t-1} - VOLUME_{t-2}$	Разница объемов за 1 день
Изменение объема лаг 2	$VOLUME\_CHANGE\_LAG2 = VOLUME_{t-2} - VOLUME_{t-3}$	Разница объемов за 2 дня
Изменение объема лаг 3	$VOLUME\_CHANGE\_LAG3 = VOLUME_{t-3} - VOLUME_{t-4}$	Разница объемов за 3 дня
Лог-норм объем	$VOLUME\_NORM = \frac{\ln(VOLUME)}{\ln(MAX(VOLUME))}$	Логарифмически нормализованный объем
Нормализованный средний объем	$VOLUME\_MA\_NORM = \frac{VOLUME\_MA}{MAX(VOLUME\_MA)}$	Относительное значение MA объема
Нормализованный всплеск объема	$VOLUME\_SPIKE\_NORM = \frac{VOLUME\_SPIKE}{MAX(VOLUME\_SPIKE)}$	Относительный уровень «всплеска» объема

Источник: составлено автором.

ISSN 2712-7435

Precision — представляет собой отношение числа наблюдений, верно классифицированных как Positive, к общему числу наблюдений с меткой Positive (распознанных правильно и неправильно). Precision измеряет точность модели при определении класса Positive.

F1-score — среднее гармоническое значение показателей Recall и Precision. Это сбалансированная мера как ложноположительных, так и ложноотрицательных результатов.

AUC-ROC — определяется как площадь под ROC-кривой. ROC-кривая показывает соотношение доли ложноположительных результатов (ось x) и доли истинно положительных результатов (ось y) для ряда различных пороговых значений в диапазоне от 0,0 до 1,0. Данная метрика показывает, насколько хорошо модель классификации способна разделять и предсказывать классы.

4. Результаты

4.1. Результаты проверки гипотезы H1

Для проверки гипотезы о том, что публикации в социальных сетях могут выступать причиной аномального роста цен и объемов торгов акциями российских компаний, приводя к эффекту социального разгона, был проведен анализ аномалий изменения цен и объемов торгов в дни, когда по акции наблюдался социальный разгон (SOCIALPUMP = 1), и в дни нормального поведения рынка (SOCIALPUMP = 0).

Отклонения рассчитывались как разность между фактическими значениями и прогнозными, полученными на основе модели нормального поведения (построенной по окну из 10 предыдущих торговых дней индивидуально для каждого эпизода): DELTA_CLOSE_PCT — отклонение цены закрытия в процентах; DELTA_VOL — показатель, отражающий, во сколько раз объем торгов отличается от его прогнозного значения.

В качестве обычных торговых дней учитывались три дня до и после социального разгона. Сравнительная статистика по периодам показала следующие результаты (табл. 3).

Таблица 3. **Сравнительная статистика показателей в дни разгона и в период 3 дня до и после разгона**

Table 3. **Comparative statistics of indicators on the days of acceleration and in the period of 3 days before and after acceleration**

Показатель	День разгона (n = 615)	Обычные дни (n = 3 690)
Среднее DELTA_CLOSE_PCT	0,019	0,0003
Среднее DELTA_VOL	4,87	0,10
Стандартное отклонение цены	0,05	0,02
Стандартное отклонение объема	36,30	1,46
Минимальное отклонение цены	–0,16	–0,19

Окончание табл. 3

Показатель	День разгона (n = 615)	Обычные дни (n = 3 690)
Максимальное отклонение цены	0,41	0,20
Среднее IndexIntensive	50,07	34,05
Среднее IndexSentiment	6,46	1,99
Стандартное отклонение IndexIntensive	75,07	73,51
Стандартное отклонение IndexSentiment	14,94	11,58

Источник: рассчитано автором.

Из данных табл. 3 видно, что изменение цены (DELTA_CLOSE_PCT) в дни разгона составляет в среднем 0,019, что в 63 раза выше, чем в обычные дни (0,0003). Это подтверждает наличие резких аномальных ценовых всплесков. Изменение объема торгов (DELTA_VOL) также резко возрастает: 4,87 против 0,10, то есть более чем в 48 раз, что свидетельствует о всплеске интереса участников рынка к акциям в такие дни. Стандартное отклонение цены увеличивается в 2,5 раза (0,05 против 0,02), а стандартное отклонение объема — более чем в 25 раз (36,3 против 1,46), что указывает на значительно более высокую волатильность в период социального разгона.

Средние значения социальных индикаторов также существенно различаются: *IndexIntensive* (интенсивность обсуждений в соцсетях) в дни разгона равен 50,07, тогда как в обычные дни — 34,05; *IndexSentiment* (оценка сентимента частных инвесторов) в дни разгона — 6,46, что более чем в три раза выше, чем в обычные дни (1,99).

Для оценки достоверности различий между днями социального разгона и обычными торговыми днями была проведена *t*-проверка средних значений по четырем ключевым показателям (табл. 4).

Результаты анализа подтверждают наличие статистически значимых отличий:

1. Для изменения цены акций (DELTA_CLOSE_PCT) наблюдается высокая статистическая значимость (t = 9,03; p < 0,000001), что указывает на устойчивое отклонение средней доходности в дни разгона. Цены акций в такие дни демонстрируют аномальный рост, выходящий за пределы обычной рыночной волатильности.
2. Объем торгов (DELTA_VOL) также значительно увеличивается в дни разгона (t = 3,33; p < 0,001), что подтверждает усиленную активность участников рынка в ответ на публикацию сообщений в социальных сетях.
3. Индикатор *IndexIntensive*, отражающий интенсивность обсуждений акций в социальных сетях, имеет статистически значимое превышение в дни разгона (t = 4,86; p < 0,001), что может свидетельствовать о внешнем согласованном воздействии.

