

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение высшего образования
«Тольяттинский государственный университет»

Институт Математики, физики и информационных технологий
(наименование института полностью)

Кафедра «Прикладная математика и информатика»
(наименование)

02.03.03 Математическое обеспечение и администрирование информационных систем
(код и наименование направления подготовки / специальности)

Web-дизайн и мультимедиа
(направленность (профиль)/специализация)

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА (БАКАЛАВРСКАЯ РАБОТА)

на тему Разработка ПО для анализа и прогноза по криптовалютам

Обучающийся

В.В. Румянцев
(Инициалы Фамилия)

(личная подпись)

Руководитель

канд.пед.наук, доцент, Т.А. Агошкова

(ученая степень (при наличии), ученое звание (при наличии), Инициалы Фамилия)

Консультант

канд.филол.наук, доцент, М.М. Бажутина

(ученая степень (при наличии), ученое звание (при наличии), Инициалы Фамилия)

Тольятти 2022

Аннотация

Тема выпускной квалификационной работы – «Разработка ПО для анализа и прогноза по криптовалютам».

Работа посвящена анализу биржевых котировок и методам прогнозирования их изменения. В исследовании рассматривается алгоритм построения прогнозных моделей, основанный на анализе временных рядов.

Объект исследования – алгоритмы анализа временных рядов.

Предмет исследования – прогнозирование изменения биржевых котировок.

Цель исследования – создание эффективной модели для прогнозирования поведения, тенденций и изменений цен финансовых инструментов на основе принципов анализа временных рядов и сверточных нейронных сетей.

Данная работа состоит из введения, трех глав, заключения и списка литературы.

Во введении определяются актуальность темы, цель и задачи, поставленные в работе, а также объект и предмет исследования.

В первой главе работы описываются методики для прогнозирования биржевых котировок, а также приводится обзор методов глубокого обучения.

Во второй главе описывается технология построения прогнозных моделей и архитектура web-сервиса для реализации предложенной технологии.

Третья глава посвящена оценке эффективности разработанного программного обеспечения.

Выпускная квалификационная работа содержит пояснительную записку объемом 48 страниц, 25 рисунков, 18 формул и список используемой литературы, состоящий из 25 источников.

Abstract

The title of the bachelor's thesis is "Development of software for cryptocurrency analysis and forecasting".

The thesis is devoted to the analysis of stock exchange quotes and methods of forecasting their changes. The study considers an algorithm for building forecasting models based on the analysis of time series.

The object of the study is time series analysis algorithms.

The subject of the study is the forecasting of changes in stock exchange quotations.

The purpose of the study is to create an effective model for predicting the behaviour, trends and price changes of financial instruments based on the principles of time series analysis and convolution neural networks.

The work consists of an introduction, three chapters, a conclusion and references.

The introduction defines the relevance of the topic, the purpose and objectives of the paper, as well as the object and subject of the research.

The first chapter of the paper describes techniques for predicting stock quotes, as well as an overview of deep learning methods.

The second chapter describes the technology for building forecasting models and the architecture of the web-service for the implementation of the proposed technology.

The third chapter is devoted to evaluating the effectiveness of the developed software burnout.

The thesis contains an explanatory note of 48 pages, 25 figures, and a list of 25 references.

Оглавление

Введение.....	5
Глава 1 Анализ методов глубокого обучения и инструментов прогнозирования биржевых котировок.....	7
1.1 Методики прогнозирования котировок на бирже	7
1.2 Обзор исследований по применению нейронных сетей для прогнозирования котировок на бирже.....	10
1.3 Исследование свёрточных нейронных сетей для прогнозирования изменения биржевых котировок	12
1.4 Анализ временных рядов	14
1.5 Стационарные временные ряды и их характеристики.....	15
1.6 Нахождение тренда временного ряда для построения регрессионной модели.....	16
1.7 Сезонные колебания временного ряда как элемент регрессионной модели	20
Глава 2 Подготовка данных и построение предсказательной модели	25
2.1 Подготовка данных для построения модели временного ряда	25
2.2 Аддитивная модель временного ряда	28
2.3 Построение аналитической модели временного ряда.....	31
2.4 Архитектура web-сервиса и web-моделирование.....	35
Глава 3 Разработка программного обеспечения и тестирование	38
3.1 Программное обеспечение для прогнозирования котировок.....	38
3.2 Апробация программного обеспечения.....	42
Заключение	44
Список используемой литературы	46

Введение

В настоящее время рост популярности инвестиции среди населения связан с желанием сформировать дополнительный источник дохода. Среди наиболее популярных финансовых активов можно отметить акции и облигации, однако в течение последних нескольких лет увеличился спрос инвесторов на криптовалюты. Инвестирование в различные финансовые активы в долгосрочной перспективе приводит к росту личного капитала. Однако в краткосрочной перспективе по причине волатильности финансовых активов существует вероятность получения убытка при неудачном выборе времени совершения сделок.

Для минимизации риска потери средств на бирже необходимо уметь предсказывать направление изменения котировок финансовых активов. Раньше анализом биржевых котировок занимались эксперты, которые, пользуясь техническим и фундаментальным анализом, предсказывали будущие цены финансовых активов. Одна с развитием технологий машинного анализа данных стало возможным автоматизировать построение прогнозов. Возможно, что за счет интеллектуального анализа данных из управления финансовыми активами можно полностью исключить человеческий фактор.

Целью данной работы является создание эффективной модели для прогнозирования поведения, тенденций и изменений цен финансовых инструментов на основе принципов анализа временных рядов и сверточных нейронных сетей.

Объект исследования – алгоритмы анализа временных рядов.

Предмет исследования – прогнозирование изменения биржевых котировок.

Для достижения поставленной цели в бакалаврской работе решаются следующие задачи:

– проведение анализа методов прогнозирования биржевых котировок финансовых активов;

- разработка программного обеспечения для прогнозирования изменения цены финансовых активов;
- проверка работы программного обеспечения.

При выполнении бакалаврской работы разработано программное обеспечение обеспечивающее выполнение для выбранного пользователем финансового актива прогнозирование изменения на трехмесячном временном интервале.

Выпускная квалификационная работа содержит пояснительную записку объемом 48 страниц, 25 рисунков, 18 формул и список используемой литературы, состоящий из 25 источников.

Глава 1 Анализ методов глубокого обучения и инструментов прогнозирования биржевых котировок

1.1 Методики прогнозирования котировок на бирже

Криптовалюта, наравне с акциями, облигациями и производными инструментами (фьючерсами и опционами) является финансовым активом. Финансовые активы в современном мире торгуются на электронных биржах. Цена на любой финансовый актив определяется балансом спроса и предложения, а потоком заявок на покупку и продажи актива. Наименьшая цена актива, при которой биржевая заявка на покупку актива может быть выполнена моментально, называется рыночной ценой. Рыночная цена на любой финансовый актив меняется во времени. Из колебаний цены актива можно извлекать прибыль, если уметь прогнозировать направление изменения его стоимости. Зарабатывать можно как на росте цены, покупая актив дешевле и продавая дороже, так и на снижении цены, занимая у биржи актив и выкупая его через некоторое время по более низкой цене.

Задача прогнозирования котировок на бирже сводится к определению на заданном временном интервале будущей рыночной цены финансового инструмента, торгуемого на бирже [1]. Успешное прогнозирование будущих цен финансовых активов позволяет не только зарабатывать участникам биржевых торгов, но и обеспечивает предсказание многих экономических, социальных и политических процессов. Например, известно, что цены на криптовалютные активы растут во время усиления инфляции. Поэтому рост цены на криптовалюты, с точки зрения прогнозирования, можно считать опережающим индикатором роста цен на товары широкого потребления [7][8].

Анализируя вопрос о возможности использования исторических данных об изменении цены анализируемого актива для предсказания его цены в будущем можно опереться на гипотезу об эффективном рынке. Эта гипотеза

утверждает, что в текущей цене актива объективно отражена вся имеющаяся позитивная и негативная экономическая информация.

Методы прогнозирования делятся на три основные категории, которые при этом могут пересекаться. Это технический анализ, фундаментальный анализ и технология интеллектуального анализа данных:

Фундаментальный анализ на предположении, что определяющим на цену актива значением обладают его фундаментальными показателями, такие, например, как капитализация криптовалюты, длительность существования актива, используемый блокчейн и т.д. [6].

Специалисты фундаментального анализа следят за изменением этих показателей, чтобы найти моменты времени, когда актив стоит ниже и выше, чем он должен стоять по фундаментальным показателям. Такой подход не соответствует гипотезе об эффективном рынке, но все равно применяется на практике [12][13].

Одним из сервисов, предоставляющий инструменты для фундаментального анализа активов, является yahoo finance (рисунок 1).

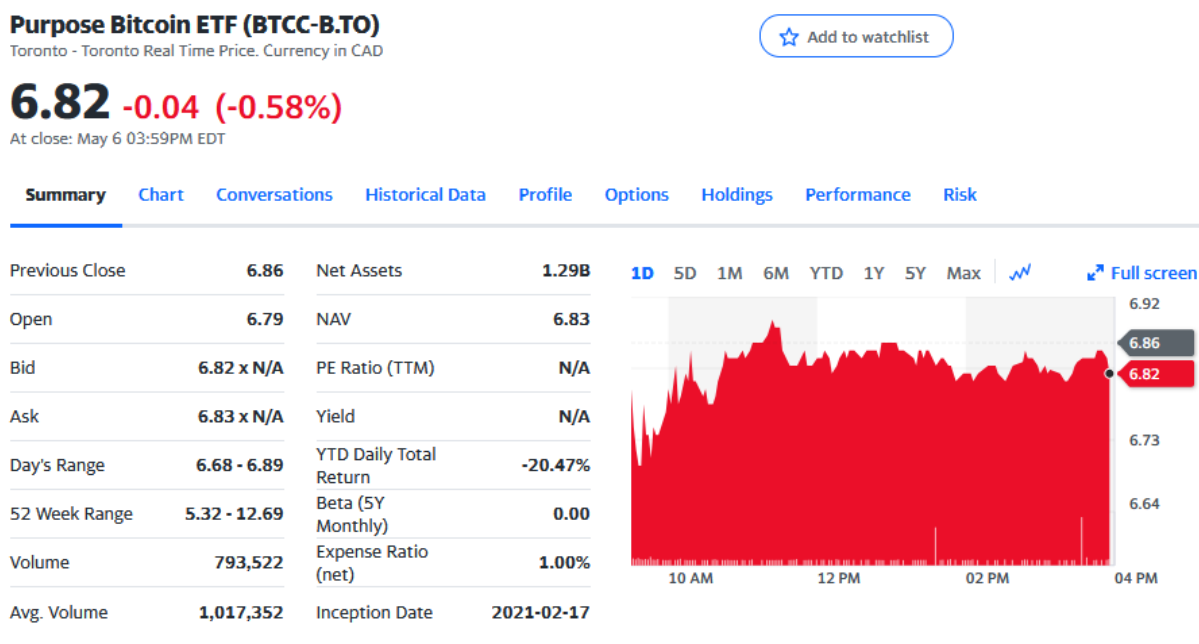


Рисунок 1 – Инструменты фундаментального анализа сервиса Yahoo finance

Технический анализ основан на предположении о том, что цена на актив формируется за счет постоянной переоценки событий участниками торгов на бирже. В этом случае цена расстраивается как реакция участников торгов на меняющиеся факторы. В техническом анализе считается, что все участники торгов следят за изменением цены актива по его графику и совершают сделки, основываясь на ожиданиях его дальнейшего изменения цены. Упрощенно технический анализ можно интерпретировать, как попытка предсказать поведения толпы на бирже, исходя из того, что толпа следит за графиком изменения актива [5].

