

Содержание

Аннотация	3
1 Введение	5
1.1 Описание предметной области	5
1.2 Постановка задач	5
1.3 Актуальность	6
2 Аналитический обзор литературы	7
2.1 Индикаторы	7
2.1.1 Индикатор Ишимоку	7
2.2 Методы машинного обучения	8
2.2.1 Метод обработки естественного языка (NLP)	8
2.2.2 Рекуррентные нейронные сети (RNN)	9
2.2.3 Метод анализа тональности	9
2.3 Статистические величины	10
2.3.1 Временные ряды	10
2.4 Анализ существующих исследований по теме	10
3 Экспериментальное исследование модели	12
3.1 Анализ особенностей модели для данного исследования	12
3.2 Постановка задач	12
3.2.1 Содержательная постановка задач	12
3.2.2 Математическая постановка задач	13
3.3 Описание предлагаемого метода	14
3.4 Теоретический анализ метода	14
3.5 Эмпирический анализ метода	15
3.5.1 Содержательная постановка	15
3.5.2 Математическая постановка	15
3.5.3 Оценка полученных результатов	17
4 Заключение	18
Список литературы	19

Аннотация

В данной работе рассматривается построение модели для оценки будущей стоимости акций компании «Сбербанк» посредством анализа данных, полученных из различных новостных источников. Система полностью основана на информации, собираемой в режиме реального времени из открытых источников — новостных порталов, аналитических ресурсов и социальных сетей, что позволяет оперативно реагировать на изменения в информационном поле.

Структура системы состоит из двух основных модулей. В первом модуле происходит сбор, предварительная обработка и анализ новостной информации с использованием методов обработки естественного языка, анализа тональности и семантического анализа. Различные модели машинного обучения и нейронные сети обучаются выявлять скрытые зависимости между событиями, отражёнными в новостях, и изменениями стоимости акций. На основе этого проводится оценка влияния новостного фона на рынок. Второй модуль включает применение временных рядов для прогнозирования будущей цены акций «Сбербанка». Полученные на первом этапе результаты интегрируются с историческими данными о котировках, что позволило сформировать комплексную оценку будущей динамики акций. Итоговый прогноз может быть использован как основа для принятия инвестиционных решений и оптимизации инвестиционного портфеля, с учетом заданных пользователем критериев риска и доходности. Таким образом, предлагаемая система позволит инвестору получать объективные и своевременные оценки перспективности акций «Сбербанка», опираясь на глубокий анализ новостного фона и финансовых показателей, что будет способствовать более обоснованному принятию инвестиционных решений.

Ключевые слова

Акции, фондовый рынок, оценка стоимости, новостной фон, анализ новостных данных, обработка естественного языка, временные ряды, индикатор Ишимоку, машинное обучение, анализ тональности.

1 Введение

1.1 Описание предметной области

В наше время компьютерные науки стремительно развиваются и находят применение в самых различных сферах, в том числе и в области инвестиций. Сегодня, когда информация распространяется с невероятной скоростью, большое количество инвесторов обращают внимание на влияние новостных потоков на финансовые рынки и принимают решения на основе анализа актуальных данных. В данном исследовании были применены современные компьютерные технологии для оценки будущей стоимости акций "Сбербанка" посредством анализа информации из различных новостных источников.

Ключевая цель нашего исследования заключается в разработке модели, которая с помощью методов машинного обучения и обработки естественного языка способна выявлять закономерности между новостными событиями и изменениями котировок акций "Сбербанка". Существует множество подходов к анализу финансовых рынков, однако применение интегрированных моделей, учитывающих не только исторические данные, но и информационные потоки, позволяет значительно повысить точность прогнозирования.

Таким образом, исследование направлено на объединение передовых технологий анализа данных и финансового моделирования с целью создания адаптивного инструмента, способного оптимально оценивать динамику акций на основе информации, поступающей из различных новостных источников.

