

КВАНТИФИКАЦИЯ ИНФОРМАЦИОННОЙ СРЕДЫ КАК ИНСТРУМЕНТ ИНВЕСТИЦИОННОГО АНАЛИЗА

© 2021 **Родионов Дмитрий Григорьевич**

доктор экономических наук, профессор, Высшая инженерно-экономическая школа
Санкт-Петербургский политехнический университет Петра Великого (СПбПУ),
Россия, Санкт-Петербург
E-mail: dmitry.rodionov@spbstu.ru

© 2021 **Карпенко Павел Алексеевич**

Высшая инженерно-экономическая школа
Санкт-Петербургский политехнический университет Петра Великого (СПбПУ),
Россия, Санкт-Петербург
E-mail: karpenko_pavel@mail.ru

© 2021 **Пашинина Полина Александровна**

кандидат экономических наук, доцент, Высшая инженерно-экономическая школа
Санкт-Петербургский политехнический университет Петра Великого (СПбПУ),
Россия, Санкт-Петербург
E-mail: pashininapolina@yandex.ru

В статье исследуется взаимосвязь между поведением инвесторов в информационной среде и ценами на акции различных компаний. С помощью дисперсионного анализа и t-статистики определено, что цены на акции разных компаний изменяются неодинаково и количество запросов в большей степени зависит от времени, т.е. от событий, произошедших в определённом времени. В статье проводится обзор исследований на данную тему, анализируются существующие методики влияния настроений инвесторов, оцениваемых с помощью методов машинного обучения, выраженных в социальных сетях, новостных форумах и тематических сайтах, на котировки акций. Научная новизна заключается в одновременном рассмотрении данного вопроса с двух сторон: оценка настроений инвесторов с использованием и тональности информационной среды, и содержательной компоненты. В результате была составлена и обоснована концептуальная модель, характеризующая две компоненты, обрабатываемые и анализируемые с помощью методов машинного обучения с использованием социальных сетей и влияющие на цены акций. Подчеркивается, что в основе концептуальной модели лежит рассмотрение влияния как положительных, отрицательных и нейтральных компонент информационного сообщения, так и упоминаний отдельных слов и их частоты на цены акций, которые, в свою очередь, обеспечивают инвестиционную привлекательность компании/актива.

Ключевые слова: информационная среда, естественная информация, квантификация естественной информации, инвестиционный анализ.

В настоящее время количество российских инвесторов возрастёт в разы. Причина такого предположения заключается в том, что вследствие экономического спада в 2020 году из-за пандемии коронавируса Центральный Банк снижал ключевую ставку, что означает и снижение процентов по банковским депозитам (в 2020 году). Также, с 1 января 2021 года введён налог на вклады, что существенно повлияло на желание инвесторов перейти на фондовый рынок, перемещая существенную долю капитала

из банков на фондовый рынок, на котором, ко всему прочему, есть своё преимущество — пользование налоговыми льготами с помощью индивидуального инвестиционного счёта. Так, важно отметить, что с 2020 года появилось более 4,2 млн. российских инвесторов (были открыты брокерские счета). Данная тенденция обуславливает повышенный интерес инвесторов к вложениям в наиболее прибыльные и наименее рискованные активы, как в популярные и востребованные способы получения заработка.

Именно поэтому возросла необходимость разработки новых инструментов для определения доходности и рискованности инвестирования в активы, а также для прогнозирования изменения основных показателей на фондовых рынках.

Предлагается также обратить внимание на количество запросов в Яндексе «Доходность акций» и похожих запросов за сентябрь 2021 года с помощью сайта wordstat.yandex.ru (Рисунок 1) [1].

Так, согласно данным на рисунке 1, количество запросов «доходность акций» в Яндексе за сентябрь 2021 года составило 22418 показов, что на 51,74% больше, чем количество этих запросов в ноябре 2019 года (12334 показа).

Таким образом, интерес к данной теме возрастает, что обуславливает необходимость рассмотрения различных факторов, влияющих на цены акций компаний. В данной статье рассма-

триваются возможные влияния социальных сетей и анализ определённых слов, высказываний людей в них, влияющих на цены, а, следовательно, и доходность акций. Сейчас нельзя отрицать, что Интернет не оказывает влияния на фондовые рынки. Так, рассмотрим пример того, как изменяется цена акций Газпром и количество запросов «Газпром» с помощью Wordstat.ru в разрезе недель [1, 2]. На рисунке 2 представлен график изменения цен закрытий акций «Газпром» на конец недели за период 23.05.2021–22.10.2021.