Приведем пример технического анализа на графике изменения цены актива Purpose Bitcoin ETF (тикер BTCC) (рисунок 2). В качестве метода анализа будем использовать Bollinger bands.



Рисунок 2 – Результат технического анализа Purpose Bitcoin ETF с использованием метода Bollinger bands (20, C, 2, MA, Y)

Данный метод позволит определять на основе цены актива (синяя линия) для каждого момента времени определять справедливое значение цены

(средняя оранжевая линия), область переоценённости актива (цветная область выше оранжевой линии) и область недооценённости актива (цветная область ниже оранжевой линии). Если текущая цена актива (синяя линия) находится в области недооценённости, то в соответствии с техническим анализом это является хорошим моментом для покупки. Предполагается, что через некоторое время цена актива вернется к справедливому значению.

С развитием компьютерной техники появился еще один подход по прогнозированию изменения котировок финансовых активов - это интеллектуальный анализ данных. Это привело к тому, что анализ котировок перешёл из разряда экономических задач в разряд технологических. Теперь поиском закономерностей в данных занимается не человек, а компьютер. Наиболее известными методами, применяемыми при интеллектуальном анализе котировок, являются искусственные нейронные сети (ANN) и генетические алгоритмы (GA) [14].

Искусственные нейронные сети можно рассматривать как механизм аппроксимации биржевых данных с помощью математических функций. Чаще всего для прогнозирования котировок используются нейронные сети прямого распространения, обученные методом обратного распространения ошибки. На втором месте по популярности находят рекуррентные нейронные сети (RNN). Среди данного типа нейронных сетей для прогнозирования котировок финансовых активов применяются сети Элмана и Джордана [3][4].

1.2 Обзор исследований по применению нейронных сетей для прогнозирования котировок на бирже

В ряде исследований посвященных использованию нейронных сетей для прогнозирования биржевых котировок утверждается, что существует 2 основных подхода. Первый подход, который называется «независимым», и он основан на использовании отдельной нейронной сети для каждого временного горизонта (1 день, 3 дня и неделя). Преимуществом такого подхода является в

том, что ошибка прогнозирования на одном временном горизонте не влияет на точность прогнозов других временных диапазонов. Так эти нейронные сети работают независимо друг от друга.

Второй подход, который называется «совместным», основан на использовании одной нейронной сети для прогнозирования изменения цен одновременно на всех временных горизонтах. Преимущество такого подхода заключается в том, что обучить одну нейронную сеть проще, чем набор нейронных сетей [2]. При использовании обоих подходов на вход нейронных сетей подается вектор значений, описывающий прошлые (исторические) значения цены актива и на выходе нейронная сеть выдает ожидаемое значение цены.

В других исследованиях, предлагается вместо предсказания с помощью нейронной сети будущей цены актива генерировать простые сигналы на покупку или продажу. При такой постановке задачи структура нейронной сети упрощается, так как на ее выходе требуется получать только два возможных значения: +1 при сигнале на покупку актива и -1 при сигнале на его продажу.

В других исследованиях, для упрощения структуры нейронной сети предлагается снизить размерность признакового пространства за счет изменения логики подачи входных значений в нейронную сеть. Вместо передачи на вход сети всех исторических значений цены актива, исследователи предлагают ориентироваться только на ключевые значения временного ряда, оценивая максимальные и минимальные значений цены.

В литературе на тему применения нейронных сетей для прогнозирования котировок отмечается следующее. «Использование нейронных сетей для анализа финансовой информации является альтернативой (или дополнением) для традиционных методов исследования, таким как статистический анализ и экспертный анализ индикаторов технического анализа, фундаментальных показателей. В силу своей адаптивности одни и те же нейронные сети могут использоваться для анализа нескольких инструментов и рынков, в то время как найденные для

конкретного инструмента закономерности из области технического анализа могут работать хуже или не работать вообще для других инструментов» [15].

«Системы, базирующиеся на искусственных нейронных сетях, в последние годы все активнее используются для прогнозирования финансовых рынков. Отличие этого подхода от технического анализа состоит в следующем» [15].

«Во-первых, нейросетевой анализ, в отличие от технического, не предполагает никаких ограничений на характер входной информации. Это могут быть как индикаторы данного временного ряда, так и сведения о поведении других рыночных инструментов. Недаром нейронные сети активно используют именно институциональные инвесторы (например, крупные пенсионные фонды), работающие с большими портфелями, для которых особенно важны корреляции между различными рынками» [15].

«Во-вторых, в отличие от технического анализа, основанного на общих рекомендациях, нейронные сети способны находить оптимальные для данного инструмента индикаторы и строить по ним оптимальную опять же для данного ряда стратегию предсказания. Более того, эти стратегии могут быть адаптивны, меняясь вместе с рынком, что особенно важно для молодых активно развивающихся рынков, в частности, российского» [15].

1.3 Исследование свёрточных нейронных сетей для прогнозирования изменения биржевых котировок

Наибольший интерес с точки зрения прогнозирования биржевых котировок представляют свёрточные нейронные сети. Это объясняется следующими причинами. Структура свёрточных нейронных сетей заточена на обработку больших массивов числовых данных [20][21]. Слои свёртки сети как предназначены для того, чтобы определять какие признаки во входных данных влияют на результат предсказаний. Если при использовании других типов нейронных сетей необходимо четко понимать, какие параметры

изменения цены актива влияют на конечный прогноз, то при использовании свёрточных сетей такой необходимости нет. В процессе настройки свёрточной сети она сама определит признаки на данных, на которые будет опираться при построении прогноза [22][23].

Алгоритм прогнозирования временных рядов при помощи нейронной сети включает следующие этапы:

- определение временного промежутка, формирование обучающей выборки;
- кодирование ввода-вывода (нейросети могут обрабатывать только числовые данные);
- нормировка данных;
- предварительная обработка данных (удаление очевидных правил из данных помогает нейросетям распознать нетривиальные шаблоны);
- обучение нескольких нейронных сетей, отличных по своему строению;
- выбор лучшей сети, которая на выходе даёт наименьшую ошибку;
- прогнозирование и принятие решений.

Свёрточная нейронная сеть (CNN) – это искусственная нейронная сеть с обратной связью, в которой паттерн связи между нейронами основан на принципах обработки зрительных сигналов корой мозга. Первичная зрительная кора (V1) обнаруживает края объектов по необработанному визуальному входу сетчатки. Вторичная зрительная кора (V2) получает краевые характеристики от V1 и извлекает простые визуальные признаки, такие как направление, пространственная частота и цвет. Зрительная область V4 (нижняя височная извилина) обрабатывает более сложные визуальные признаки объекта. Все обработанные визуальные признаки переносятся в целевую область для распознавания объектов. Упрощенный математический аппарат обработки сигналов на участке между V1 и V4, по сути, и является свёрточной нейронной сетью (рисунок 3).

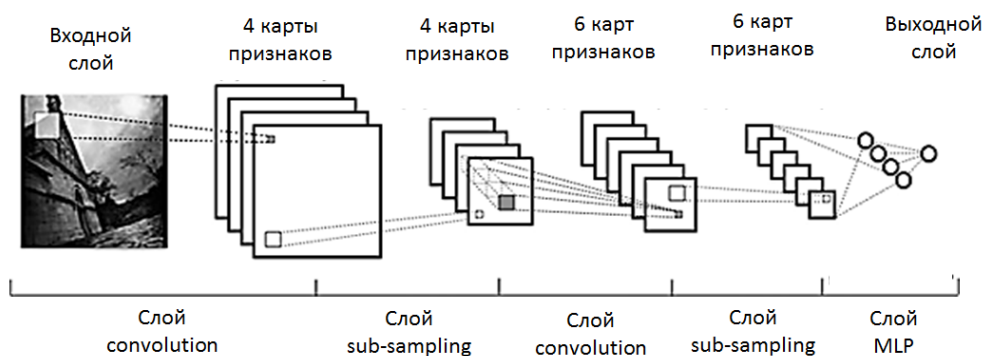


Рисунок 3 – Свёрточная нейронная сеть

Хоть сверточная нейронная сеть и была предназначена для обработки изображений, но фактически это модель может работать с любыми данными, представленными в числовом виде. Например, сверточные нейронные сети успешно используются для классификации текстов. В нашем случае сверточная сеть будет применяться для анализа временных рядов.

1.4 Анализ временных рядов

Временной ряд представляет собой набора наблюдений, полученных путем регулярного измерения одной переменной в течение некоторого периода времени. Форма данных для типичного временного ряда – это одна последовательность или список наблюдений, выполненных через равные промежутки времени [9].

Одна из наиболее важных причин проведения анализа временных рядов – это попытка спрогнозировать будущие значения ряда. Модель ряда, объясняющая предыдущие значения может также предсказать поведение следующих значений, будут ли они увеличиваться или уменьшаться и насколько сильно. Очевидно, что возможность таких успешных предсказаний очень важна для любой области бизнеса или науки [10].

Существуют две основные цели анализа временных рядов: определение природы ряда и прогнозирование (предсказание будущих значений

временного ряда по настоящим и прошлым значениям). Обе эти цели требуют, чтобы модель ряда была идентифицирована и, более или менее, формально описана. Как только модель определена, становится возможным с ее помощью интерпретировать рассматриваемые данные. На основе найденной модели возможна экстраполяция ряда, т.е. предсказание его будущих значений [11].