1.2 Постановка задач

Для достижения поставленной цели необходимо организовать сбор новостной информации из множества источников, включая официальные публикации, аналитические отчёты и социальные сети, а также провести предварительную обработку данных, включающую их очистку и нормализацию для дальнейшего анализа. Далее следует разработать и применить алгоритмы анализа тональности, позволяющие определить эмоциональную окраску новостей, а также выделить ключевые события и факторы, способные влиять на динамику котировок акций. После этого интегрировать обработанные новостные данные с историческими данными по ценам акций "Сбербанка" и применить методы машинного обучения для выявления закономерностей и построения прогностической модели. Завершающим этапом является проведение валидации и оптимизации модели с помощью бэктестов и сравнительного анализа полученных данных с действительными значениями стоимости акций "Сбербанка" в

определенный период времени, что позволит оценить точность и надёжность разработанного инструмента, а также корректировать его для повышения адаптивности к изменяющимся рыночным условиям.

1.3 Актуальность

В условиях стремительного развития информационных технологий и глобализации финансовых рынков всё более актуальной задачей для инвесторов, аналитиков и руководителей компаний становится правильно оценить будущую стоимость акций, особое значение приобретает применение новых методов анализа данных, позволяющих учитывать влияние информационных потоков на динамику котировок акций. Данное исследование является своевременным и значимым, поскольку оно совмещает современные подходы обработки больших данных и машинного обучения с традиционными методами финансового анализа. Использование новостных данных для прогнозирования цен акций позволяет выявлять малозаметные взаимосвязи между событиями в информационном поле и колебаниями рыночных индексов, что открывает новые перспективы для оптимизации инвестиционных стратегий и минимизации финансовых рисков. Кроме того, в условиях повышенной рыночной волатильности и усиливающейся конкуренции на мировых финансовых рынках, способность оперативно реагировать на информационные сигналы становится ключевым конкурентным преимуществом.

2 Аналитический обзор литературы

2.1 Индикаторы

2.1.1 Индикатор Ишимоку

Ишимоку Кинко Хё — это комплексный технический индикатор, разработанный для одновременного определения направления тренда, уровней поддержки и сопротивления, а также оценки силы движения цены. Он включает в себя несколько линий, таких как Tenkan-sen (линия конверсии), Kijun-sen (базовая линия), Senkou Span A и Senkou Span B (формирующие "облако" или "кумо") и Chikou Span (запаздывающая линия). Облако, которое создаётся на основе двух ведущих линий, помогает быстро определить доминирующий тренд: если цена находится выше облака, рынок считается бычьим, а если ниже — медвежьим. Такая многогранность позволяет видеть не только текущее состояние рынка, но и прогнозировать возможные изменения в его динамике.

Использование индикатора Ишимоку в данном исследовании обосновано тем, что он даёт объективные сигналы, отражающие технические тенденции рынка, что особенно важно при интеграции количественного анализа с качественными новостными данными. При сопоставлении рыночных трендов, выявляемых по Ишимоку, с эмоциональной окраской и содержанием новостных сообщений можно более точно прогнозировать поведение цены акций.

Определения и формулы для важных составляющих индикатора Ишимоку

Tenkan-sen (линия конверсии) — короткая линия тренда, значения которой определяются как среднее между наибольшим максимумом и наименьшим минимумом за короткий период в s (small) интервалов [5]:

$$\text{Tenkan}_t = \frac{\max_{i \in [t-s:t]} (\text{high}_i) + \min_{i \in [t-s:t]} (\text{low}_i)}{2} \quad (1)$$

Kijun-sen (базовая линия) — линия стандарта, значения которой определяются как среднее между наибольшим максимумом и наименьшим минимумом за период в m (medium) интервалов [5]:

$$\text{Kijun}_t = \frac{\max_{i \in [t-m:t]} (\text{high}_i) + \min_{i \in [t-m:t]} (\text{low}_i)}{2}. \quad (2)$$

Senkou Span A – среднее значение между Tenkan и Kijun, сдвинутое вперед на среднее m количество интервалов [5]:

$$\begin{aligned} \text{SenkouA}_t &= \frac{\text{Tenkan}_{t-s} + \text{Kijun}_{t-m}}{2} = \\ &= \frac{\max_{i \in [t-m-s:t-m]} (\text{high}_i) + \min_{i \in [t-m-s:t-m]} (\text{low}_i) + \max_{i \in [t-2m:t-m]} (\text{high}_i) + \min_{i \in [t-2m:t-m]} (\text{low}_i)}{4}. \end{aligned} \quad (3)$$