Таблица 4. **Результаты статистической проверки значимости отклонения показателей**

Table 4. **Results of statistical testing of the significance of deviation of indicators**

Показатель	t-статистика	p-значение	Вывод
DELTA_CLOSE_PCT	9,03	< 0,000001	Среднее изменение цен в дни социального разгона акций статистически значимо выше нуля. В дни разгона цены акций растут аномально сильно и это не является случайным явлением
DELTA_VOL	3,33	< 0,001	Объем торгов в дни социального разгона значимо превышает норму — наблюдается резкий всплеск активности торгов и роста объемов
IndexIntensive	4,86	<0,001	Интенсивность обсуждений статистически значимо выше в дни разгона
IndexSentiment	7,29	<0,001	Сентимент сообщений в социальных сетях статистически значимо выше в дни разгона

Источник: рассчитано автором.

4. Показатель *IndexSentiment*, отражающий эмоциональную окраску публикаций, также демонстрирует выраженное различие ($t = 7{,}29$; $p < 0{,}001$), фиксируя преимущественно положительный эмоциональный фон в дни социального разгона.

Таким образом, *t*-тест подтвердил, что все ключевые рыночные и социальные признаки имеют статистически значимые отличия в дни предполагаемого информационного воздействия, что позволяет обоснованно говорить о наличии эффекта социального разгона.

Наличие статистической значимости между показателями, а также значительный рост цен и объемов торгов сами по себе еще не говорят о том, что именно публикации сообщений в социальных сетях стали причиной данного роста. Поэтому при интерпретации результатов важно учитывать, что было проделаны дополнительные шаги (подробно указаны в разделе с методологией), направленные на исключение влияния дополнительных факторов: исключены дни с фундаментальными событиями (отчеты, дивиденды, инсайдерские сделки), добавлено предсказания ожидаемого поведение цены и объема, используя данные до разгона (окна в 10 торговых дней) и все отклонения (DELTA_CLOSE_PCT, DELTA_VOL) рассчитаны от модели нормального фона, что исключает естественные рыночные отклонения как причину.

Результаты подтверждают первую гипотезу: в дни социального разгона акции демонстрируют существенно более высокие отклонения от нормального поведения по сравнению с обычными днями. Это проявляется как в ценовых аномалиях, так и в резком росте объема торгов. Статистическая значимости изменений подтверждена.

4.2. Результаты проверки гипотезы H2

В рамках проверки гипотезы H2 о том, что модели машинного обучения способны более эффективно выявлять случаи социального разгона акций по сравнению с традиционными эконометрическими подходами, были построены четыре модели машинного обучения (CatBoost, SVM, Random Forest, KNN), а также модель логистической регрессии. Сравнительный анализ их результативности представлен на рис. 1 и в табл. 5.

Catboost и KNN демонстрируют наилучшие результаты: они правильно обнаруживают более 90 % случаев социального разгона (Recall). Из всех предсказанных случаев разгона 77 % у KNN и 86 % у Catboost действительно являются разгоном акций (метрика Precision). Общий показатель качества AUC-ROC у данных моделей превышает 90 %, что является отличным результатам для моделей классификации.

Random Forest и SVM демонстрируют более низкие результаты: они обнаруживают лишь 72 и 82 % случаев социального разгона соответственно, при этом их точность (79–89 %) несколько ниже, чем у Catboost.

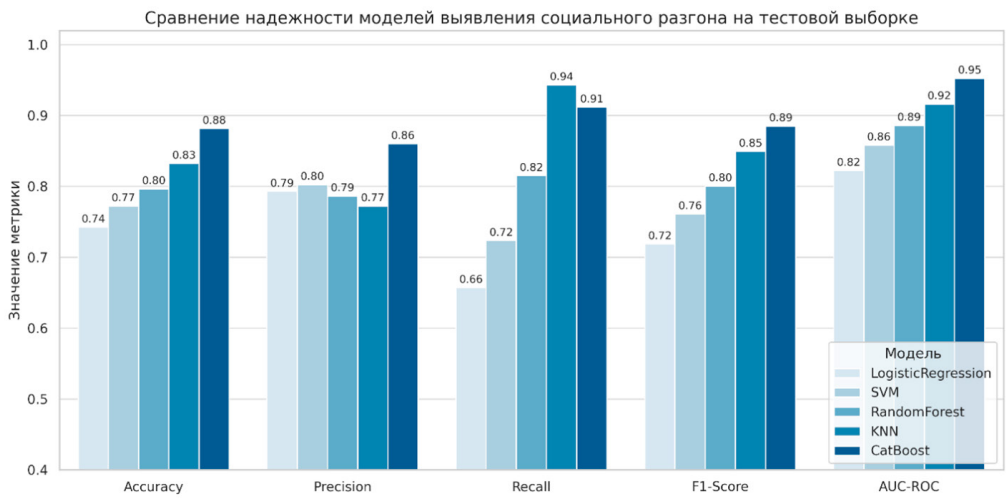

Рис. 1. Результаты надежности моделей выявления социального разгона на тестовой выборке

Figure 1. Results of the reliability of models for identifying social dispersal on a test sample

Источник: составлено автором.