«Как и большинство других видов анализа, анализ временных рядов предполагает, что данные содержат систематическую составляющую (обычно включающую несколько компонент) и случайный шум (ошибку), который затрудняет обнаружение регулярных компонент. Большинство методов исследования временных рядов включает различные способы фильтрации шума, позволяющие увидеть регулярную составляющую более отчетливо» [9].

Наиболее распространенные компоненты временных рядов делятся на две категории: это компоненты тренда или сезонные компоненты. Тенденции обычно представляют собой линейные или нелинейные компоненты в системе, которые могут меняться со временем. Сезонные компоненты являются повторяющимися компонентами. Эти два типа стандартных компонента обычно совместно участвуют в формировании значений ряда. Например, прибыль компании может расти с каждым годом, но при этом зависеть от сезонных факторов.

1.5 Стационарные временные ряды и их характеристики

Стационарные временные ряды, вероятностные характеристики которых не меняются со временем, очень важны при анализе временных рядов.

Стационарные временные ряды используются при описании случайной составляющей анализируемого процесса. Если совместные распределения вероятностей n наблюдений x_1, x_2, \dots, x_n одинаковы, то временной ряд x_t , ($t = 1, 2, \dots, n$) называется строго стационарным [17].

Другими словами, характер такого ряда x_t не зависит от момента k , то есть закон распределения и его числовые характеристики не зависят от k . Поэтому, математическое ожидание и стандартное отклонение ряда оцениваются из x_t ($t = 1, 2, \dots, n$) по следующим формулам (1) и (2):

$$\bar{x}_t = \frac{\sum_{t=1}^n x_t}{n}, \quad (1)$$

$$\sigma_t^2 = \frac{\sum_{t=1}^n (x_t - \bar{x}_t)^2}{n}, \quad (2)$$

где \bar{x}_t – математическое ожидание;

σ_t^2 – стандартное отклонение ряда;

x_t – значение ряда;

n – количество значений в ряду.

Простейшим примером стационарного временного ряда с нулевым математическим ожиданием и некоррелированными ошибками является белый шум.

Следовательно, можно сказать, что помеха (ошибка) e_t в классической модели линейной регрессии образует белый шум, а в случае ее нормального распределения – это нормальный (гауссовский) белый шум.

1.6 Нахождение тренда временного ряда для построения регрессионной модели

Есть три основных типа трендов.

Первый и наиболее очевидный тип тренда – это тренд среднего, при котором временной ряд выглядит так, как будто он колеблется вокруг медленно увеличивающихся или убывающих значений.

Второй тип тренда – это тренд дисперсии. В этом случае амплитуда колебаний переменной изменяется со временем. Другими словами, это гетероскедастический процесс.

Как правило, экономические процессы с увеличением средних значений также имеют увеличенную дисперсию.

Третий, более тонкий тренд не всегда наблюдается визуально – это изменение корреляции между текущим значением последовательности и предыдущим значением, то есть тенденция автоковариации и автокорреляции.

Когда мы разбиваем ряд на несколько компонентов, то, что мы обычно называем трендом, относится к изменению среднего значения переменной, то есть тенденции среднего.

В рамках анализа среднего тренда выделяются следующие основные методы приближенных временных рядов и соответствующие основные типы средних трендов.

Полиномиальный тренд, пример которого показан на рисунке 4, определяется по формуле (3):

$$y = a_0 + a_1 t + \dots + a_p t^p \quad , \quad (3)$$

где y – значение тренда;

t – входной параметр тренда;

$a_0 \dots a_p$ – коэффициенты полинома;

p – степень полинома.

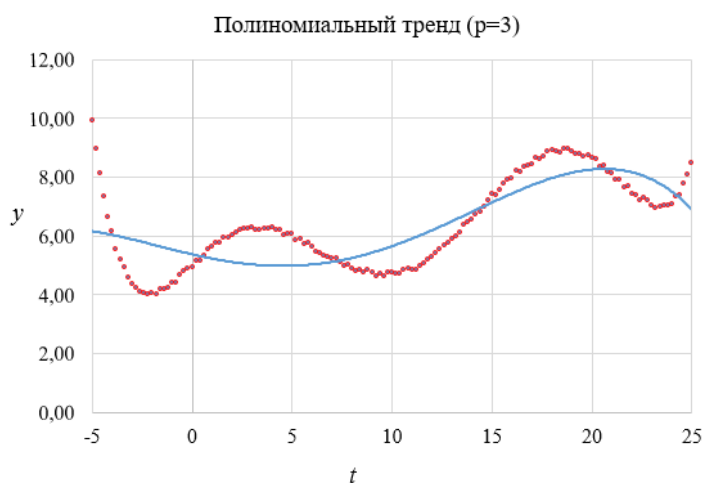


Рисунок 4 – Полиномиальный тренд (линия тренда показана синим)

Полиномиальные тренды используются для представления периодически возрастающих и убывающих значений временных рядов. Полиномы идеально подходят для анализа таких данных, как объемы сезонных продаж.

Экспоненциальный тренд – это характеристика процесса развития в среде, которая не накладывает никаких ограничений на горизонтальный рост. Видно, что на практике он может развиваться только в течение ограниченного времени, потому что рано или поздно любая среда будет иметь ограничения, и любые ресурсы в итоге будут исчерпаны.

Экспоненциальный тренд, пример которого показан на рисунке 5, определяется по формуле (4):

$$y = e^{a_0} \cdot e^{a_1 t} \cdot \dots \cdot e^{a_p t^p}, \quad (4)$$

где y – значение тренда;

t – входной параметр тренда;

$a_0 \dots a_p$ – коэффициенты экспоненциального тренда;

p – степень экспоненциального тренда.

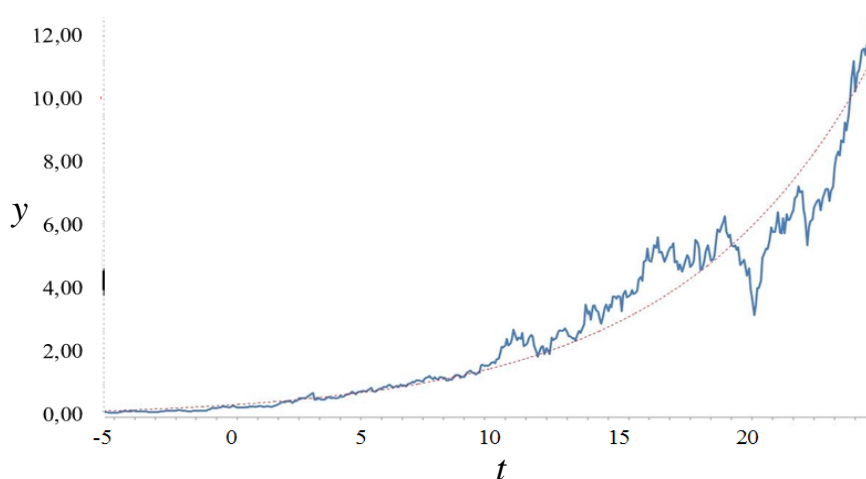


Рисунок 5 – Экспоненциальный тренд (тренд показан пунктирной линией)

Гармонический тренд, пример которого показан на рисунке 6, определяется по формуле (5):

$$y = R \cdot \cos(\omega t + \varphi), \quad (5)$$

где y – значение тренда;

t – входной параметр тренда;

ω, φ – коэффициенты гармонического тренда.

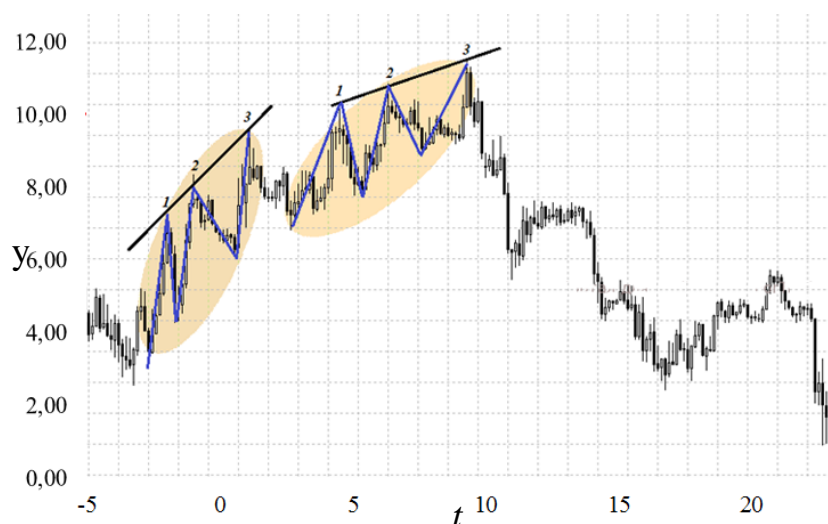


Рисунок 6 – Гармонический тренд (линия тренда показана синим)

Гармонический тренд состоит из периодических функций. В отличие от процесса выбора порядка полиномиального тренда, гармонический тренд не имеет порядка собственных частот (все гармоники равны).

Тренд может быть выражен логистической функцией (пример тренда показан на рисунке 7). Формула данного тренда (6):

$$y = \frac{k}{1+b \cdot e^{-at}}, \quad (6)$$

где y – значение тренда;

t – входной параметр тренда;

k, α – коэффициенты логистической функции.

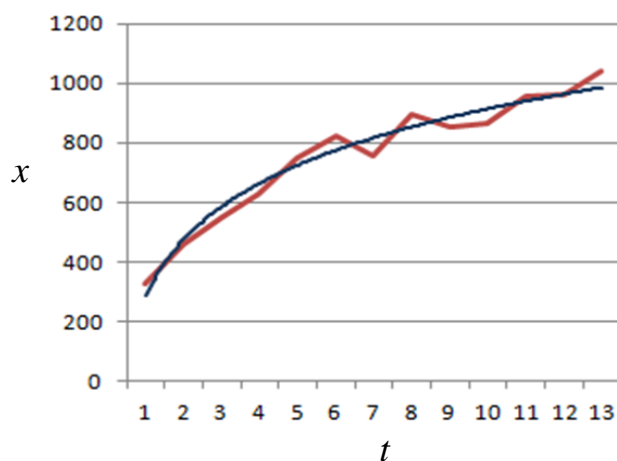


Рисунок 7 – Логистическая кривая (линия тренда показана синим цветом)

Рассмотрены наиболее известные виды трендов. Теперь перейдем к рассмотрению структуры временного ряда.