Senkou Span B – среднее значение между минимумом и максимумом за период в l (large) интервалов, сдвинутое на m интервалов вперед [5]:

$$\text{SenkouB}_t = \frac{\max_{i \in [t-m-l:t-m]} (\text{high}_i) + \min_{i \in [t-m-l:t-m]} (\text{low}_i)}{2}. \quad (4)$$

Chikou Span (запаздывающая линия) – сдвинутое назад на средний промежуток времени значение цены [5]:

$$\text{Chikou}_t = \text{close}_{t+m}. \quad (5)$$

Облако Ишимоку (Kumo) – область между линиями Senkou Span A и Senkou Span B: так как эти линии сдвинуты вперед на средний промежуток времени, по ним можно прогнозировать будущие изменения цены [5].

2.2 Методы машинного обучения

2.2.1 Метод обработки естественного языка (NLP)

Метод обработки естественного языка (NLP) представляет собой совокупность алгоритмов и техник, направленных на автоматический анализ, интерпретацию и генерацию текстовой информации, что позволяет структурировать неструктурированные данные, выявлять ключевые понятия, оценивать тональность высказываний и проводить тематическую классификацию [4].

В контексте исследования применение данного метода является особенно актуальным, поскольку новости содержат огромный объем информации, представленной в текстовом формате, который трудно анализировать вручную. Использование методов NLP позволяет оперативно обрабатывать и анализировать массовые текстовые данные, выявлять тренды и корреляции между упоминаниями компании в медиа и изменениями в ценах акций, а также проводить оценку эмоциональной окраски публикаций, что может служить индикатором общественного мнения и влияния на рыночные процессы.

2.2.2 Рекуррентные нейронные сети (RNN)

Рекуррентные нейронные сети представляют собой класс алгоритмов, специально разработанных для обработки последовательных данных, поскольку они обладают способностью запоминать и учитывать предыдущую информацию при анализе новых входных сигналов [1]. Такая особенность позволяет эффективно работать с временными рядами и текстовыми последовательностями, где важны как краткосрочные, так и долгосрочные зависимости.

В контексте исследования применение рекуррентных нейронных сетей становится оправданным, так как новости и финансовые показатели являются динамическими и изменяются во времени. RNN могут выявлять скрытые закономерности в потоке новостной информации, прогнозировать изменения и учитывать влияние исторических событий на будущие рыночные тенденции. Благодаря способности моделировать временные зависимости, такие сети помогают определить, какие информационные паттерны в новостях могут предшествовать изменению стоимости акций, что делает их ценным инструментом для создания более точных и адаптивных прогнозов в условиях высокой неопределенности и изменчивости рыночных данных.

2.2.3 Метод анализа тональности

Метод анализа тональностей представляет собой инструмент обработки естественного языка, позволяющий автоматически определять эмоциональную окраску текстовой информации, выделяя положительные, отрицательные и нейтральные оценки [4].

В контексте исследования его применение обосновано тем, что новости и публикации существенно влияют на восприятие компании инвесторами и, соответственно, на динамику котировок акций. С помощью анализа тональностей можно выявлять настроения, отраженные в медийном поле, что позволяет не только оценить текущее состояние мнения о компании, но и спрогнозировать возможные изменения на рынке. Такой подход способствует более глубокому пониманию влияния информационного поля на финансовые показатели и может служить дополнительным инструментом для принятия обоснованных инвестиционных решений, снижая риски и повышая точность прогнозов.

2.3 Статистические величины

2.3.1 Временные ряды

Временные ряды представляют собой последовательности наблюдений, собранных через равные интервалы времени, что позволяет выявлять тенденции, сезонные колебания и случайные колебания в динамике исследуемых параметров [3].

Применение анализа временных рядов в исследовании является необходимым, поскольку цены акций подвержены влиянию множества факторов, в том числе экономических новостей, событий на рынке и внутренних корпоративных изменений, проявляющихся во времени. Такой подход позволяет объединить исторические данные о ценах акций с информацией, извлечённой из новостных источников, для выявления закономерностей и построения прогностических моделей, способных предсказать будущую динамику рынка. Анализ временных рядов помогает не только определить общие тренды и периоды повышенной волатильности, но и установить корреляции между новостными событиями и изменениями в стоимости акций, что существенно повышает точность и обоснованность прогнозов.