Таблица 5. **Результаты надежности классификационных моделей выявления социального разгона на обучающей и тестовой выборках**

Table 5. **Results of the reliability of classification models for identifying social dispersal on training and test samples**

Модель	Датасет	Accuracy	Precision	Recall	F1-Score	AUC-ROC
CatBoost	Train	0,981	0,973	0,990	0,981	0,998
CatBoost	Test	0,882	0,860	0,912	0,885	0,952
RandomForest	Train	0,844	0,833	0,862	0,847	0,923
RandomForest	Test	0,797	0,787	0,815	0,801	0,886
LogisticRegression	Train	0,739	0,779	0,666	0,718	0,826
LogisticRegression	Test	0,743	0,793	0,657	0,719	0,822
KNN	Train	0,880	0,819	0,974	0,890	0,976
KNN	Test	0,833	0,772	0,943	0,849	0,916
SVM	Train	0,797	0,814	0,771	0,792	0,883
SVM	Test	0,773	0,803	0,724	0,761	0,858

Источник: рассчитано автором.

Модель логистической регрессии (Logistic Regression) работает значительно хуже и уступает другим моделям практически во всех показателях, кроме precision = 79 %, но из-за того, что логистическая регрессия пропускает много случаев социального разгона (recall = 66 %, 34 % случаев не выявлено), высокое значение показателя precision не несет практической ценности.

По результатам тестирования моделей видно, что модели машинного обучения превосходят модель логистической регрессии по большинству показателей, особенно заметно различие по итоговому комплексному показателю ROC-AUC, который на 5–16 % выше у моделей машинного обучения.

Наглядное отличие эффективности можно увидеть при сравнении результатов CatBoost с Logistic Regression: полнота (recall) выше на 25,5 процентных пункта (65,7 против 91,2 %), а интегральный показатель качества AUC-ROC выше на 13 процентных пунктов. Это означает, что модель CatBoost на практике обнаруживает на 25–30 % больше эпизодов социального разгона, при этом с меньшей долей ложных срабатываний. Даже более простые модели, такие как Random Forest и KNN, демонстрируют прирост показателя recall от 16 до 29 процентных пунктов, что делает их значительно более надежными средствами автоматического выявления подозрительной торговой активности в результате социального разгона.

Таким образом гипотеза о том, что модели машинного обучения позволяют более точно выявлять случаи социального разгона акций по сравнению с традиционными эконометрическими моделями, подтвердилась.

5. Обсуждение

5.1. Анализ результатов

Полученные результаты демонстрируют, что в дни социального разгона акции действительно демонстрируют аномально высокие значения прироста цен и объема торгов. В частности, среднее изменение цен оказалось более чем в 60 раз выше по сравнению с обычными днями, а объемы торгов — в 48 раз. Кроме того, показатели социальной активности (интенсивность обсуждений и сентимент инвесторов) также оказались статистически значимо выше в дни разгона. Это подтверждает гипотезу о том, что публикации сообщений в социальных сетях могут являться причиной рыночных аномалий, создавая значительное краткосрочное изменение динамики торгов акций конкретных компаний.

Сравнение с предыдущими исследованиями подтверждает достоверность полученных результатов. Semenova, Winkler [33] доказали, что волны позитивных обсуждений в Reddit способны приводить к резкому росту цен акций. Ranco et al. [35] выявили, что резкий рост упоминаний и обсуждений отдельных компаний индекса Dow Jones в социальной сети Twitter статистически значимо способно вызывать аномальный рост цен акций на 1–2 % даже при отсутствии фундаментальных событий, что подтверждает причинную связь публикаций в социальных сетях с рыночными аномалиями.

Гипотеза о наличии причинной связи между публикациями в социальных сетях и рыночными аномалиями подтвердилась. Рыночные и социальные признаки в дни социального разгона показали статистически значимое отклонение от нормального хода торгов. Это дает основание утверждать, что публикации в социальных сетях могут выступать причиной аномалий в изменение цен и объемов торгов при отсутствии фундаментальных рыночных причин.

Результаты исследования показали, что модели машинного обучения, особенно CatBoost и SVM, демонстрируют высокую точность в задаче выявления эпизодов социального разгона, значительно превосходя надежность модели логистической регрессии. Значения ROC-AUC превышают 0,9, а F1-score по дням социально разгона достигает 0,85–0,89, что свидетельствует о высокой способности моделей распознавать неочевидные аномалии, обусловленные информационным воздействием.

Подобные подходы применялись в зарубежной литературе. Nam, Skillicorn [28] протестировали шесть моделей выявления манипуляций по схеме «pump and dump», наилучшие показатели продемонстрировала модель CNN (сверточная нейронная сеть) с показателем accuracy равным 85 % и F1-score равным 62 %. Li et al. [27] тестировали семь моделей классификации для выявления случаев манипуляций акциями на рынке Китая. Наибольшую эффективность по всем показателям удалось достичь моделям KNN (метод k ближайших соседей) и DT (дерево решений), при этом общий разброс показателей accuracy и AUC-ROC по всем моделям составил

от 89 до 99 %. Это показывает, что включение сентимент-анализа, интенсивности обсуждений и комплексных рыночных признаков, которые были добавлены в рамках данной работы, позволяет достичь более высокой надежности моделей выявления социального разгона акций.