1.7 Сезонные колебания временного ряда как элемент регрессионной модели

Существует несколько способов анализа структуры временных рядов, содержащих сезонные или циклические колебания. Самый простой метод – построение аддитивной или мультипликативной модели временных рядов. В общем случае аддитивная модель определяется формулой (7):

$$Y = T + S + E, \quad (7)$$

где Y – аддитивная модель на основе суммы компонентов;

T – компонент, отвечающий за тренд;

S – сезонная составляющая;

E – случайная составляющая.

Модель предполагает, что каждый уровень временного ряда может быть выражен как сумма тренда T , сезонной S и случайной составляющей E .

В общем случае аддитивная модель определяется формулой (8):

$$Y = TSE, \quad (8)$$

где Y – аддитивная модель на основе произведения компонентов;

T – компонент, отвечающий за тренд;

S – сезонная составляющая;

E – случайная составляющая.

Эта модель предполагает, что каждый уровень временного ряда может быть выражен как произведение тренда T , сезонной S и случайной составляющей E . Выбор одной из двух моделей основан на анализе структуры сезонных колебаний. Если диапазон колебаний приблизительно постоянен, устанавливается модель аддитивных временных рядов, предполагающая, что значение сезонной составляющей остается постоянным в разные периоды. Если амплитуда сезонных колебаний увеличивается или уменьшается, устанавливается модель умножения временного ряда, которая определяет уровень ряда в соответствии со значением сезонной составляющей. Построение моделей сложения и умножения упрощается для вычисления значений T , S и E для каждого уровня последовательности.

Теоретические дисперсии $D(b_0)$, $D(b_1)$ зависят от дисперсии случайной ошибки σ^2 . Согласно данным выборки, отклонение ε_i и его дисперсия σ^2 неизвестны, поэтому они заменяются остатком наблюдения e_i и его дисперсией выборки. Несмещенная оценка дисперсии σ^2 – это величина, которая используется для измерения дивергенции линии регрессии и определяется формулой (9):

$$S^2 = \frac{\sum e_i^2}{n-2}, \quad (9)$$

где S – стандартная ошибка регрессии;

e_i – остаток i -ого наблюдения,

n – количество наблюдений.

Значение $S = \sqrt{S^2}$ называется стандартной ошибкой регрессии. Используя оценочное значение S^2 для замены неизвестного σ^2 в теоретической дисперсии, мы получаем оценку дисперсии (10):

$$S_{b_0}^2 = \frac{\bar{x}^2 \cdot S^2}{n \cdot (\bar{x}^2 - \bar{x}^2)}, \quad S_{b_1}^2 = \frac{S^2}{n \cdot (\bar{x}^2 - \bar{x}^2)}, \quad (10)$$

где $S_{b_0}^2$ – дисперсия величины b_0 ;

$S_{b_1}^2$ – дисперсия величины b_1 ;

S – стандартная ошибка регрессии;

\bar{x} – среднее отклонение;

n – количество наблюдений;

Значение $S_{b_0} = \sqrt{S_{b_0}^2}$, $S_{b_1} = \sqrt{S_{b_1}^2}$ – это стандартное отклонение случайных величин b_0 и b_1 , которое называется коэффициентом регрессии. Знаменатель дроби (10) содержит сумму квадратов отклонений значений x_i от среднего \bar{x} . На рисунке 8, к примеру, через две пары точек (1, 3) и (2, 3) проведена одна и та же прямая. Однако расстояние (1, 3) шире расстояния (2,3). В случае, если вместо точки 3 взять точку 3_a или 3_b (т.е. при изменении выборки), то наклон прямой для пары (1, 3) изменится гораздо меньше, чем для пары (2, 3).

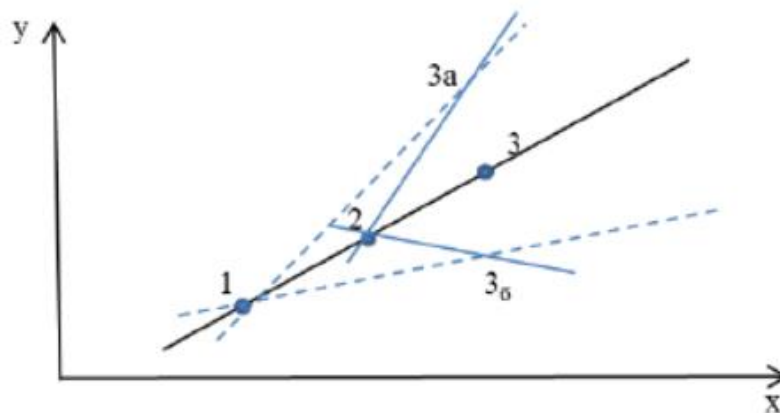


Рисунок 8 – Стандартные ошибки регрессии

Выборочная дисперсия свободного члена уравнения регрессии прямо пропорциональна дисперсии $S_{b_1}^2$ (11):

$$S_{b_0}^2 = S_{b_1}^2 \cdot \frac{\sum x_i^2}{n}. \quad (11)$$

где $S_{b_0}^2$ – дисперсия величины b_0 ;

$S_{b_1}^2$ – дисперсия величины b_1 ;

x_i – отклонение i -ого значения;

n – количество значений.

Фактически, чем больше изменение наклона прямой, проведенной через определённую точку (\bar{x}, \bar{y}) , тем больше разброс значения пересечения, которое характеризует пересечение прямой и оси OY .

На рисунке 9 прямая проходит через две пары точек (1, 2) и (3, 4) и пересекает ось OY в точке $(0, b_0)$. У одной из этих пар абсолютное значение значения переменной X (X и Y изменяются в одном диапазоне) меньше, чем у другой.

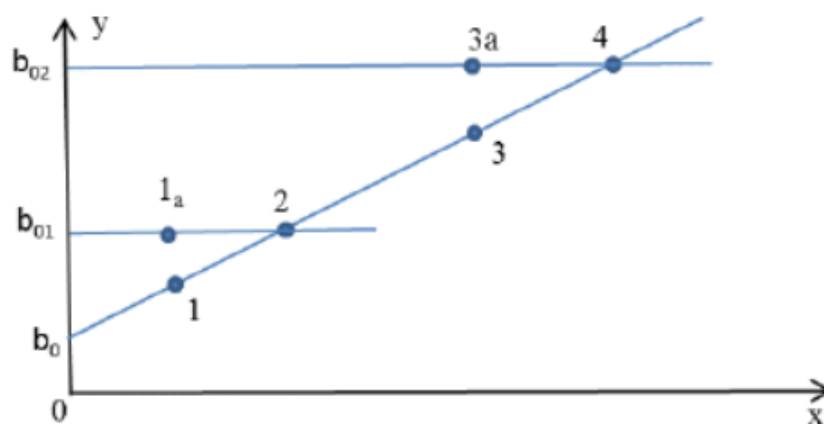


Рисунок 9 – Зависимость разброса значений свободного члена от средней величины x^2

Если точки 1 и 3 в этих парах переместить на одинаковое расстояние (1_a , 3_a), наклоны полученных линий (1_a , 2) и (3_a , 4) будет совпадать. Однако разница между свободным членом b_{01} и b_0 первой прямой будет намного меньше, чем между свободным членом b_{02} и b_0 для второй прямой.

Выводы по первой главе

Задача прогнозирования котировок финансовых активов представлена в виде анализа данных временного ряда. При этом прогнозирование будущих значений котировок основано на анализе предыдущих исторических значений. В общем случае, требуется построить регрессионную модель, которая максимально точно описывала бы исторические данные котировок.

Показаны компоненты регрессионной модели, участвующие в формировании прогнозируемого значения ряд. Прогнозирование основано на экстраполяции модели на будущие значения котировок.

Рассмотрены основные компоненты регрессионной модели, такие как тренд, сезонная и случайная составляющая.

В качестве регрессионной модели было предложено использовать свёрточную нейронную сеть. Показано, как свёрточная нейронная сеть может применяться для анализа временных рядов, представленных в векторном виде.

Глава 2 Подготовка данных и построение предсказательной модели

2.1 Подготовка данных для построения модели временного ряда

Одной из первоначальных задач прогнозирования котировок финансовых активов является получение доступа к их историческим данным. Существует специальные сервисы – поставщики данных, которые хранят и предоставляют по запросу информацию с прошлыми значениями котировок выбранного финансового актива. Одним из таких сервисов является Quandl, главная страница которого расположена по адресу demo.quandl.com (рисунок 10).

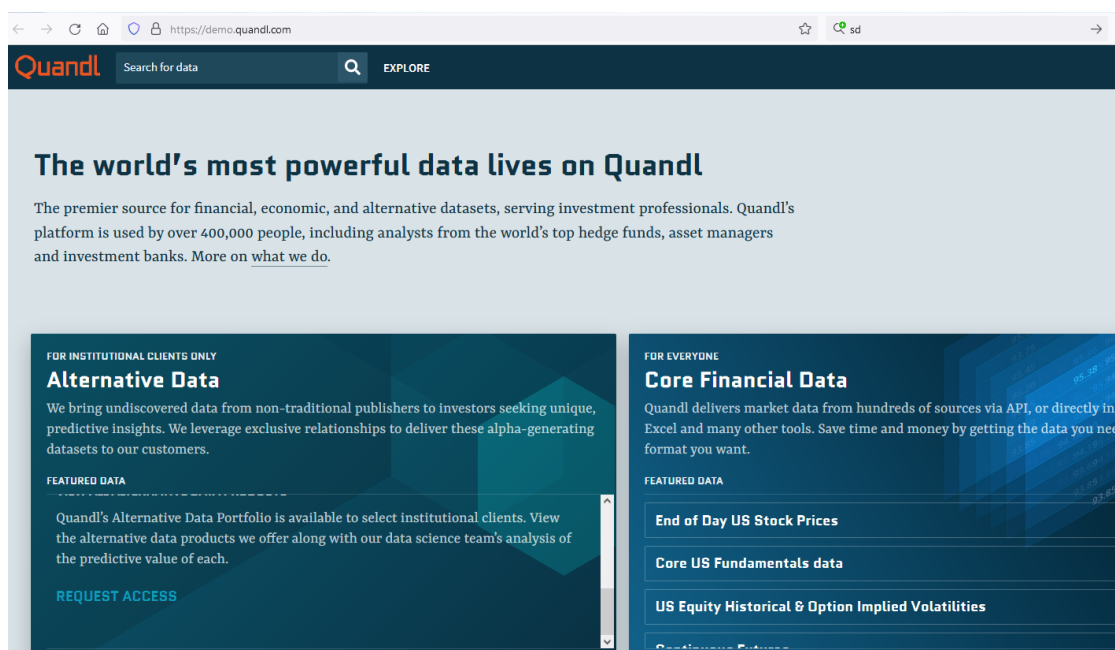


Рисунок 10 – Главная страница сервиса Quandl

Данный сервис обладает программным интерфейсом, обеспечивающим программистам возможность запроса и получения требуемых исторических финансовых данных. Схема работы программного интерфейса Quandl, представленной в официальной документации показана на рисунке 11.