2.4 Анализ существующих исследований по теме

В качестве теоретико-экономической основы для данного исследования была взята статья Владимировой О.А. "Влияние новостного фона на стоимость компании: обзор литературы и направления будущих исследований"[2], а в качестве практической - выпускная квалификационная работа А.Ю. Негрескула на тему "Математическое, алгоритмическое и программное обеспечение для анализа и прогнозирования курсов криптовалют"¹.

В статье О.А. Владимировой, как и в данном исследовании, рассматривается влияние информационного фона на капитализацию компании. В источнике подчеркивается, что новости способны вызывать эмоциональные реакции у инвесторов, что приводит к иррациональным колебаниям спроса и предложения – может отразиться на стоимости акций компании. Были приведены примеры, когда громкие новостные заголовки сильно отражались на биржевых котировках - например, в понедельник после сердечного приступа Эйзенхауэра в сентябре 1955 года рынок снизился на 6,62%; переход президентской власти после убийства президента Кеннеди совпал с ростом рынка на 3,98%, в то время как собственно новости о его убийстве привели к снижению стоимости акции почти на 3%. В дальнейшем в статье упоминается, что примерно в половине случаев очень сложно оценить влияние новостного

¹ВКР доступна по ссылке: <https://drive.google.com/file/d/1N46MeHNFp3zbcX3lUyzsG2QEEHmena3/view>, дата обр. 24.02.2025

фона на стоимость акций. Далее указывается, что из-за постоянного информационного шума на рынках является достаточно очевидным факт, что искажённые цены обнаруживаются довольно часто, хотя прогнозы были совсем противоположные – это и говорит о том, цены непосредственно реагируют на новости. В статье также упоминается, что компания должна выставлять себя в положительном свете в публичном поле, про неё должно быть достаточно хороших новостей в газетах и новостных источниках: всё это для того, чтобы поддерживать хорошую репутацию, а, следовательно, меньше зависеть от каких-то негативных вбросов, потому что даже фейковые плохие новости могут нанести колоссальные финансовые и имиджевые потери компании, что уж говорить о настоящих. Итогом статьи стал вывод, что влияние новостей на капитализацию компании достаточно спорно.

В работе А.Ю. Негрескула больший интерес представляют прикладные методы, а также использованные на практике метрики, упомянутые ранее уже в данном исследовании. В качестве основополагающей метрики был взят индикатор Ишимоку, далее с помощью алгоритма Случайный лес было произведено обучение модели с функцией потерь – критерием Джини.

3 Экспериментальное исследование модели

3.1 Анализ особенностей модели для данного исследования

В настоящее время исследования влияния новостного фона на стоимость акций компании ведутся достаточно активно, но какого-то определенного метода анализа выявлено не было – ключевых причин этого несколько: сама отрасль исследования сравнительно молодая – работы были начаты всего около 10 лет назад; разные подходы к выявлению зависимостей между новостями и стоимостью акций – какие-то исследователи считают, что нужно отслеживать влияние только негативных новостей, другие – только позитивных, третьи – абсолютно всех; огромное количество инструментов для анализа – в реалиях большого выбора всевозможных алгоритмов машинного обучения, анализа тональности текста и прогнозирования временных рядов кратно уменьшается вероятность прийти к какому-то единому методу.

3.2 Постановка задач

3.2.1 Содержательная постановка задач

На основании проведенного анализа существующих публикаций было сделано заключение, что в данный момент проведено большое количество исследований фондового рынка с применением различных математических и алгоритмических моделей, однако исследований, направленных на измерение влияния сторонних факторов на фондовый рынок было проведено не так уж и много, как уже было сказано выше. Поставлена задача – оценить влияние новостных сводок для прогнозирования фондового рынка не только на основании предыдущих значений индекса акции, но и на основании тональности новостей. В качестве общего новостного источника был выбран сервис NewsAPI, который позволяет находить новости по заданному ключевому слову и работать с ними в среде программирования. Были поставлены следующие подзадачи:

- 1 Получить исторические данные котировок акций компании "Сбербанк" и подсчитать для них технический индикатор Ишимоку;
- 2 Получить новостные данные о компании "Сбербанк" и провести анализ их тональности;
- 3 Составить из двух полученных массивов данных общий временной ряд, используемый для прогнозирования;

4 Разработать модель для прогноза изменения курса акции.