Таким образом, в сравнении с предшествующими работами, предложенные модели машинного обучения обеспечивают более высокую точность и полноту выявления случаев социального разгона.

Гипотеза о превосходстве моделей машинного обучения по сравнению с эконометрическими для задач выявления случаев социального разгона акций также подтвердилась: ROC-AUC лучших моделей превышал 90 %, тогда как логистическая регрессия показала существенно меньшую точность. Это говорит о том, что учет нелинейных взаимодействий и комплексных признаков позволяет достовернее выделять случаи социального разгона.

5.2. Ограничения исследования

Исследование имеет ряд ограничений.

Отсутствие прямых данных о действиях конкретных участников торгов не позволяет с полной уверенностью говорить о преднамеренно социальном разгоне акций отдельным лицом или группой лиц.

Социальные признаки агрегированы по двум социальным сетям «Телеграм» и «Пульс», анализ не учитывает особенности платформ и типов контента, а также публикации по другим социальным сетям и форумам инвесторов (например Smartlab).

Модель нормального поведения торговых показателей акций строилась на окне в 10 дней, что может не учитывать сезонные или внешнеэкономические эффекты.

Построенные модели обучены на исторических данных и не предназначены для непосредственного применения в реальном времени без регулярного обновления выборки и повторного обучения с учетом изменения рыночной среды и активности инвесторов в социальных сетях.

Полученные результаты не гарантируют универсальной применимости за пределами российского фондового рынка или в других временных интервалах, поскольку поведение инвесторов и влияние социальных сетей могут существенно отличаться в других странах и рыночных условиях.

6. Заключение

Цель исследования заключалась в выявлении и эмпирической проверке феномена социального разгона на российском фондовом рынке, то есть случаев, когда публикации в социальных сетях становятся причиной аномального роста цен и объемов торгов акций отдельных компаний. В рамках работы были сформулированы и проверены две гипотезы, а также разработана и протестирована методология, сочетающая рыночные и поведенческие признаки для выявления случаев социального разгона.

Основные результаты исследования содержат следующие выводы:

1. В дни социального разгона акции демонстрируют статистически значимое превышение средних значений прироста цен и объемов торгов. Аномальная динамика торгов в случаях социального разгона связана с публикациями сообщений в социальных сетях и сопровождается резким увеличением интенсивности обсуждений, а также позитивного сентимента частных инвесторов.
2. Построенные модели машинного обучения (CatBoost, SVM) достигли высоких значений ROC-AUC (более 0,9) и F1-score, что позволяет с высокой точностью выявлять случаи социального разгона и отличать их от обычной рыночной активности.

Теоретическая значимость исследования. Работа вносит вклад в развитие поведенческой и институциональной теории финансов, акцентируя внимание на нетрадиционных источниках рыночной информации, таких как социальные сети и сентимент частных инвесторов. Обоснован и эмпирически подтвержден факт влияния публикаций в социальных сетях на аномальное изменение цен и объемов торгов российскими акциями, что расширяет теоретическое понимание природы краткосрочных аномалий на российском фондовом рынке.

Практическая значимость. Разработанные в рамках исследования методы и модели могут быть использованы различными группами участников финансового рынка для повышения прозрачности рынка, предупреждения рыночных манипуляций и построения торговых стратегий.

Финансовый регулятор (Банк России) может использовать предложенную методику для автоматического мониторинга социальных сетей и выявления признаков социального разгона акций. Полученные результаты особенно актуальны для системы превентивного контроля за рыночными манипуляциями, поскольку позволяют оперативно выявлять всплески социальной активности и случаи отклонения рыночной динамики от нормального поведения, не объясняемые фундаментальными факторами. Интеграция предложенной методики с данными об операциях конкретных участников рынка может быть использована как основа для разработки системы автоматического выявления манипуляций, совершаемых с использованием социальных сетей.

Инвестиционные компании и трейдеры могут использовать выявленные признаки социального разгона и методику их определения как часть торговых сигналов, отслеживающих всплески активности в социальных сетях по конкретным бумагам. Это позволит своевременно выявлять потенциальные случаи социального разгона для выбора оптимального времени закрытия позиции, а также для использования краткосрочного импульса в спекулятивных стратегиях покупки акций. Разработчики торговых систем могут внедрить индексы социальной активности и классификаторы социальных аномалий в аналитические модули торговых платформ для оценки вероятности краткосрочного отклонения цен акций от справедливого уровня.

Список использованных источников

1. *Теплова Т. В., Соколова Т. В., Томтосов А. Ф., Бучко Д. В., Никулин Д. Д.* Сентимент частных инвесторов в объяснении различий в биржевых характеристиках акций российского рынка // Журнал Новой экономической ассоциации. 2022. № 1. С. 53–84. https://doi.org/10.31737/2221-2264-2022-53-1-3

2. *Teplova T., Tomtosov A., Sokolova T.* A retail investor in a cobweb of social networks // PLOS ONE. 2022. Vol. 17, Issue 12. e0276924. https://doi.org/10.1371/journal.pone.0276924

3. *Коновалова М. Е.* Инсайдерские сделки и способы их предотвращения на рынке ценных бумаг в России // Креативная экономика. 2023. Т. 17, № 10. С. 3817–3832. https://doi.org/10.18334/ce.17.10.119204