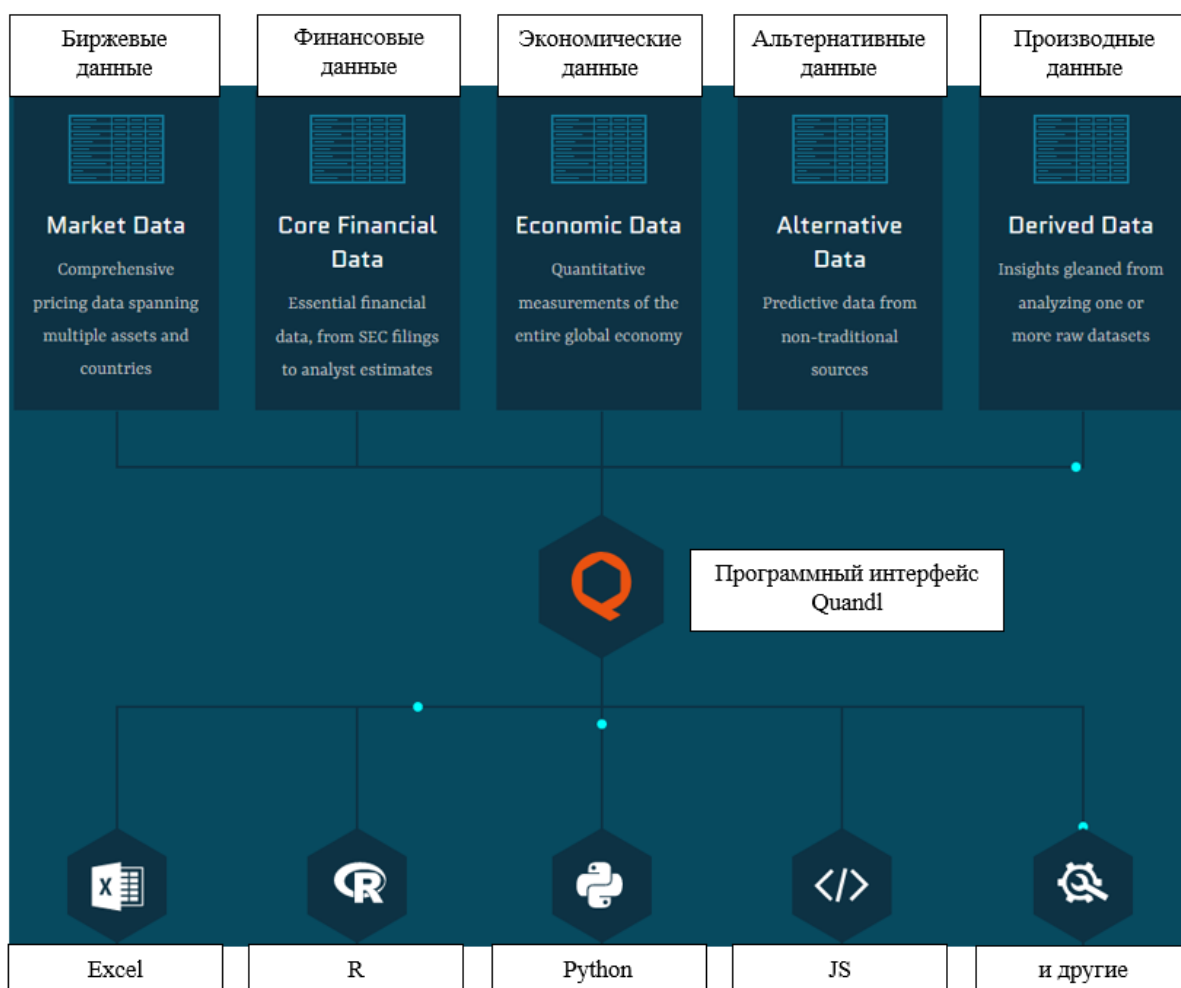
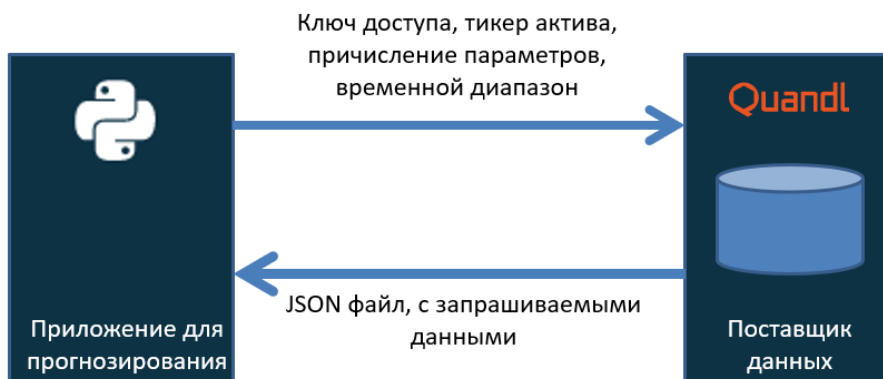


Рисунок 11 – Схема работы программного интерфейса Quandl

Для получения от такого сервиса информации о финансовом активе в виде json файла необходимо через программный интерфейс передать следующую информацию:

- ключ доступа к данным, который можно купить или получить бесплатно через страницу сервиса
- тикер финансового актива (это короткое зарегистрированное обозначение актива)
- перечисление параметров финансового актива, которые требуется получить, примеры параметров - цена актива на начало дня (Open), максимальная цена за день (High), минимальная цена за день (Low) и т.д. (рисунок 12).



Пример запроса на получение данных:
['aba3306518af...', 'TSLA', ['Close'], 'All']

Рисунок 12 – Схема получения данными с сервисом Quandl

В ответ на сформированный запрос Quandl собирает и отправляет данные в виде json файла. Пример полученных от Quandl данных показан на рисунке 13.

```
[ "Date", "Open", "High", "Low", "Close", "Volume", "Dividend", "Split", "Adj_Open", "Adj_High", "Adj_Low",
Series", "premium":true, "limit":null, "transform":null, "column_index":null, "start_date": "2013-09-0
28", 99.52, 99.62, 99.12, 99.4, 9763936.0, 0.0, 1.0, 92.95151589542357, 93.04491573305275, 92.5779165550078
27", 99.56, 99.6, 98.8299, 99.26, 5140793.0, 0.0, 1.0, 92.98887582946514, 93.0262357635067, 92.30696363337
26", 98.35, 99.44, 98.35, 99.16, 4295891.0, 0.0, 1.0, 91.85873782470767, 92.87679602734043, 91.85873782470
22", 98.14, 98.42, 97.68, 98.21, 5478016.0, 0.0, 1.0, 91.66259817098944, 91.92411770928042, 91.23295892951
21", 99.32, 99.33, 97.98, 98.06, 7368128.0, 0.0, 1.0, 92.76471622521572, 92.77405620872612, 91.51315843482
20", 99.48, 99.65, 98.45, 98.75, 10729605.0, 0.0, 1.0, 92.91415596138201, 93.07293568105868, 91.9521376598
19", 99.78, 99.91, 98.29, 98.8, 12011446.0, 0.0, 1.0, 93.19435546669378, 93.31577525232886, 91.80269792364
18", 97.44, 98.17, 97.36, 97.9, 7964156.0, 0.0, 1.0, 91.00879932526199, 91.69061812152063, 90.934079457178
15", 98.04, 98.49, 96.91, 97.11, 16141258.0, 0.0, 1.0, 91.56919833588553, 91.98949759385317, 90.5137801992
14", 98.415, 98.57, 97.1, 97.13, 9384839.0, 0.0, 1.0, 91.91944771752524, 92.06421746193631, 90.69123988590
13", 96.66, 97.91, 96.55, 97.76, 7877938.0, 0.0, 1.0, 90.28028061145139, 91.44777855025042, 90.17754079283
12", 96.85, 97.24, 96.6, 96.7, 7810617.0, 0.0, 1.0, 90.45774029814883, 90.82199965505414, 90.2242407103890
11", 96.44, 97.27, 96.29, 96.93, 8928919.0, 0.0, 1.0, 90.07480097422277, 90.85001960558532, 89.93470122156
08", 96.67, 96.7, 96.138, 96.55, 5820690.0, 0.0, 1.0, 90.28962059496179, 90.31764054549296, 89.79273347220
07", 97.06, 97.58, 96.65, 96.78, 7501473.0, 0.51, 1.0, 90.65387995186708, 91.13955909440747, 90.2709406279
06", 97.99, 98.14, 97.2401, 97.28, 7139465.0, 0.0, 1.0, 91.04273200664325, 91.18209734801478, 90.345998210
05", 97.0, 97.94, 96.92, 97.83, 9009407.0, 0.0, 1.0, 90.12292075359113, 90.99627689285273, 90.048592571526
04", 97.73, 97.96, 96.675, 97.01, 8955511.0, 0.0, 1.0, 90.8011654149326, 91.01485893836893, 89.82096251395
01", 97.61, 97.71, 96.23, 97.35, 8296075.0, 0.0, 1.0, 90.68967314183537, 90.78258336941639, 89.40751200121
30", 98.09, 98.44, 96.51, 97.23, 12853312.0, 0.0, 1.0, 91.13564223422428, 91.46082803075784, 89.6676606384
29", 97.22, 98.26, 96.88, 97.56, 11423531.0, 0.0, 1.0, 90.32732325426937, 91.29358962111202, 90.0114284804
28", 96.59, 96.885, 95.82, 96.77, 8709645.0, 0.0, 1.0, 89.74198882050895, 90.01607399187296, 89.0265800681
27", 96.85, 97.65, 96.3, 96.62, 8904600.0, 0.0, 1.0, 89.9835554122196, 90.72683723286778, 89.4725491605239
```

Рисунок 13 – Пример полученных данных

Сначала идут заголовки данных в двойных кавычках, перечисленных через запятую, а затем идут, сами значения запрашиваемых параметров, перечисленные через запятую.

Цены на акции представляют собой временные ряды длины N , определенные как p_0, p_1, \dots, p_{N-1} , в которых p_i - цена закрытия в день i , $0 \leq i < N$.

Если взять из полученных данных информацию о цене актива и визуализировать его, то можно получить график, показанный на рисунке 14.