3.2.2 Математическая постановка задач

1 Сформировать временной ряд с признаками $\mathbf{x}(t)$ и целевой переменной $y(t)$, состоящий из векторов признаков для каждого момента времени t вида:

$$\mathbf{x}(t) = \left[C(t), \text{tenkan}(t), \text{kijun}(t), \text{Senkou_A}(t), \text{Senkou_B}(t), \text{Chikou}(t), s(t) \right] \in \mathbb{R}^7,$$

где:

- $C(t)$ — цена закрытия в момент времени t ;
- $\text{tenkan}(t)$, $\text{kijun}(t)$, $\text{Senkou_A}(t)$, $\text{Senkou_B}(t)$, $\text{chikou}(t)$ — линии индикатора Ишимуку (вычисление производится по формулам 1, 2, 3, 4 и 5 со стандартными параметрами);
- $s(t)$ — усреднённая оценка тональности новостных заголовков за t .

Целевая переменная определяется как:

$$y(t) = C(t);$$

2 Для каждого признака $x_j(t)$ применить Min-Max масштабирование:

$$\tilde{x}_j(t) = \frac{x_j(t) - \min\{x_j\}}{\max\{x_j\} - \min\{x_j\}},$$

где $\min\{x_j\}$ и $\max\{x_j\}$ — минимальное и максимальное значения признака j по всему датасету.

После масштабирования сформировать последовательность длиной L (например, $L = 30$):

$$X(t) = \left[\tilde{\mathbf{x}}(t - L + 1), \tilde{\mathbf{x}}(t - L + 2), \dots, \tilde{\mathbf{x}}(t) \right] \in \mathbb{R}^{L \times 6},$$

при этом целевая переменная для данной последовательности определяется как:

$$y(t) = \tilde{C}(t),$$

где $\tilde{C}(t)$ — масштабированное значение цены закрытия;

3 Обучить модель

$$f_{\theta} : \mathbb{R}^{L \times 7} \rightarrow \mathbb{R},$$

с параметрами θ , на обучающей выборке $\{X(t_i), y(t_i)\}_{i=1}^M$. Критерий обучения — среднеквадратичная ошибка (MSE):

$$\mathcal{L}(\theta) = \frac{1}{M} \sum_{i=1}^M \left(\tilde{C}(t_i) - f_{\theta}(X(t_i)) \right)^2.$$

Минимизация $\mathcal{L}(\theta)$ позволяет подобрать оптимальные параметры θ модели;

4 Спрогнозировать цену акций компании "Сбербанк" с использованием рекурсивного подхода и отмасштабировать его для координат дата – стоимость акций:

Пусть τ — индекс последнего наблюдения в исходном временном ряду. Для прогнозирования на F будущих периодов ($k = 1, 2, \dots, F$) применяется следующий алгоритм: Для каждого будущего момента $\tau + k$ формируется входная последовательность:

$$X(\tau + k) = \left[\tilde{\mathbf{x}}(\tau - L + k + 1), \dots, \tilde{\mathbf{x}}(\tau + k) \right].$$

Модель предсказывает масштабированное значение:

$$\hat{y}(\tau + k) = f_{\theta}(X(\tau + k)).$$

Для получения предсказанного значения цены закрытия в исходном масштабе применяется обратное преобразование:

$$\widehat{C}(\tau + k) = \min\{C\} + \hat{y}(\tau + k) \cdot (\max\{C\} - \min\{C\}).$$

3.3 Описание предлагаемого метода

Модель, построенная для данного исследования, основывается на использовании индикатора Ишимоку, обработке и анализе новостных сводок про "Сбербанк" с помощью NLP и анализа тональностей и обучении рекуррентной нейронной сети.