4. *Шаманина Е. И., Сероштан Е. С.* Манипулятивные сделки: сущность, анализ рисков и подходы к выявлению недобросовестных практик на российском рынке ценных бумаг // Вестник Алтайской академии экономики и права. 2024. № 7–2. С. 389–394. https://doi.org/10.17513/vaael.3611

5. *Sridhar S., Sanagavarapu S.* ELM-AD: Extreme Learning Machine Framework for Price and Volume Anomaly Detection in Stock Markets // Proceedings 2021 International Conference on Computing and Communications Applications and Technologies (I3CAT). IEEE, 2021. Pp. 44–51. https://doi.org/10.1109/I3CAT53310.2021.9629409

6. *Tallboys J., Zhu Y., Rajasegarar S.* Identification of Stock Market Manipulation with Deep Learning // Proceedings of 17th International Conference on Advanced Data Mining and Applications (ADMA 2022). Part I / edited by B. Li, L. Yue, J. Jiang, W. Chen, X. Li, G. Long, F. Fang, H. Yu. Springer Cham, 2022. Pp. 408–420. https://doi.org/10.1007/978-3-030-95405-5_29

7. *Uslu N. C., Akal F.* A Machine Learning Approach to Detection of Trade-Based Manipulations in Borsa Istanbul // Computational Economics. 2022. Vol. 60. Pp. 25–45. https://doi.org/10.1007/s10614-021-10131-8

8. *Ueda K., Suwa H., Yamada M., Ogawa Y., Umehara E., Yamashita T., Tsubouchi K., Yasumoto K.* SSCDV: Social media document embedding with sentiment and topics for financial market forecasting // Expert Systems with Applications. 2024. Vol. 245. 122988. https://doi.org/10.1016/j.eswa.2023.122988

9. *Li Y., Bu H., Li J., Wu J.* The role of text-extracted investor sentiment in Chinese stock price prediction with the enhancement of deep learning // International Journal of Forecasting. 2020. Vol. 36, Issue 4. Pp. 1541–1562. https://doi.org/10.1016/j.ijforecast.2020.05.001

10. *Lin P., Ma S., Fildes R.* The extra value of online investor sentiment measures on forecasting stock return volatility: A large-scale longitudinal evaluation based on Chinese stock market // Expert Systems with Applications. 2024. Vol. 238, Part B. 121927. https://doi.org/10.1016/j.eswa.2023.121927

11. *Mili M., Sahut J., Teulon F., Hikkerova L.* A multidimensional Bayesian model to test the impact of investor sentiment on equity premium // Annals of Operations Research. 2024. Vol. 334. Pp. 919–939. https://doi.org/10.1007/s10479-023-05165-0

12. *Dixon M. F., Halperin I., Bilokon P.* Machine Learning in Finance: From Theory to Practice. Springer, 2020. 548 p. https://doi.org/10.1007/978-3-030-41068-1

13. *Cai Y., Tang Z., Chen Y.* Can real-time investor sentiment help predict the high frequency stock returns? // North American Journal of Economics and Finance. 2024. Vol. 72. 102147. https://doi.org/10.1016/j.najef.2024.102147

14. *Gupta R., Nel J., Pierdzioch C.* Investor confidence and forecastability of US stock market realized volatility: Evidence from machine learning // Journal of Behavioral Finance. 2023. Vol. 24, Issue 1. Pp. 111–122. https://doi.org/10.1080/15427560.2021.1949719

15. *Chen H., De P., Hu Y. J., Hwang B. H.* Wisdom of crowds: The value of stock opinions transmitted through social media // The Review of Financial Studies. 2013. Vol. 27, Issue 5. Pp. 1367–1403. https://doi.org/10.1093/rfs/hhaa114

16. *Fama E. F.* Efficient Capital Markets: A Review of Theory and Empirical Work // The Journal of Finance. 1970. Vol. 25, No. 2. Pp. 383–417. https://doi.org/10.2307/2325486

17. *Samuelson P. A.* Proof that Properly Anticipated Prices Fluctuate Randomly // Industrial Management Review. 1965. Vol. 6, No. 2. Pp. 41–49. URL: https://capital-gain.ru/assets/docs/Proof-that-properly-anticipated-prices-fluctuate-randomly.pdf

18. *Shiller R. J.* Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? // The American Economic Review. 1981. Vol. 71, No. 3. Pp. 421–436. URL: https://www.aeaweb.org/aer/top20/71.3.421-436.pdf

19. *Baker M., Wurgler J.* Investor sentiment and the cross-section of stock returns // The Journal of Finance. 2006. Vol. 61, Issue 4. Pp. 1645–1680. https://doi.org/10.1111/j.1540-6261.2006.00885.x

20. *Baker M., Wurgler J.* Investor Sentiment in the Stock Market // Journal of Economic Perspectives. 2007. Vol. 21, No. 2. Pp. 129–151. https://doi.org/10.1257/jep.21.2.129

21. *Niu H., Pan Q., Xu K.* Hybrid deep learning models with multi-classification investor sentiment to forecast the prices of China's leading stocks // PLoS ONE. 2023. Vol. 18, Issue 11. e0294460. https://doi.org/10.1371/journal.pone.0294460

22. *Liu Q., Lee W. S., Huang M., Wu Q.* Synergy between stock prices and investor sentiment in social media // Borsa Istanbul Review. 2023. Vol. 23, Issue 1. Pp. 76–92. https://doi.org/10.1016/j.bir.2022.09.006