Так, можно заметить, что цены на акции «Газпром» меняются еженедельно, что может быть обусловлено рядом факторов, и наличие динамики является основанием для определения этих факторов и направленности, и степени их влияния, а в рамках данного исследования оценивается эффект поведения инвесторов на цены акций.

Что искали со словом «доходность акций» — 22 418 показов в месяц		Запросы, похожие на «доходность акций»	
Статистика по словам	Показов в месяц	Статистика по словам	Показов в месяц
доходность акций	22 418	самое выгодное вложение денег на сегодняшний день	8 485
дивидендная доходность акций	4 212	куда вложить деньги чтобы получать ежемесячный доход	1 012
доходность акций компании	2 402	инвестиционный доход	12 964
цена акций доходность акций	1 758	доход.ру дивидендный календарь	685
текущая доходность акции	1 718	чистый дисконтированный доход	3 748
доходность российских акций	1 509	куда можно вложить деньги и получить прибыль	1 696
доходность акций 2021	1 326	накопленный купонный доход	1 216
доходность дивидендов +по акциям	1 303	чистый дисконтированный доход это	547
годовая доходность акций	1 149	вложение денег с целью получения дохода	519
ожидаемая доходность акции	1 137	куда вложить деньги чтобы получить прибыль	514
рыночная доходность акций	1 133		

Рис. 1. Количество поисковых запросов «Доходность акций» и похожих запросов в России за сентябрь 2021 года.

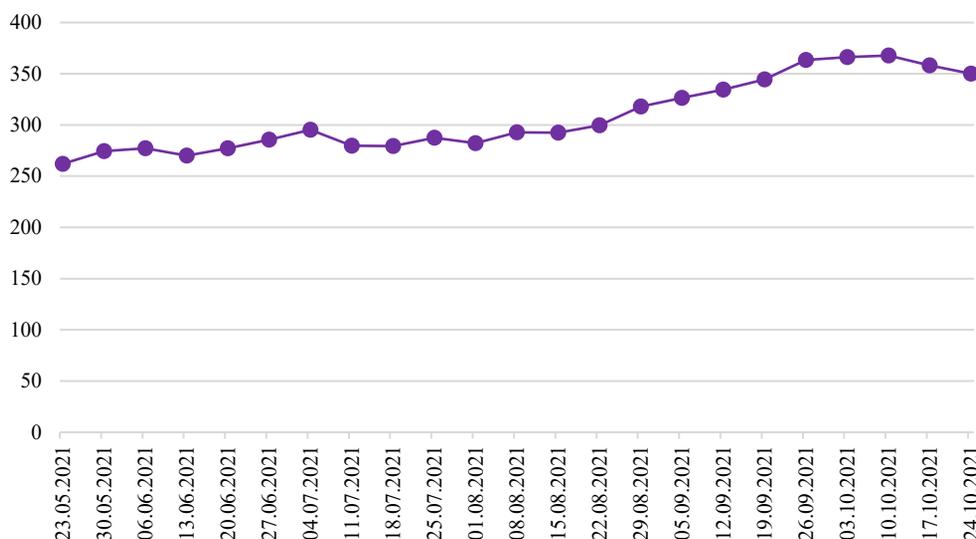


Рис. 2. График изменения акций на цены «Газпром» за 23.05.2021–22.10.2021

В предположении, что информационная среда неоднозначно влияет на цены разных компаний, проведём однофакторный дисперсионный анализ для определения, существует ли разница в распределении цен акций разных компаний. Для примера рассмотрим цены акций «Газпром» и «Магнит». Предположим, что нулевая гипотеза — отсутствие разницы между изменениями цен на акции «Газпром» «Магнит», и, соответственно, альтернативная — наличие разницы. Рассчитаем *t*-критерий Стьюдента и приведём результаты в Таблице 1.

Расчётное значение *t*-критерия больше табличного, что говорит о том, что существует разница между изменениями цен на акции «Газпром» и «Магнит». Так, нулевая гипотеза отвергается, а альтернативная подтверждается.

Немаловажно также рассмотреть однофакторный дисперсионный анализ по показателю цена акции. Формируем нулевую гипотезу: принадлежность акции конкретной компании не оказывает никакого влияния на цену акции. Результаты дисперсионного анализа представлены в таблице 2.