3.4 Теоретический анализ метода

В разработанной модели используются: алгоритм VADER для анализа тональностей текста – выбор пал на этот алгоритм из-за его возможности работать с достаточно малень-

кими выборками, ведь тональность рассматривалась для заголовков статей и из-за быстрой работы: методы машинного обучения не рассматривались из-за сложности в обучении на малом объеме данных, необходимости разметки данных и требовательности к вычислительным ресурсам; индикатор Ишимоку – он достаточно прост в подсчёте, а также обогащает данные, на которых обучается модель, что помогает ей лучше улавливать сложные технические паттерны и, как следствие, повышает точность прогнозов; рекуррентные нейронные сети, основанные на LSTM-слоях – данные о ценах акций представляют собой временные ряды, в которых порядок и длительные зависимости между наблюдениями имеют решающее значение, а LSTM специально разработаны для работы с последовательностями и способны «запоминать» важную информацию на протяжении длительного времени, что помогает преодолевать проблему затухающего градиента, присущую обычным RNN.

3.5 Эмпирический анализ метода

3.5.1 Содержательная постановка

На вход модели поступают исторические котировки акций "Сбербанка" с помощью функционала библиотеки `ufinance`, где запрашиваются данные за максимально возможный период, после чего полученные данные сохраняются в CSV-файл, а также новостные данные, получаемые через `NewsAPI`, где по ключевому слову «Сбербанк» на русском языке извлекаются заголовки и даты публикаций; эти данные агрегируются по дате с объединением заголовков и сохраняются в отдельный CSV-файл.

На выходе же модель выдаёт прогнозируемое скалированное значение цены закрытия, то есть цены акции в этот день, для определенного периода времени, которое затем обратным масштабированием преобразуется в реальное значение в рублях.

3.5.2 Математическая постановка

В самом начале происходит предобработка данных: считываются данные котировок, сортируются по дате, и для них вычисляются компоненты индикатора Ишимоку – линия конверсии (`tenkan-sen`), базовая линия (`kijun-sen`), ведущий спан A (`senkou span A`), ведущий спан B (`senkou span B`) и запаздывающий спан (`chikou span`) (по формулам 1, 2, 3, 4 и 5) с использованием скользящих окон, при необходимости заполняются пропуски. Одновременно загружаются новостные данные, сортируются по дате и для каждого заголовка с помощью алгоритма `SentimentIntensityAnalyzer` из библиотеки `nltk` вычисляется тональность, затем тональности усредняются для каждой даты. После этого данные котировок и усреднённые

значения тональности объединяются по дате, пропущенные значения тональности заменяются на ноль, а строки с отсутствующими данными, возникшими из-за расчёта скользящих окон, удаляются.

Далее определяются ведущие признаки – закрывающая цена, технические индикаторы Ишимоку и тональность, а также целевая переменная, равная закрывающей цене. Признаки масштабируются с использованием `MinMaxScaler` из библиотеки `sklearn`, после чего из масштабированных данных формируются временные последовательности фиксированной длины (30 шагов), где для каждой последовательности целевым значением является закрывающая цена следующего дня. Полученные последовательности разделяются на обучающую и тестовую выборки в пропорции 80/20.

Далее строится рекуррентная нейронная сеть на базе LSTM из библиотеки `tensorflow`, которая состоит из двух последовательных слоев LSTM, между которыми применяется `Dropout` для регуляризации, и завершается плотным слоем, выдающим прогнозируемую цену. Модель компилируется с оптимизатором `Adam` и функцией потерь в виде среднеквадратичной ошибки, после чего обучается в течение 50 эпох с размером пакета 32, используя тестовую выборку для контроля точности.

После обучения модели осуществляется прогнозирование на тестовой выборке, где масштабированные прогнозы преобразуются обратно в реальные значения с использованием обратного масштабирования. Результаты сравнения фактических и прогнозируемых цен визуализируются на графике с указанием дат.