23. *Cookson J. A., Lu R., Mullins W., Niessner M.* The social signal // Journal of Financial Economics. 2024. Vol. 158. 103870. https://doi.org/10.1016/j.jfineco.2024.103870

24. *Islam S. R., Ghafoor S. K., Eberle W.* Mining illegal insider trading of stocks: A proactive approach // Proceedings 2018 IEEE International Conference on Big Data. IEEE, 2018. Pp. 1397–1406. https://doi.org/10.1109/BigData.2018.8622303

25. *Liu Q., Wang C., Zhang P., Zheng K.* Detecting stock market manipulation via machine learning // International Review of Financial Analysis. 2021. Vol. 78. 101887. https://doi.org/10.1016/j.irfa.2021.101887

26. *Chong E., Han C., Park F. C.* Deep learning networks for stock market analysis and prediction // Expert Systems with Applications. 2017. Vol. 83. Pp. 187–205. https://doi.org/10.1016/j.eswa.2017.04.030

27. *Li A., Wu J., Liu Z.* Market manipulation detection based on classification methods // Procedia Computer Science. 2017. Vol. 122. Pp. 788–795. https://doi.org/10.1016/j.procs.2017.11.438

28. *Nam D., Skillicorn D. B.* Detecting pump & dump stock market manipulation from online forums // arXiv. 2023. 2301.11403. https://doi.org/10.48550/arXiv.2301.11403

29. *Domashova J., Yakimov D., Bredikhin D., Gorbunov K., Slavik R., Kadyrov I.* Detection and analysis of atypical stock transactions with possible misuse of insider information and market manipulation: methods and models // Procedia Computer Science. 2022. Vol. 213. Pp. 165–174. https://doi.org/10.1016/j.procs.2022.11.052

30. *Münster J., Reichenbach M., Walther T.* Robinhood, Reddit, and the news: The impact of traditional and social media on retail investor trading // Journal of Financial Markets. 2024. Vol. 71. 100929. https://doi.org/10.1016/j.finmar.2024.100929

31. *Warkulat M., Pelster M.* Social media attention and retail investor behavior: Evidence from r/wallstreetbets // International Review of Financial Analysis. 2024. Vol. 96, Part B. 103721. https://doi.org/10.1016/j.irfa.2024.103721

32. *Keasey K., Lambrinoudakis C., Mascia D. V., Zhang Z.* The impact of social media influencers on the financial market performance of firms // European Financial Management. 2025. Vol. 31, Issue 2. Pp. 745–785. https://doi.org/10.1111/eufm.12513

33. *Semenova M., Winkler A.* Reddit's self-organized bull runs: Social contagion, sentiment, and risk // arXiv 2021. 2104.01847. https://doi.org/10.48550/arXiv.2104.01847

34. *Xu L., Xue C., Zhang J.* The impact of investor sentiment on stock liquidity of listed companies in China // Investment Management and Financial Innovations. 2024. Vol. 21, Issue 2. Pp. 1–14. https://doi.org/10.21511/imfi.21(2).2024.01

35. *Ranco G., Aleksovski D., Caldarelli G., Grčar M., Mozetič I.* The effects of Twitter sentiment on stock price returns // PLOS ONE. 2015. Vol. 10, Issue 9. e0138441. https://doi.org/10.1371/journal.pone.0138441

ИНФОРМАЦИЯ ОБ АВТОРЕ

Хазиев Глеб Андреевич

Аспирант базовой кафедры инфраструктуры финансовых рынков факультета экономических наук, стажер-исследователь Центра финансовых исследований и анализа данных Национального исследовательского университета «Высшая школа экономики». г. Москва, Россия (109028, г. Москва, Покровский бульвар, 11); ORCID https://orcid.org/0000-0003-3346-0006 e-mail: glebhaziev@mail.ru

ДЛЯ ЦИТИРОВАНИЯ

Хазиев Г. А. Выявление социального разгона акций с использованием машинного обучения: эмпирический анализ российского рынка // Journal of Applied Economic Research. 2025. Т. 24, № 4. С. 1445–1474. https://doi.org/10.15826/vestnik.2025.24.4.047

ИНФОРМАЦИЯ О СТАТЬЕ

Дата поступления 9 июня 2025 г.; дата поступления после рецензирования 30 июля 2025 г.; дата принятия к печати 13 августа 2025 г.

Detecting Social Stock Pumping Using Machine Learning: Empirical Evidence from the Russian Market

Gleb A. Khaziev
National Research University Higher School of Economics, Moscow, Russia
glebhaziev@mail.ru

Abstract. In recent years, the Russian stock market has experienced a growing influence of social media on the behavior of retail investors, which creates preconditions for short-term anomalies in stock prices and trading volumes. While the role of social media publications in driving stock dynamics has intensified since 2019, no comprehensive academic or practical studies have yet explored this phenomenon in depth, which highlights the relevance of the present research. The study aims to identify and quantitatively assess cases of social stock pumping — situations where spikes in discussions and sentiment on social media precede abnormal increases in prices and trading volumes of individual securities. The research tests two hypotheses: first, that social media activity can serve as a causal factor in market anomalies; and second, that machine learning models outperform traditional econometric approaches in detecting such cases. The analysis is based on data for 104 Russian companies covering the period from 2019 to 2025, incorporating market features along with aggregated indices of discussion intensity and investor sentiment. The methodology includes constructing a model of normal stock behavior based on historical data and calculating deviations on days identified as potential social pumping events. For classification, four machine learning models (CatBoost, SVM, Random Forest, KNN) and logistic regression were applied. The results confirm the existence of persistent anomalies in price and volume behavior on social pumping days, along with a statistically significant rise in social indicators. Machine learning models achieved ROC-AUC scores exceeding 0.9, thus outperforming many earlier published approaches. The findings contribute to the theoretical development of behavioral finance and offer practical tools for regulators, analysts, and traders to detect market anomalies and incorporate social influence into trading strategies.