Дисперсионный анализ позволяет вычислить ту часть дисперсии, которая обусловлена дифференциацией конкретного фактора, а критерий Фишера позволит оценить статистическую значимость этого фактора. Таким образом, расчётное значение *F*-критерия больше табличного (при этом, *p*-значимость крайне мала), что означает, что цена акций разных компаний изменяется абсолютно по-разному. Иными словами, принадлежность акции той или иной компании-эмитенту данной акции играет важную роль для данного актива акций. При этом распределение цен во времени оказывает менее сильное влияние цен на акции в сравнении с принадлежностью

акции определённой компании.

Подобный анализ для количества запросов «Газпром акции» и в зависимости от компании, а именно — её популярности на данный момент, обсуждаемости, что может обуславливать её успех и активные действия на рынке, а так же от некоторой информации в интернете (например, публикация какого-либо анализа цен на акции данной компании или распространение ложной информации), количество запросов может быть разным (исходя из сложившихся условий в информационных каналах). Дисперсионный анализ показал, что при этом распределение запросов во времени оказывает более сильное влияние на количество запросов про акцию чем то, какая компания-эмитент используется в запросе. Это говорит о том, что количество запросов про акции (с названием компании-эмитента) больше зависит от такого, какая ситуация сложилась на рынке в данный период времени.

Так, было определено, что цены на акции разных компаний изменяются неодинаково и количество запросов акций компаний изменяется так же в разных направлениях, однако в большей степени зависит от времени, т.е. от событий, произошедших в определённом времени.

Кроме того, необходимо рассмотреть поведение инвесторов на рынке. Поведение инвестора на рынке показывает, как в той или иной ситуации он себя поведет: в страхе начнет продавать дешевающие активы, переждет падение или докупит падающие акции. Однако, падение или повышение цен на активы не является единственной «лакмусовой бумагой», определяющей действия инвесторов. Также существенное влияние на их поведение оказывают информационные потоки, проявляющиеся в виде СМИ, общения с другими трейдерами, слив инсайдерской

Таблица 1. Расчёт *t*-критерия Стьюдента

Показатель	Расчётное значение	Табличное значение
<i>t</i> -критерий (акции «Газпром» и «Магнит»)	62,65	2,0739

Таблица 2. Результаты дисперсионного анализа (цена акции)

Источник вариации	Дисперсия	<i>F</i> расчётное	<i>F</i> табличное
Цена акции	325599758,7	4103,98	4,06
Время (неделя)	3490851,34		
Итого	329090610		

информации и т.д. Для этого важно понимать, какие именно «новости» подталкивают инвесторов на те или иные действия на фондовом рынке, что и является основой данной научно-исследовательской работы. Поведение инвесторов и существующие методики их анализа представлены в литературном обзоре.

Таким образом, согласно проведённым анализам, можно предположить, что цены компаний изменяются в разных направлениях за определённый промежуток времени, и эти изменения обусловлены рядом факторов. Однако, в настоящее время развитие информационных каналов и рост числа непрофессиональных инвесторов обуславливают необходимость исследования влияния поведения инвесторов, выраженного в количественных характеристиках, на цены акций. Иными словами, внезапное увеличение количества людей, которые хотят инвестировать в акции, но ввиду отсутствия опыта и достаточных знаний в этой области не могут делать это самостоятельно и профессионально, и появление множества сайтов и социальных сетей, на которых как профессиональные, так и непрофессиональные инвесторы могут делиться своим мнением, а новостные источники публиковать как достоверную, так и недостоверную информацию, обеспечивает необходимость изучения влияния поведения инвесторов на котировки акций для того, чтобы оценить, является ли то или иное движение цены акции обоснованным или подверженным сфальсифицированным мнениям или новостям. Так, немаловажным элементом, влияющим на движение цен на акции, являются инвесторы, которые, обладая уникальными личностями и психологическими характеристиками, склонны поступать тем или образом при любом изменении на фондовом рынке. Исходя из последних событий инвесторы всё больше заинтересованы в альтернативном способе вложения денег, для чего необходимо уметь анализировать тенденции фондового рынка и прогнозировать наиболее вероятные изменения, поэтому рынок переполнен непрофессиональными инвесторами, которые действуют нерационально, именно поэтому особенно важно понимать психологию инвесторов и предполагать их действия в ответ на изменения на рынке. Эмоции и особенности темперамента человека могут принести большие финансовые потери и изменения на рынке, что обуславливает необходимость разработки инструментов, способных с некоторой вероятностью предсказать

изменения котировок акций (Севумян Э.Н.) [3]. Кроме того, рациональные инвесторы сдерживают искусственно вызванные движения цен путём следования субъективным оценкам, не поддающихся влиянию интернет-сайтов, однако обсуждение события в социальных сетях, провоцирующее непрофессиональных инвесторов поступать тем или иным образом, способствует сильным движениям доходностей акций, которые не могут быть скорректированы поведением рациональных инвесторов (Сюн Сюн, Юнцян Мэн, Натан Лаэль Джозеф, Дэхуа Шенгуан) [4].