Рисунок 14 - График изменения актива с тикером TSLA

Форма графика похожа на типичный случайный процесс, но это не совсем так. На самом деле в изменении цены актива присутствуют закономерности, которые необходимо выявить, чтобы иметь возможность прогнозировать изменение графика, или другими словами - временного ряда, на несколько месяцев вперед.

2.2 Аддитивная модель временного ряда

Любой временной ряд содержит два необходимых элемента: уровень ряда – значение какого-либо статистического показателя и время определенной статистики, представленной моментом или периодом уровня.

Прогнозирование – это описание возможного будущего состояния объекта.

Время (период) прогноза – промежуток времени с момента отображения последней статистики опроса до времени, связанного с прогнозом.

По времени прогнозирования экономические прогнозы делятся на:

- долгосрочные (более 5 лет);
- среднесрочные (от 1 года, до 5 лет);
- краткосрочные (от нескольких месяцев до года);
- оперативные (до одного месяца).

Каждая модель временного ряда состоит из тренда (Т), периодичности (S) и ошибки (E).

В аддитивной модели временного ряда уровень ряда выражается как сумма указанных компонентов (12):

$$Y = T + S + E, \quad (12)$$

где Y – аддитивная модель на основе суммы компонентов;

T – компонент, отвечающий за тренд;

S – сезонная составляющая;

E – случайная составляющая.

В мультипликативной модели временного ряда уровни ряда представлены в виде произведения указанных компонент (13):

$$Y = T \cdot S \cdot E, \quad (13)$$

где Y – аддитивная модель на основе произведения компонентов;

T – компонент, отвечающий за тренда;

S – сезонная составляющая;

E – случайная составляющая.

Автокорреляция уровней ряда – это корреляция между прошлыми и настоящими значениями уровня временного ряда (14):

$$r_i = \frac{cov(y_{t-i}; y_t)}{\sigma_{y_{t-i}} \sigma_{y_t}} = \frac{\overline{y_{t-i} \cdot y_t} - \overline{y_{t-i}} \cdot \overline{y_t}}{\sqrt{\overline{y_{t-i}^2} - \overline{y_{t-i}}^2} \cdot \sqrt{\overline{y_t^2} - \overline{y_t}^2}}, \quad (14)$$

где cov – ковариация;

σ – несмещенная оценка дисперсии;

y – значение ряда;

\bar{y} – среднее значение ряда.

Компоненты \bar{y}_i и $\overline{y_{t-i}}$ определяются по формулам (15) и (16):

$$\bar{y}_i = \frac{1}{n-i} \sum_{t=i+1}^n y_t, \quad (15)$$

$$\overline{y_{t-i}} = \frac{1}{n-i} \sum_{t=1}^{n-i} y_t, \quad (16)$$

где y_t – значение ряда на временном промежутке t ,

n – количество временных промежутков.

Построение аддитивной модели осуществляется в следующем порядке:

- выровнять исходный ряд методом скользящей средней;
- оценить сезонную компоненту S ;
- устранить сезонную составляющую из исходного уровня ряда и получить выровненные данные;
- аналитически выровнять уровни ряда с использованием полученного уравнения тренда;
- рассчитать полученные по модели значения;
- рассчитать абсолютные и относительные ошибки.

Таким образом, рассмотрен алгоритм построения аддитивной модели.

2.3 Построение аналитической модели временного ряда

Прогнозирование временных рядов - очень популярная задача анализа. Существует множество различных методов прогнозирования временных рядов. В этой работе для прогнозирования временных рядов используется библиотека Prophet [19].

Библиотека Prophet позволяет моделировать различные процессы, меняющиеся во времени, и обеспечивает возможность прогнозирования [18].

По умолчанию Prophet использует линейную модель для своего прогноза. При прогнозировании роста обычно существует некоторая максимально достижимая точка: общий размер рынка, общая численность населения и т.д. Это называется пропускной способностью, и прогноз должен достигать насыщения в этой точке.

По сути, это аддитивная регрессионная модель, которая выглядит следующим образом (17):

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t, \quad (17)$$

где $y(t)$ – регрессионная модель;

$g(t)$ – функция, задающая тренд;

$s(t)$ – функция, описывающая сезонную составляющую;

$h(t)$ – функция, описывающая нерегулярную составляющую;

ϵ_t – компонента, содержащая ошибку модели.

$s(t)$ - сезонная компонента, моделирующая циклические изменения, связанные с сезонностью. Например, недельная сезонность моделируется с использованием фиктивных переменных. Добавлено 6 дополнительных атрибутов, например [понедельник, вторник, среда, четверг, пятница, суббота], значения 0 и 1 в соответствии с датой. Атрибут воскресенья в модель не добавляется, так как он линейно зависит от других дней недели. Годовая сезонность моделируются при помощи ряда Фурье.

Тренд $g(t)$ – это кусочно-линейная или логистическая функция вида (18):

$$g(t) = \frac{c}{1+e^{-k(t-b)}}, \quad (18)$$

где k, b – коэффициенты логистической функции.

При таком тренде скорость роста уменьшается по мере увеличения показателя. Например, увеличение продаж или прибыли компании.

Библиотека также способна выбрать лучшую точку изменения тренда на основе исторических данных. Однако их можно установить самостоятельно, например, если известна дата выхода финансового отчёта компании, который может существенно повлиять на ключевые показатели.

Компонента $h(t)$ отвечает за аномальные дни, включая нерегулярные, такие как Черная пятница.

Ошибка ε_t содержит информацию, не обрабатываемую моделью.

В данном исследовании роль регрессионной модели играет свёрточная нейронная сеть со структурой, представленной на рисунке 15.

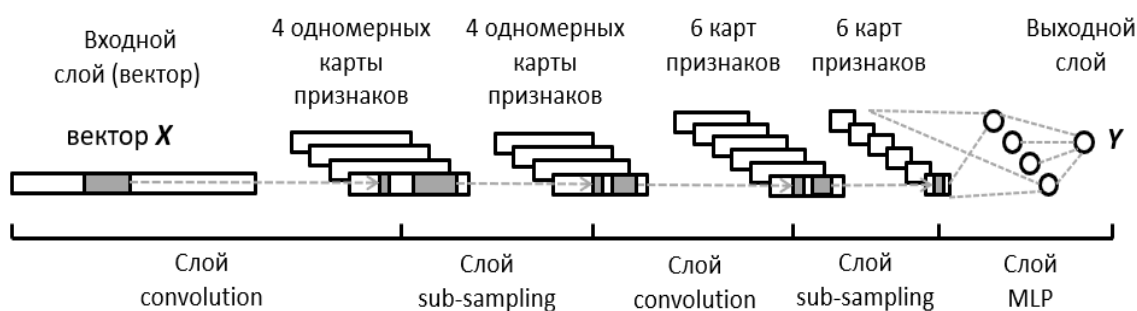


Рисунок 15 – Структура свёрточной нейронной сети для анализа временных рядов

На вход нейронной сети подается вектор значений X , который является фрагментом временного ряда длиной 120 значений. Каждое значение – это средняя цена актива за день. Затем с помощью слоев свертки (convolution) и

выборки (sub-sampling) производится извлечение макропризнаков из фрагмента временного ряда. Затем с помощью многослойного персептрона (MLP) производится прогнозирование средней цены на следующий день.

Запустив нейронную сеть еще раз и подав на ее вход временной ряд из 119 известных значений и одного спрогнозированного значения можно сгенерировать прогноз еще на один следующий день (на второй день). Аналогичным образом можно получить прогноз на 4, 5, 6 дни и т.д.

Чтобы обучить нейронную сеть необходимо сформировать обучающую выборку. Формировка обучающей выборки осуществляется с использованием метода сканирующего окна. Окно шириной 120 значений движется слева на право по историческим данным изменения цены актива, смещаясь на каждом шаге на одно значение. На каждом шаге 120 значений заносится в обучающую выборку как входные данные X , а 121 значение, как выходное значение Y . Таким образом постепенно генерируется обучающая выборка, которая в последствии используется для обучения нейронной сети с использованием алгоритма Adam (рисунок 16).



Рисунок 16 – Применение метода сканирующего окна для формирования обучающей выборки

Библиотека Prophet имеет интерфейс, аналогичный sklearn. Сначала мы создаем модель, затем вызываем для нее метод настройки модели, а затем получаем прогноз.

Входными данными для Prophet всегда является фрейм данных с двумя столбцами: ds и y. Столбец ds – время, тип этого поля должен быть date или datetime. Столбец y должен быть числовым, представляющим размер, который мы хотим предсказать.

Чтобы измерить качество полученных прогнозов, мы удалим данные за последний год из обучения и сделаем прогнозы на их основе.

```
predictions = 365  
train = df[:-predictions]
```

Далее мы создаем объект класса Prophet и обучаем его.

```
model = self.create_model()  
model.fit(train)
```

Используя функцию Prophet.make_future_dataframe, мы можем создать фрейм данных, который содержит исторические временные точки и еще 365 дней, которые мы хотим спрогнозировать.

Для прогнозирования мы вызываем функцию прогнозирования модели и передаем ей полученный на предыдущем шаге фрейм данных.

```
future = m.make_future_dataframe(periods=predictions)  
forecast = m.predict(future)
```

Для оценки результатов построения модели используем встроенные в библиотеку Prophet инструменты визуализации (рисунок 17).

Как видно из рисунка 17, программа достаточно точно предсказывает периоды роста и снижения стоимости акций, хоть и не справляется с точным предсказанием цены. Как и ожидалось, со временем неопределенность будет увеличиваться. Бурный рост котировок в 2020 году обусловлен новостями о росте онлайн-продаж компании, которые модель обработать не способна.

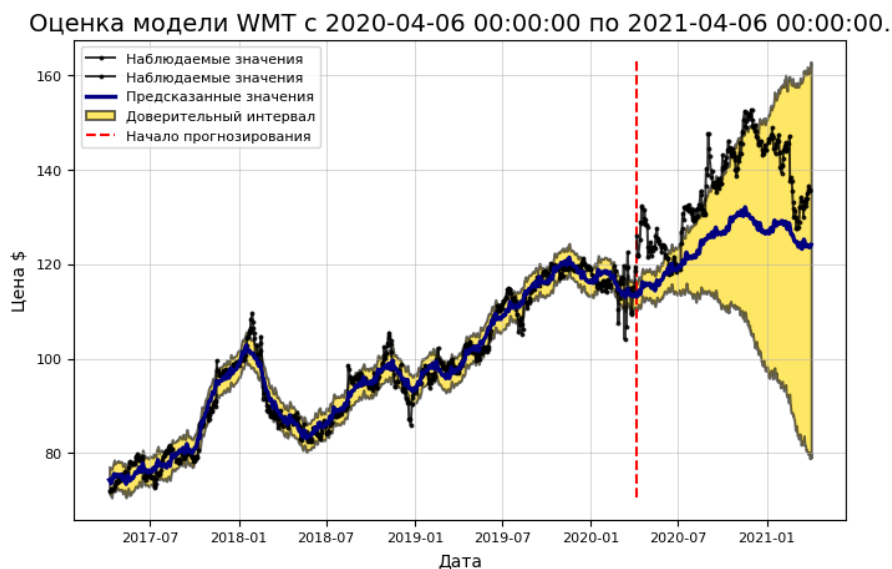


Рисунок 17 – Прогноз изменения стоимости актива WMT на 1 год вперед.