Key words: stock market; stocks; manipulation; social pump; machine learning; investor sentiment.

JEL G12, G14

References

1. Teplova, T.V., Sokolova, T.V., Tomtosov, A.F., Buchko, D.V., Nikulin, D.D. (2022). The sentiment of private investors in explaining the differences in the trade characteristics of the Russian market stocks. *Journal of the New Economic Association,* 53–84. (In Russ.). https://doi.org/10.31737/2221-2264-2022-53-1-3

2. Teplova, T., Tomtosov, A., Sokolova, T. (2022). A retail investor in a cobweb of social networks. *PLOS ONE,* Vol. 17, Issue 12. e0276924. https://doi.org/10.1371/journal.pone.0276924

3. Konovalova, M.E. (2023). Insider trading and methods of its prevention in the Russian securities market. *Creative Economy,* Vol. 17, No. 10, 3817–3832. (In Russ.). https://doi.org/10.18334/ce.17.10.119204

4. Shamanina, E.I., Seroshtan, E.S. (2024). Manipulative transactions: Essence, risk analysis, and approaches to identifying unfair practices in the Russian securities market. *Bulletin of the Altai Academy of Economics and Law,* No. 7–2, 389–394. (In Russ.). https://doi.org/10.17513/vaael.3611

5. Sridhar, S., Sanagavarapu, S. (2021). ELM-AD: Extreme Learning Machine Framework for Price and Volume Anomaly Detection in Stock Markets. *Proceedings 2021 International Conference on Computing and Communications Applications and Technologies (I3CAT).* IEEE, 44–51. https://doi.org/10.1109/I3CAT53310.2021.9629409

6. Tallboys, J., Zhu, Y., Rajasegarar, S. (2022). Identification of Stock Market Manipulation with Deep Learning. *Proceedings of 17th International Conference on Advanced Data Mining and Applications (ADMA 2022).* Part I. Edited by B. Li, L. Yue, J. Jiang, W. Chen, X. Li, G. Long, F. Fang, H. Yu. Springer Cham, 408–420.. https://doi.org/10.1007/978-3-030-95405-5_29

7. Uslu, N.C., Akal, F. (2022). A Machine Learning Approach to Detection of Trade-Based Manipulations in Borsa Istanbul. *Computational Economics,* Vol. 60, 25–45. https://doi.org/10.1007/s10614-021-10131-8

8. Ueda, K., Suwa, H., Yamada, M., Ogawa, Y., Umehara, E., Yamashita, T., Tsubouchi, K., Yasumoto, K. (2024). SSCDV: Social media document embedding with sentiment and topics for financial market forecasting. *Expert Systems with Applications,* Vol. 245, 122988. https://doi.org/10.1016/j.eswa.2023.122988

9. Li, Y., Bu, H., Li J., Wu, J. (2020). The role of text-extracted investor sentiment in Chinese stock price prediction with the enhancement of deep learning. *International Journal of Forecasting,* Vol. 36, Issue 4, 1541–1562. https://doi.org/10.1016/j.ijforecast.2020.05.001

10. Lin, P., Ma, S., Fildes, R. (2024). The extra value of online investor sentiment measures on forecasting stock return volatility: A large-scale longitudinal evaluation based on Chinese stock market. *Expert Systems with Applications,* Vol. 238, Part B, 121927. https://doi.org/10.1016/j.eswa.2023.121927

11. Mili, M., Sahut, J., Teulon, F., Hikkerova, L. (2024). A multidimensional Bayesian model to test the impact of investor sentiment on equity premium. *Annals of Operations Research,* Vol. 334, 919–939. https://doi.org/10.1007/s10479-023-05165-0

12. Dixon, M.F., Halperin, I., Bilokon, P. (2020). *Machine Learning in Finance: From Theory to Practice.* Springer, 548 p. https://doi.org/10.1007/978-3-030-41068-1

13. Cai, Y., Tang, Z., Chen, Y. (2024). Can real-time investor sentiment help predict the high frequency stock returns? *North American Journal of Economics and Finance,* Vol. 72, 102147. https://doi.org/10.1016/j.najef.2024.102147

14. Gupta, R., Nel, J., Pierdzioch, C. (2023). Investor confidence and forecastability of US stock market realized volatility: Evidence from machine learning. *Journal of Behavioral Finance,* Vol. 24, Issue 1, 111–122. https://doi.org/10.1080/15427560.2021.1949719

15. Chen, H., De, P., Hu, Y.J., Hwang, B.H. (2013). Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies,* Vol. 27, Issue 5, 1367–1403. https://doi.org/10.1093/rfs/hhaa114

16. Fama, E.F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance,* Vol. 25, No. 2, 383–417. https://doi.org/10.2307/2325486

17. Samuelson, P.A. (1965). Proof that Properly Anticipated Prices Fluctuate Randomly. *Industrial Management Review,* Vol. 6, No. 2, 41–49. Available at: https://capital-gain.ru/assets/docs/Proof-that-properly-anticipated-prices-fluctuate-randomly.pdf