Несмотря на это большинство инвесторов анализируют информацию о рынке и компании, прежде чем покупать акции. Информация о рынке доступна в социальных сетях, а также в новостях, блогах и отзывах клиентов компаний. Инвесторы просматривают профили компаний, исторические торговые данные, новости, мнения аналитиков, финансовые отчеты и сообщения, чтобы определить, стоимость акций какой компании увеличилась или уменьшилась. Ручной анализ этой информации подвержен ошибкам, поэтому применяются методы машинного обучения, которые используются для поиска связей между параметрами (настроение инвесторов и цены на акции) и прогнозирования будущих значений зависимых переменных (цен на акции). В работах Мажар Джавед Аван, Мохд Шафри Мохд Рахим, Хайтам Нобани и других с помощью различных методов машинного обучения подтверждается наличие существенной и значимой связи между котировками и акций и тем, какие данные публикуются в информационных источниках, а также обосновывается необходимость использования методов машинного обучения для анализа зависимостей и прогнозирования цен на акции [5]. Авторы Пейран Жоао, Андре Вейга, Ангар Вальтер, Юлин Ли, Хуэй Бу, Цзяхун Лик, Цзюньцзе Ву, Ди Ван обращаются к регрессионному анализу и моделям, где независимая переменная — настроение инвесторов, зависимая — цена на акцию [6, 7, 8]. Регрессионный анализ многих исследований показал, что высокий коэффициент детерминации обеспечивает большой процент объяснения изменения цен под влиянием информационной среды, представленной настроениями инвесторов (сообщениями, новостями). Важно отметить, что упоминание названий компаний, использование определённых слов (токенов) и выражений также несут в себе определённый посыл, который непременно найдёт отражение в настроениях

инвесторов. Иными словами, количество упоминаний каких-либо компаний, ценных бумаг, событий на фондовом рынке также связано с движением цен на акции, которые тем или иным образом связаны с компанией, событием и т.д. из новости. Так, авторами Пейран Жоао, Андре Вейга, Ангар Вальтер в статье «Social media, news media and the stock market» было определено, что высокая обсуждаемость в социальных сетях предсказывает более высокую будущую волатильность, и эта взаимосвязь является статистически значимой. Важным замечанием является то, что обсуждаемость в социальных сетях вызывает большую волатильность, чем обсуждаемость в средствах массовой информации [6]. Однако, не только упоминания конкретных компаний влияют на будущие котировки их акций, но и упоминания связанных компаний, событий или акционеров, и работы с использованием сопутствующих слов имеют более точные результаты в определении влияния настроений инвесторов на цены акций (Бин Ли, Кит Си Си Чан, Андрес Гарсия-Мединай, Хосе Антонио Нуньес Мора) [9, 10].

Елин Ли, Хуэй Бу, Цзяхун Лик, Цзюньцзе Ву углубляются в изучение влияния комментариев инвесторов на котировки акций, используя модели долгой краткосрочной памяти (LSTM), метода опорных векторов и наивной модели Байеса, и приходят к выводам о том, что ежедневные настроения инвесторов содержат прогнозирующую информацию только для цен открытия, в то время как почасовые настроения имеют наиболее точный прогноз для цен закрытия. Кроме того, существует и идея о том, что чем больше новостей публикуется, тем активнее ведётся покупка/продажа активов на фондовом рынке, однако, кроме того, отмечается, что покупка ведётся активнее [7]. Исследования, разделяющие настроения инвесторов на позитивные и негативные, пришли к выводу о том, что пессимистические настроения инвесторов влияют на котировки сильнее, чем оптимистические, и также фирмы с оптимистическими настроениями инвесторов имеют значительно более высокую доходность акций в текущем месяце, чем фирмы с пессимистическими настроениями (Хао Фан, Цзянь-Пин Чун, Ян-Чен Лу, Йен-Сянь Ли) [11]. Кроме того, существуют и работы, подтверждающие то, что связь между публикациями на интернет-сайтах общения инвесторов и средней еженедельной доходностью акций в основном относится к периодам, когда общение инвесторов было неактивным, а более

высокая доходность акций не зависит от настроений инвесторов (Сюн Сюн, Юнцян Мэн, Натан Лаэль Джозеф, Дэхуа Шенгуан) [4].