Фактически, если мы применим описанный метод к реальным транзакциям, мы будем обучать новую модель каждый день и делать прогнозы не более чем за один день.

2.4 Архитектура web-сервиса и web-моделирование

Приведенная на рисунке 18 схема показывает архитектуру веб-сервиса для прогнозирования котировок финансовых активов.



Рисунок 18 – Архитектура веб-сервиса для прогнозирования временных рядов

Код, написанный на языке python, выполняется на сервере Google colab. Программа запрашивает исторические данные по биржевым котировкам на сервисе Quandl. Эти запросы выполняются через так называемые удаленные вызовы процедур. Удаленные вызовы процедур (RPC) — это вызовы методов, которые размещаются в соответствующей веб-службе [16].

При использовании данной архитектуры веб-сервиса пользователь через браузер видит изображение с прогнозом цены выбранного финансового актива.

Приведем моделирование работы облачного сервера. У каждого приложения работающего в облаке есть уникальная web-ссылка. Открыв такую ссылку через браузер, пользователь получает доступ к интерфейсу приложения. Обмен данными между пользователем и сервером облачных вычислений осуществляется через стандартные протоколы – HTTP и Websockets. Работа сервера облачных вычислений строится на взаимодействии трех элементов, схема показана на рисунке 19:

- программного кода (ipynb-файла), хранящего в себе инструкции для вычислений и описание графического интерфейса;
- вычислительного ядра (kernel), контролирующего аппаратные узлы, участвующие в вычислениях, инициируемых приложением;
- сервера управления вычислениями, согласующего работу между пользователем, программным кодом облачного приложения и вычислительным ядром (kernel).

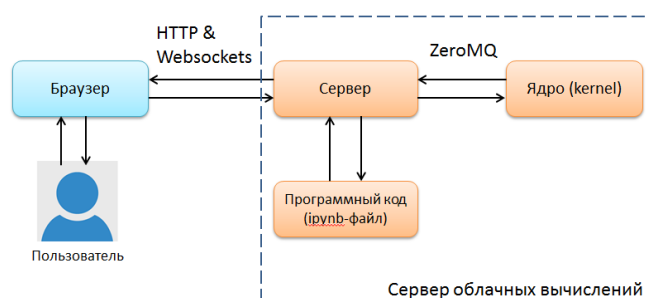


Рисунок 19 – Моделирование работы сервера облачных вычислений

Взаимодействие сервера управления с вычислительным ядром осуществляется с использованием протокола асинхронного обмена сообщениями ZeroMQ, ориентированного на использование в распределённых и параллельных вычислениях [24][25]. Результаты вычислений транслируются пользователю облачного приложения через браузер.

Выводы по второй главе

В качестве анализируемого временного ряда предложено использовать набор значений, показывающих среднюю цену актива в течение каждого дня. В качестве прогнозной модели используется свёрточная нейронная сеть, состоящая из нескольких слоев свертки (convolution) и выборки (subsampling), а также многослойного персептрона (mlp). Формирование обучающей выборки обеспечивается за счет использования метода сканирующего окна смещающегося по историческим данным изменения цены актива слева на право. При этом ширина окна равна 120 значениям.

Разработана структура веб-сервиса для прогнозирования цены активов, которая включает в себя получение исторических данных из сервиса Quandl, обработке этих данных на облачном сервисе Colab и предоставление прогнозов по выбранному активу через браузер в виде изображений.

Показаны основные компоненты сервера облачных вычислений, которые задействуются в разрабатываемом приложении для анализа финансовых активов. Описаны протоколы, используемые для обмена данными между пользователем и сервером управления вычислениями, а также между сервером управления вычислениями и вычислительным ядром.

Глава 3 Разработка программного обеспечения и тестирование

3.1 Программное обеспечение для прогнозирования котировок

При выполнении исследований было разработано программное обеспечение, разработанное на языке Python и функционирующее на сервере облачных вычислений. Приложение не использует базы данных. Разработанное приложение реализует следующий функционал:

- загрузка исторических данных по изменению цены активов из сервиса Quandl;
- формирование на основе загруженных исторических данных обучающей выборки
- использования обучающей выборки для обучения сверточной нейронной сети;
- тестирования обученной нейронной сети на исторических данных для определения доверительного интервала прогнозов;
- прогнозирование изменения цены выбранного актива на 4 месяца вперед;
- визуализация фактических (исторических) данных, смоделированных нейронной сетью данных и доверительного интервала посредством графика.

Разработанное программное обеспечение обладает удобным и простым графическим интерфейсом, который позволяет:

- задавать API-ключи доступа к данным с сервиса Quandl (это необходимо так как доступ к криптовалютным котировкам предоставляется только на платной основе);
- задавать тикер анализируемого актива (тикер – это зарегистрированное короткое обозначение актива);
- инициировать запуск анализа данных.

При разработке приложения использовался язык программирования Python. Интерфейс приложения основан на использовании виджетов.

Графический интерфейс пользователя разработанного приложения показан на рисунке 20. Интерфейс логически поделен на 2 блока. Первый блок «API-ключи для доступа к данным Quandl» предназначен для настройки подключений к сервису-поставщику данных. Второй блок «Выбор тикера актива» предназначен для выбора анализируемого актива и визуализации результатов анализ в виде графика.

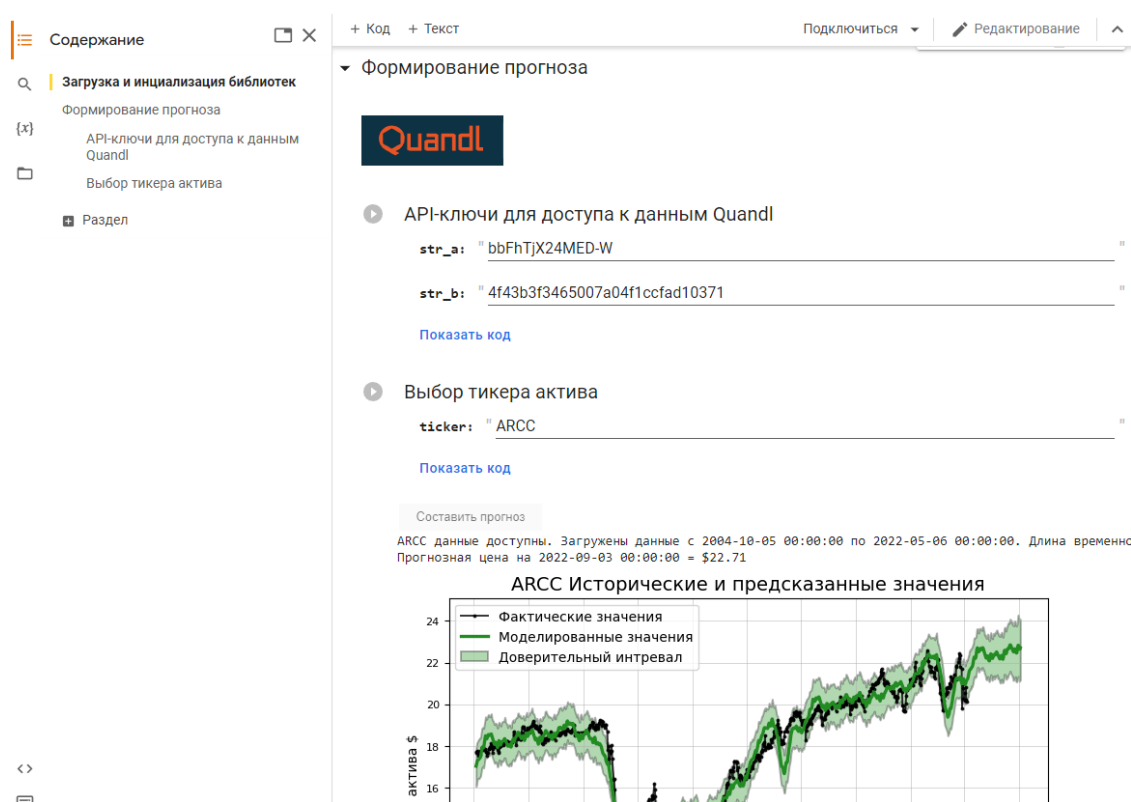


Рисунок 20 – Графический интерфейс приложения

Рассмотрим некоторые особенности программного кода. Для работы с данными от сервиса Quandl необходимо подключить к проекту одноименную библиотеку. В файле `sroker.py` содержатся вспомогательные функции, отвечающие за моделирование нейронной сети, поэтому этот файл также требуется подключить к проекту с помощью инструкции `import`. Для

построения графиков используется метод `plt()` реализация которого содержится в библиотеке `matplotlib` (рисунок 21).

```
#!/rm -rf Data-Analysis-master
!unzip gdrive/MyDrive/Colab\ Notebooks/Data-Analysis-tiingo.zip
!pip install --upgrade quandl
!pip install pytrends
%cd Data-Analysis-stocker_add_tiingo
%cd stocker
!rm stocker.py
%cd content/
import shutil
shutil.copy('/content/gdrive/MyDrive/Colab Notebooks/stocker_ru/stocker.py', "/content/Data-A
from stocker import Stocker
import matplotlib.pyplot as plt
import quandl
def nodatafig(ticker, type_t, num):
    fig, ax = plt.subplots(1, 1)

    ax.text(0.04, 0.43, 'Не доступно с бесплатным api-ключом', transform=ax.transAxes, rotation
plt.legend(loc = 2, prop={'size': 10}); plt.xlabel('Дата'); plt.ylabel('Цена актива $');
plt.grid(linewidth=0.6, alpha = 0.6)
plt.title(ticker + ' Исторические и предсказанные значения');
plt.show()
fig.savefig("test.png", dpi=fig.dpi, bbox_inches='tight')
return True
```

Рисунок 21 – Программный код с подключением требуемых библиотек и функцией обработки ошибки API-ключа

Одной из важных особенностей сервиса Quandl является возможность покупки доступа к историческим данным всех имеющихся в мире финансовых активов (криптовалют, акций, облигаций, etf-фондов, фьючерсов, опционов, фиатных валют). С бесплатным ключом Quandl предоставляет доступ к данным только 500 активов. Чтобы пользователь знал, что прогнозирование данных невозможно из-за используемого типа API-ключа была разработана функция `nodatafig()`, которая выводит соответствующее сообщение на экран пользователя (рисунок 22).