В заключительной части алгоритма реализуется сценарный прогноз на 10 будущих дней. Для каждого из этих дней определяется новая дата, формируется входная последовательность из последних 30 наблюдений, масштабируется и подаётся на модель для получения прогнозируемой цены. При этом моделируется, что цена открытия равна предыдущей закрывающей, а максимальное и минимальное значения устанавливаются с поправкой в +1% и -1% от прогнозируемой цены соответственно; объём остаётся неизменным, а тональность новостного фона задаётся как среднее значение за последние 10 дней. Новые данные добавляются в основную `DataFrame`, после чего пересчитываются технические индикаторы Ишимоку с заполнением пропусков для обновлённого набора данных. Полученные прогнозируемые цены для будущих дней сохраняются и отображаются на графике, сравнивающем последние 50 реальных наблюдений с прогнозными значениями, а также выводятся на печать с указанием соответствующих дат.

3.5.3 Оценка полученных результатов



Рис. 3.1: Реальные и спрогнозированные стоимости акций "Сбербанка"

Как можно видеть на Рисунке 3.2, используемая модель достаточно хорошо спрогнозировала истинные значения стоимости акций – есть погрешность, но она не критична. В январе-феврале 2022 года наблюдается достаточно сильное расхождение прогноза и реальности, что связано с резкими событиями, которые не находили отражения в новостях до их происхождения.



Рис. 3.2: Прогноз стоимости акций "Сбербанка на конкретный период"

Из Рисунка 3.2 можно сделать вывод о стоимости акций на период с 10-15 по 20-25 апреля 2022 года – 120-123 рубля за одну акцию, что является достаточно хорошим приближением к реальной стоимости, которая за тот же период составляла 121-122 рубля за акцию.

В данной модели представлены данные до 2022 года, так как источники, использованные в данном исследовании, в большинстве своем иностранные, а в них данные по российским компаниям обновляются либо долго, либо не обновляются в принципе.

4 Заключение

В данном исследовании была успешно реализована модель, которая с помощью методов машинного обучения и обработки естественного языка способна выявлять закономерности между новостными событиями и изменениями котировок акций "Сбербанка". По визуализированным результатам работы данной модели можно сделать дополнительный вывод, который лишь частично совпадает с выводом из статьи О.А. Владимировой – влияние новостного фона на стоимость акций компании безусловно есть, однако положительный новостной фон в основном будет только подтверждать рост курса акций, который вызван другими причинами, а вот негативные же новости будут достаточно сильно влиять на котировки, что мы и видим на примере января-февраля 2022 года. В конечном итоге следует согласиться с выводами из статьи и сказать, что новостной фон имеет спорное влияние на курс акций.

Исследование может быть продолжено в двух различных плоскостях: либо улучшать точность и производительность модели, которая тогда сможет максимально точно выявлять зависимости между новостями и котировками, следствием чего станет возможность превращения её в полноценное программное обеспечение для помощи в прогнозировании курса акций различных компаний с отличной точностью; либо углубиться в проблему конкретного информационного поля – положительного, нейтрального или отрицательного, ведь каждая из этих областей достаточно плохо изучена, откуда следует невозможность в настоящее время точно определить корреляцию между областью новостного фона и биржевыми котировками

Список литературы

- [1] Андросова Е.Е. “Применение рекурсивных рекуррентных нейронных сетей”. В: *Новые информационные технологии в автоматизированных системах* 1.1 (2016), с. 5—7.
- [2] Владимирова О.А. “Влияние новостного фона на стоимость компании: обзор литературы и направления будущих исследований”. В: *Стратегические решения и риск-менеджмент* 1.1 (2018), с. 1—5.
- [3] Елизарова Л.В. Рапаков Г.Г. Горбунов В.А. “Исследование LSTM-нейросетевого подхода при моделировании временных рядов”. В: *Вестник Череповецкого государственного университета* 1.1 (2023), с. 2—7.
- [4] Джурабоев А.Э.У. Самигулин Т.Р. “Анализ тональности текста методами машинного обучения”. В: *Научный результат. Информационные технологии* 1.1 (2021), с. 2—5.
- [5] Чернышова О. Н. Черкашнев Р. Ю. Федорова А. Ю. “Прогнозирование динамики цены актива на основе синергии индикаторов «Облако Ишимоку» и «Аллигатор Билла Вильямса»”. В: *Бюллетень науки и практики* 1.1 (2018), с. 3—4.