Важный вопрос — кто именно оказывает существенное и значимое влияние на поведение инвесторов и, следовательно, котировки акций. Так, некоторые инвесторы полагают, что высказывания лидеров мнений оказывают сильное влияние на массы, являясь причиной начала какого-либо тренда на фондовом рынке, что обеспечивает составление факторов, определяющих силу сообщения: (1) факторы получателя сообщения (черты личности получателя, которые характеризуют степень его подверженности влиянию социальных медиа); (2) свойства технологии (к технологическим различиям социальных медиа) (А.В.Дубко, К.Буассье) [12]. В продолжении вопроса о том, кто или что оказывает более существенное влияние на поведение инвесторов, котировки акций, существует подтверждённое с помощью методов машинного обучения предположение о том, что публикации от пользователей с большим количеством подписчиков оказывают большее влияние на доходность в тот же день, в то время как публикации от пользователей с небольшим количеством подписчиков оказывают большее влияние на будущие доходы (Хонки Сул, Алан Р.Деннис, Линьяо Айви Юань) [13]. Авторы других работ придерживаются мнения, что необязательно давать людям ложную информацию, чтобы побудить их выполнить то или иное действие или изучить авторитетность источников, количество их подписчиков, так как для этого также можно просто найти сообщников, которые осознанно поступят так, как объективно действовать не стоит. Данная идея иллюстрируется на примере, когда в Telegram канале была выложена новость о том, что сейчас рекомендуется приобрести акции «Распадской угольной компании», после чего цена акций компании стала расти и в итоге достигла самых больших значений за последние 8 лет, хотя обоснованных причин на приобретение акций данной компании не было, участникам чата была дана ранее известная положительная информация о данной компании. Такие искусственные вмешательства на рынок акций чаще всего происходят из-за насыщенности в 2020–2021 годах фондового рынка непрофессиональными инвесторами, которые в силу своей неопытности не могут отличить «адекватную» информацию от той, которая является своего рода манипуляцией (Н.А.Звягинцева, К.А.Ов-

чинникова) [14].

Важным замечанием является то, что существующие методики анализа влияния информационной среды (применяемые в вышеописанных работах), в частности — настроений инвесторов в социальных сетях на цены акций в большинстве случаев являются похожими, однако недавно опубликованные исследования (в 2020 и 2021 году) имеют усовершенствованные инструменты в связи с ростом интереса к данной теме.

Во многих работах так или иначе проводился сбор первичных данных. Так, анализировались сайты инвесторов, новостные ресурсы и социальные сети (Twitter). Комментарии, «твиты» (термин сервиса микроблоггинга Twitter, обозначающий запись, публикуемую пользователем) и новости — для целей автоматизированного парсинга данной информации часто применяется язык программирования Python 3 и ряд библиотек. Для анализа данных, а чаще всего — о прогнозировании цен на акции, используется машинное обучение (ML) — это область компьютерных наук, с помощью которой компьютерные системы могут придавать смысл

данным точно так же, как это делают люди. Проще говоря, ML — это тип искусственного интеллекта, который извлекает шаблоны из необработанных данных, используя алгоритм или метод. Наиболее часто для создания программ ML применяются языки R, Python, Scala и Julia. [15]. В работах авторов, упомянутых выше, нередко использовались такие методы ML как дерево решений (этот алгоритм классифицирует объекты с помощью вопросов, касающихся их атрибутов, и ответов на них, двигаясь по «ветвям» из каждого разветвления, пока не будет дан ответ на последний вопрос, «случайный лес» (составляет множество деревьев решений («комитет») с разным набором атрибутов, формируются разные классы с результатами каждого дерева и выбирается наиболее вероятный класс (мудрость толпы)), и кластеризация (это группирование элементов данных, имеющих сходные характеристики, с помощью статистических алгоритмов).

На рисунке 3 представлена одна из возможных моделей сбора и анализа данных социальных сетей для прогнозирования рынка в большинстве исследований, оценивающих влияние настроений инвесторов на цены акций.