18. Shiller, R.J. (1981). Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? *The American Economic Review,* Vol. 71, No. 3, 421–436. Available at: https://www.aeaweb.org/aer/top20/71.3.421-436.pdf

19. Baker, M., Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance,* Vol. 61, Issue 4, 1645–1680. https://doi.org/10.1111/j.1540-6261.2006.00885.x

20. Baker, M., Wurgler, J. (2007). Investor Sentiment in the Stock Market. *Journal of Economic Perspectives,* Vol. 21, No. 2, 129–151. https://doi.org/10.1257/jep.21.2.129

21. Niu, H., Pan, Q., Xu, K. (2023). Hybrid deep learning models with multi-classification investor sentiment to forecast the prices of China's leading stocks. *PLoS ONE,* Vol. 18, Issue 11, e0294460. https://doi.org/10.1371/journal.pone.0294460

22. Liu, Q., Lee, W.S., Huang, M., Wu, Q. (2023). Synergy between stock prices and investor sentiment in social media. *Borsa Istanbul Review,* Vol. 23, Issue 1, 76–92. https://doi.org/10.1016/j.bir.2022.09.006

23. Cookson, J.A., Lu, R., Mullins, W., Niessner, M. (2024). The social signal. *Journal of Financial Economics,* Vol. 158, 103870. https://doi.org/10.1016/j.jfineco.2024.103870

24. Islam, S.R., Ghafoor, S.K., Eberle, W. (2018). Mining illegal insider trading of stocks: A proactive approach. *Proceedings 2018 IEEE International Conference on Big Data.* IEEE, 1397–1406. https://doi.org/10.1109/BigData.2018.8622303

25. Liu, Q., Wang, C., Zhang, P., Zheng, K. (2021). Detecting stock market manipulation via machine learning. *International Review of Financial Analysis,* Vol. 78, 101887. https://doi.org/10.1016/j.irfa.2021.101887

26. Chong, E., Han, C., Park, F.C. (2017). Deep learning networks for stock market analysis and prediction. *Expert Systems with Applications,* Vol. 83, 187–205. https://doi.org/10.1016/j.eswa.2017.04.030

27. Li, A., Wu, J., Liu, Z. (2017). Market manipulation detection based on classification methods. *Procedia Computer Science,* Vol. 122, 788–795. https://doi.org/10.1016/j.procs.2017.11.438

28. Nam, D., Skillicorn, D.B. (2023). Detecting pump & dump stock market manipulation from online forums. *arXiv,* 2301.11403. https://doi.org/10.48550/arXiv.2301.11403

29. Domashova, J., Yakimov, D., Bredikhin, D., Gorbunov, K., Slavik, R., Kadyrov, I. (2022). Detection and analysis of atypical stock transactions with possible misuse of insider information and market manipulation: methods and models. *Procedia Computer Science,* Vol. 213, 165–174. https://doi.org/10.1016/j.procs.2022.11.052

30. Münster, J., Reichenbach, M., Walther, T. (2024). Robinhood, Reddit, and the news: The impact of traditional and social media on retail investor trading. *Journal of Financial Markets,* Vol. 71. 100929. https://doi.org/10.1016/j.finmar.2024.100929

31. Warkulat, M., Pelster, M. (2024). Social media attention and retail investor behavior: Evidence from r/wallstreetbets. *International Review of Financial Analysis,* Vol. 96, Part B, 103721. https://doi.org/10.1016/j.irfa.2024.103721

32. Keasey, K., Lambrinoudakis, C., Mascia, D.V., Zhang, Z. (2025). The impact of social media influencers on the financial market performance of firms. *European Financial Management,* Vol. 31, Issue 2. 745–785. https://doi.org/10.1111/eufm.12513

33. Semenova, M., Winkler, A. (2021). Reddit's self-organized bull runs: Social contagion, sentiment, and risk. *arXiv,* 2104.01847. https://doi.org/10.48550/arXiv.2104.01847

34. Xu, L., Xue, C., Zhang, J. (2024) The impact of investor sentiment on stock liquidity of listed companies in China. *Investment Management and Financial Innovations,* Vol. 21, Issue 2, 1–14. https://doi.org/10.21511/imfi.21(2).2024.01

35. Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., Mozetič, I. (2015). The effects of Twitter sentiment on stock price returns. *PLOS ONE,* Vol. 10, Issue 9, e0138441. https://doi.org/10.1371/journal.pone.0138441

INFORMATION ABOUT AUTHOR

Gleb Andreevich Khaziev

Post-Graduate Student, Basic Department of Financial Markets Infrastructure, Faculty of Economic, Junior Research Fellow at the Center for Financial Research and Data Analysis, National Research University Higher School of Economics, Moscow, Russia (109028, Moscow, Pokrovsky Boulevard, 11); ORCID https://orcid.org/0000-0003-3346-0006 e-mail: glebhaziev@mail.ru

FOR CITATION

Khaziev, G.A. (2025). Detecting Social Stock Pumping Using Machine Learning: Empirical Evidence from the Russian Market. *Journal of Applied Economic Research,* Vol. 24, No. 4, 1445–1474. https://doi.org/10.15826/vestnik.2025.24.4.047

ARTICLE INFO

Received June 9, 2025; Revised July 30, 2025; Accepted August 13, 2025.