Рис. 3. Модель сбора и анализа данных социальных сетей для прогнозирования рынка

Глубокое обучение — это метод ML, который позволяет обучать нейросеть прогнозировать результаты по входному датафрейму (структура, в которой данные хранятся в табличной форме). Продвинутое моделирование глубокого обучения, такие как Long Short Term Memory Networks (LSTM), способны улавливать закономерности в данных временных рядов, следовательно, возможно их применение для прогнозирования будущего тренда данных [16]. В последних исследованиях вводится модель глубокого обучения, которая называется долгой кратковременной памятью (LSTM), для построения прогнозов доходности акций.

Практически во всех проанализированных исследованиях для анализа настроений инвесторов используется наивный байесовский алгоритм — алгоритм классификации, основанный на теореме Байеса с допущением о независимости признаков. Сообщения, новости, твиты подразделяются на три категории: негативные (о снижении акции), позитивные (о повышении) и нейтральные (отсутствие изменений). Отрицательные и позитивные твиты применяются для построения индекса настроений в какой-либо промежуток времени, и далее рассчитывается корреляция индекса и изменения цен на акции. В некоторых работах данный алгоритм применялся, чтобы классифицировать сообщения на три другие категории: покупать, держать и продавать.

При этом помимо прямого выделения в сообщениях выражений об ожидании снижения или роста цены акции необходимо выделить их эмоциональную составляющую. Так, тональность сообщений исследуется с помощью интеллектуального анализа текста. Однако, в некоторых исследованиях либо тональность не учитывается, либо анализируется специальными финансовыми аналитиками (рассматривается только содержательная компонента, без эмоциональной). Для анализа тональности авторы некоторых научных статей использовали OpinionFinder (система, которая обрабатывает документы, автоматически идентифицирует субъективные предложения и измеряет позитивное и негативное настроение) и Google Profile of Mood States (алгоритм, который измеряет настроение и классифицирует его по 6 параметрам: спокойствие, тревожность, уверенность, энергичность, вид и счастье).

Таким образом, большинство работ, рас-

сматривающих взаимосвязь содержания текстовых сообщений инвесторов и движения цен на акции, содержат похожие инструменты для оценки этих параметров и прогнозирования будущих трендов. Также немаловажно отметить, что все способы оценки влияния настроений инвесторов на цены активов тем или иным образом сводятся к сбору данных по динамике цен, текстов сообщений/публикаций/твитов и анализу их эмоциональной и содержательной компоненты. Именно параметры цен и тональность или частота упоминание слов и их корреляция с ценами являются основными компонентами всех исследований, в том числе и данной работы. Так, автор исследования предлагает рассмотреть эти компоненты подробнее.

Цены на акции. Цена акции — цена, по которой могут заключаться сделки с акциями. На фондовом рынке у акции имеются только две оценки, являющиеся предметом соглашения:

- номинальная стоимость акции — это оценка акции, как предмет эмиссионного соглашения между эмитентом и инвестором.
- рыночная цена акции — это оценка акции, как предмет инвестиционного соглашения между торговцами на рынке [17]. Это текущая цена, обозначенная продавцом или покупателем, по которой участники торгов согласны совершить сделку с ценной бумагой. Так, инвесторы, способные идти на риск ради получения дохода, увлечены не столько номинальной стоимостью акций, сколько рыночной. Анализируя изменения рыночных цен ценных бумаг, инвесторы могут занимать длинные и короткие позиции на рынке и получать из этого прибыль [18].

Тональность текста. Тональность — текстовая категория, в которой находит отражение эмоционально-волевая установка автора текста при достижении конкретной коммуникативной цели, психологическая позиция автора по отношению к излагаемому тексту, а также к адресату и ситуации общения [19].

Анализ тональности — это область компьютерной лингвистики, которая занимается изучением мнений и эмоций в текстовых документах.

Цель анализа тональности — поиск эмоционально-окрашенных слов/словосочетаний в тексте и соотнесение их с теми или иными классами тональности.

В анализе текстов внимание может уделяться двум основным аспектам:

1. Ключевая тема в тексте

2. Тональность — отношение автора к обсуждаемой теме

Число классов тональности конкретного текста определяется самим исследователем, заинтересованным в анализе тональности. Часто рассматривается задача бинарной классификации тональности, т.е. классов всего два: «положительный» и «отрицательный».

Существуют и другие варианты для обучения классификатора:

- Плоская классификация — обучаем один классификатор для всех классов.
- Иерархическая классификация — классы делятся на группы и происходит обучение нескольких классификаторов для определения групп. Например, существует 5 классов («сильно положительный», «средне положительный», «нейтральный», «средне отрицательный», «сильно отрицательный»). Для начала необходимо обучить бинарный классификатор.
- Регрессия — обучаем классификатор для получения численного значения тональности, например от 1 до 10, где большее значение означает более положительную тональность [20].

Содержательная компонента. Как выше было указано, в тексте сообщения мы можем анализировать эмоцию автора или тему самого текста. Так, разбиение текста на отдельные слова и дальнейшая работа с ними также может помочь в определении взаимосвязи настроения инвесторов, выраженного в использовании тех или иных слов, и цен акций. Например, прямое упоминание в сообщении слова «упадёт» в чате общения инвесторов может обозначить снижение цен на акции, так как некоторые непрофессиональные инвесторы, доверяя источнику информации и не обладая навыками анализа событий на фондовом рынке, могут продать акции, что действительно может способствовать снижению цен на акции. Однако, перед тем как анализировать текст, необходимо выполнить его предварительную обработку. Необходимо выделить следующие шаги для этого этапа:

- Токенизация — преобразование предложений в слова;
- Удаление ненужных знаков препинания, тегов;
- Удаление стоп-слов — часто встречающиеся слова, которые не имеют определенной се-

мантики;

- Основа — слова сводятся к корню путем удаления перегиба путем удаления ненужных символов, обычно суффикса;

- Лемматизация — еще один подход к устранению перегиба путем определения части речи и использования подробной базы данных языка. Для дальнейшей работы подходят классические ML-подходы («Наивный Байес» или «Машины опорных векторов») [21].

После проведения вышеуказанных шагов текст будет пригоден для перевода его в числовой формат, чтобы продолжать выделение признаков (квантификации полученного списка слов). Далее возможно составление различных индексов, расчёт частоты наличия определённых слов и их корреляция с движением цен на акции.

Концептуальная модель. Исходя из вышеописанных параметров, концептуальная модель обозначенной темы может выглядеть образом, показанным на рисунке 4.

Таким образом, проведены анализы возможностей исследования акций различных эмитентов во времени по компаниям, имеющие в итоге идею о различной динамике доходностей акций, которые поддаются влиянию информационных потоков. Литературный обзор существующих исследований на тему изучения влияния настроений инвесторов на цены акций подтвердил наличие взаимосвязи и воздействия сообщений профессиональных и непрофессиональных участников фондового рынка на котировки акций. Существующие методики основываются на методах машинного обучения, позволяющих существенно упростить обработку информации, и именно поэтому за основу изучения влияния в концептуальной модели берется квантификация информационной среды, а именно числовая оценка эмоциональной составляющей участников среды и отдельных токенов (содержательная компонента). Так, в основе концептуальной модели лежит рассмотрение влияния как положительных, отрицательных и нейтральных компонент информационного сообщения, так и упоминаний отдельных слов и их частоты на цены акций, которые, в свою очередь, обеспечивают инвестиционную привлекательность компании/актива.



Рис. 4. Концептуальная модель, описывающая влияние настроения инвесторов на цены акций

Библиографический список

1. Подбор слов. — Текст: электронный // wordstat: [сайт]. — URL: <https://wordstat.yandex.ru/> (дата обращения: 15.10.2021).
2. Итоги торгов. — Текст: электронный // Московская биржа: [сайт]. — URL: <https://www.moex.com/> (дата обращения: 15.10.2021) Подбор слов. — Текст: электронный // wordstat: [сайт]. — URL: <https://wordstat.yandex.ru/> (дата обращения: 15.10.2021).
3. Севумян, Э.Н. ПРОГНОЗИРОВАНИЕ ДИНАМИКИ ЦЕННЫХ БУМАГ В УСЛОВИЯХ ПОВЫШЕННОГО ИНТЕРЕСА ИНВЕСТИТОРОВ К БИРЖЕВОЙ ТОРГОВЛЕ Текст научной статьи по специальности «Экономика и бизнес» / Э.Н. Севумян. — Текст: электронный // ФРиБ. — 2021. — № 6. — С. 68–70.
4. Stock mispricing, hard-to-value stocks and the influence of internet stock message boards / Xiong Xiong, Meng Yongqiang, L.J.Nathan, Shen Dehua. — Текст: электронный// International Review of Financial Analysis. — 2020. — № 72.
5. Mazhar, J.A. Social Media and Stock Market Prediction: A Big Data Approach / J.A.Mazhar, Shafry, Mohd Mohd, Nobanee, и Haitham. — Текст: электронный // — 2021.
6. Peiran, Jiao social media, news media and the stock market / Jiao Peiran, Veiga André, Walther Ansgar. — Текст: электронный // Journal of Economic Behavior & Organization. — 2020. — № 176. — С. 63–90.
7. The role of text-extracted investor sentiment in Chinese stock price prediction with the enhancement of deep learning / Li Yelin, Bu Hui, Lic Jiahong, Wu Junjie. — Текст: электронный // International Journal of Forecasting. — 2020. — № 26. — С. 1541–1562.
8. Exploiting investors social network for stock prediction in China's market / Xi Zhang Jiawei Shi, Di Wang, Binxing Fang. — Текст: электронный // International Review of Financial Analysis. — 2021. — № 77.

9. *Bing, Li* Discovering public sentiment in social media for predicting stock movement of publicly listed companies / Li Bing, C, C Keith, Ou, и, Sun Carol. Текст: электронный // Information Systems. — 2017. — № 69. — С. 81–92.
10. *Roma'n, Alejandro, Mendoza* Measuring information flux between social media and stock prices with Transfer Entropy / Alejandro, Mendoza Roma'n, Garcí'a-Medina I Andre's, A. N. Jose'. — Текст: электронный // Modeling in Finance and Econometrics from the paradigm of Econophysics. — 2021.
11. The impacts of investors' sentiments on stock returns using fintech approaches / Fang Hao, Chung Chien-Ping, Lu Yang-Cheng [и др.]. — Текст: электронный // Journal of Computational Science. — 2018. — № 28. — С. 294–303.
12. Дубко А.В. ВЛИЯНИЕ СОЦИАЛЬНЫХ МЕДИА НА ВОЛАТИЛЬНОСТЬ ЦЕН НА АКЦИИ: СОСТОЯНИЕ ПРОБЛЕМЫ / А. В. Дубко, К. Буассье. — Текст: электронный.
13. *Hongkee, Sul* Trading on Twitter: The Financial Information Content of Emotion in social media/ Sul Hongkee, R. D. Alan, I. Y. Lingyao. — Текст: электронный // 47th Hawaii International Conference on System Sciences. — 2014. — № 3.
14. *Звягинцева Н.А.* КОНЪЮНКТУРА РЫНКА ЦЕННЫХ БУМАГ: ФЕНОМЕН ВЛИЯНИЯ ЧАСТНЫХ ИНВЕСТОРОВ С ИСПОЛЬЗОВАНИЕМ СОЦИАЛЬНЫХ СЕТЕЙ / Н. А. Звягинцева, К. А. Овчинникова. — Текст: электронный // Baikal Research Journal. — 2021. — № 3.
15. MACHINE LEARNING. — Текст: электронный // bigdataschool.ru: [сайт]. — URL: <https://www.bigdataschool.ru/wiki/machine-learning> (дата обращения: 24.10.2021).
16. Прогнозирование временных рядов с LSTM в Python. — Текст: электронный // Python по байтам: [сайт]. — URL: <https://pythobyte.com/time-series-prediction-using-lstm-with-pytorch-in-python-521ce3ed/> (дата обращения: 24.10.2021).
17. Коллектив, Авторы Книга: Рынок ценных бумаг: Шпаргалка — Коллектив Авторы /Шпаргалки от РИОРа. — Москва: РИОР, 2010. — с. — Текст: электронный.
18. Котировки акций — что это значит, как их читать инвестору. — Текст: электронный // fnam.ru: [сайт]. — URL: <https://www.fnam.ru/education/likbez/kotirovki-akciiy-cto-eto-znachit-kak-ix-chitat-investoru-20210526-16360/> (дата обращения: 05.11.2021).
19. Стилистический энциклопедический словарь Русского языка / Л. М. Алексеева, В. И. Аннушкин, Е. А. Баженова [и др.]. — 2-е издание, стереотипное: «ФЛИНТА», 2011. — 696 с. — Текст: электронный.
20. Машинное обучение — обработка текста. — Текст: электронный // machinelearningmastery.ru: [сайт]. — URL: <https://www.machinelearningmastery.ru/machine-learning-text-processing-1d5a2d638958/> (дата обращения: 30.10.2021).
21. Обучаем компьютер чувствам. — Текст: электронный // habr.com: [сайт]. — URL: <https://habr.com/ru/post/149605/> (дата обращения: 30.10.2021).