WARNING:matplotlib.legend.No handles with labels found to put in legend.
Error Retrieving Data.
(Status 404) (Quandl Error QECX02) You have submitted an incorrect Dataset code. Please check

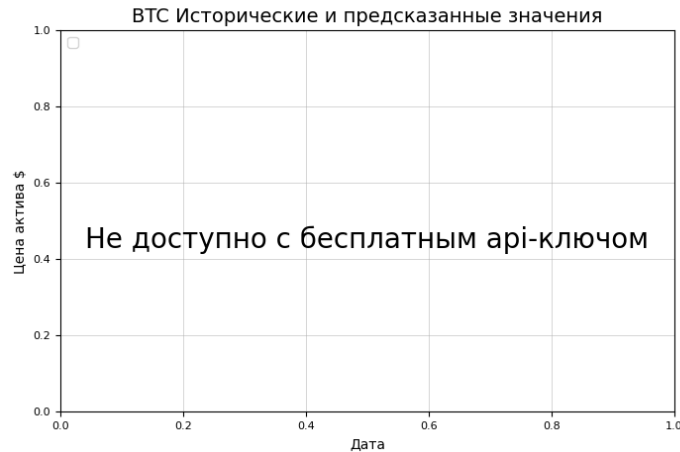


Рисунок 22 – Предупреждение о невозможности получить данные по активу с использованием бесплатного API-ключа

Для построения графика с прогнозом реализована кнопка «составить прогноз» при нажатии на которую для выбранного актива производится сбор данных из сервиса Quandl, обучение нейронной сети и моделирование поведения временного ряда. Программный код, вызываемой при нажатии кнопки показан на рисунке 23.

```
#@title Выбор тикера актива
ticker = 'MO' #@param {type:"string"}

import ipywidgets as widgets
from IPython.display import display
button = widgets.Button(description="Составить прогноз")
output = widgets.Output()

def on_button_clicked(b):
    try:
        share=Stocker(ticker, str_a="Hq[REDACTED]", str_b="17d6[REDACTED]a04f1ccfad10:
        model, model_data = share.create_prophet_model(days=120, saveimg="test.png")
        #share.predict_future(days=30)
    except AttributeError:
        nodatafig(ticker = ticker, type_t = "d", num = 1)
        print (i+ ' тикер не поддерживается')

button.on_click(on_button_clicked)
display(button, output)
```

Рисунок 23 – Программный код для запуска анализа и прогнозирования котировок по выбранному активу

Разработано программное обеспечения для анализа прогноза котировок финансовых активов. Web-приложение работает в облаке и доступ к нему осуществляется по ссылке через браузер. Теперь рассмотрим примеры использования приложения.

3.2 Апробация программного обеспечения

Для апробации программного обеспечения были выбраны два актива с тикерами ARCC и MO. Результат прогноза по изменению актива с тикером ARCC показан на рисунке 24. Прогноз для актива с тикером MO показан на рисунке 25.

На рисунках черным цветом отмечены фактические (исторические) значения цены, зеленая линия – это модельные данные, выдаваемые нейронной сетью, светло-зеленой областью обозначен доверительный интервал.

ARCC данные доступны. Загружены данные с 2004-10-05 00:00:00 по 2022-05-06 00:00:00. Длина
Прогнозная цена на 2022-09-03 00:00:00 = \$22.71



Рисунок 24 – Прогноз для актива с тикером ARCC

МО данные доступны. Загружены данные с 1970-01-02 00:00:00 по 2022-05-06 00:00:00. Длина :
Прогнозная цена на 2022-09-03 00:00:00 = \$56.78



Рисунок 25 – Прогноз для актива с тикером ARCC

На рисунке 24 видно, что нейронная сеть прогнозирует рост актива с текущего уровня в 20\$ до значения в диапазоне от 21 до 24\$. Такой актив можно покупать с целью его продажи по более высокой цене. На рисунке 25 видно, что с течением времени доверительный интервал очень сильно расходится к диапазону 51-62\$. С учетом того, что текущая цена актива около 56\$, покупать его на текущих уровнях рискованно.

Выводы по третьей главе

Разработано и протестировано программное обеспечение, реализующее технологию прогнозирования цен на активы, основанную на анализе временных рядов с помощью свёрточной нейронной сети. Показаны примеры работы программного обеспечения на данных финансовых активов с тикерами МО и ARCC.

Заключение

Анализ литературных источников показал, что существуют исследования по применению нейронных сетей для прогнозирования котировок финансовых активов, торгуемых на биржах.

Так же установлено, что перспективным направлением развития данного направления является применения сверточных нейронных сетей для решения задачи прогнозирования, так как использование такого типа сетей позволяет избежать необходимости ручного поиска значимых признаков в числовых значениях котировок (нейронная сеть выполняет этот шаг самостоятельно в процессе обучения).

Задача прогнозирования котировок финансовых активов представлена в виде анализа данных временного ряда.

При этом прогнозирование будущих значений котировок основано на анализе предыдущих исторических значений. В общем случае, требуется построить регрессионную модель, которая максимально точно описывала бы исторические данные котировок.

Прогнозирование основано на экстраполяции модели на будущие значения котировок.

В качестве регрессионной модели было предложено использовать сверточную нейронную сеть.

В качестве анализируемого временного ряда предложено использовать набор значений, показывающих среднюю цену актива в течение каждого дня.

В качестве прогнозной модели используется свёрточная нейронная сеть, состоящая из нескольких слоев свертки (convolution) и выборки (subsampling), а также многослойного персептрона (mlp).

Формирование обучающей выборки обеспечивается за счет использования метода сканирующего окна смещающегося по историческим данным изменения цены актива слева на право. При этом ширина окна равна 120 значениям.

Разработана структура веб-сервиса для прогнозирования цены активов, которая включает в себя получение исторических данных из сервиса Quandl, обработке этих данных на облачном сервисе Colab и предоставление прогнозов по выбранному активу через браузер в виде изображений.

Разработано и протестировано программное обеспечение, реализующее технологию прогнозирования цен на активы, основанную на анализе временных рядов с помощью свёрточной нейронной сети.

Показаны примеры работы программного обеспечения на мистических данных финансовых активов с тикерами MO и ARCC.

Результаты ВКР представляют научно-практический интерес и могут быть рекомендованы веб-дизайнерам и аналитикам, занимающимся разработкой программного обеспечения для финансовой сферы.

Список используемой литературы

1. Албегов, М.М. Краткосрочное прогнозирование регионального развития в условиях неполной информации / М.М. Албегов. - Москва: Наука, 2016. - 335 с.
2. Афанасьев, В. Н. Анализ временных рядов и прогнозирование / В.Н. Афанасьев, М.М. Юзбашев. - М.: Финансы и статистика, Инфра-М, 2016. - 320 с.
3. Бокс, Дж. Анализ временных рядов прогноз и управление (часть 2) / Дж. Бокс, Г. Дженкинс. - М.: [не указано], 2018. - 688 с.
4. Бокс, Дж. Анализ временных рядов прогноз и управление. Выпуск 1 / Дж. Бокс, Г. Дженкинс. - М.: Мир, 2016. - 408 с.
5. Конкина, Вера Анализ временных рядов и прогнозирование затрат / Вера Конкина. - М.: Palmarium Academic Publishing, 2019. - 136 с.
6. Лефевр Воспоминания биржевого спекулянта / Лефевр, Эдвин. - М.: ЗАО Олимп-Бизнес, 2017. - 544 с.
7. Лукашин, Ю. П. Адаптивные методы краткосрочного прогнозирования временных рядов / Ю.П. Лукашин. - М.: Финансы и статистика, 2018. - 416 с.
8. Пивоваров С.Э. Методология комплексного прогнозирования развития отрасли / С.Э. Пивоваров. - М.: Наука, 2019. - 193 с.
9. Подкорытова О.А. Анализ временных рядов 2-е изд., пер. и доп. Учебное пособие для бакалавриата и магистратуры - М.: Юрайт, 2017. - 713 с.
10. Подкорытова, О. А. Анализ временных рядов. Учебное пособие / О.А. Подкорытова, М.В. Соколов. - М.: Юрайт, 2016. - 266 с.
11. Садовникова, Н. А. Анализ временных рядов и прогнозирование. Учебник / Н.А. Садовникова, Р.А. Шмойлова. - Москва: ИЛ, 2016. - 152 с.
12. Смиттен Жизнь и смерть величайшего биржевого спекулянта / Смиттен, Ричард. - М.: Омега-Л, 2018. - 384 с.

13. Тарп, Ван; Бартон Д.Р. Биржевые стратегии. Игры без риска / Тарп, Ван; Д.Р. Бартон, С. Сьюггеруд. - М.: СПб: Питер, 2018. - 400 с.
14. Тихомиров, Н.П. Демография. Методы анализа и прогнозирования / Н.П. Тихомиров. - М.: Экзамен, 2019. - 256 с.
15. Филин В.В., Леденева Т.М. Нейросетевое прогнозирование курса акций с применением методов нелинейной динамики // Вестник Воронежского государственного технического университета. - 2012. - №4. - С. 57-60.
16. Шмитт, Кристофер CSS. Рецепты программирования / Кристофер Шмитт. - М.: Русская Редакция, БХВ-Петербург, 2017. - 672 с.
17. Юзбашев, М.М. Анализ временных рядов и прогнозирование. Учебник. Гриф УМО МО РФ / М.М. Юзбашев. - М.: Финансы и статистика, 2019. - 592 с.
18. Chollet F. Deep Learning with Python – Manning Publications, 2017. – 384 p.
19. Geron Au. Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems Second Edition (Third Release) – O'Reilly Media, 2019. – 856 p.
20. Kong, X. Principal Component Analysis Networks and Algorithms / Xiangyu Kong, Changhua Hu, Zhansheng Duan. – Springer Singapore, 2017 – 323 p.
21. Liu, J. Radial Basis Function (RBF) Neural Network Control for Mechanical Systems: Design, Analysis and Matlab Simulation. – Springer Berlin Heidelberg, 2014 – 365 p.
22. Min, F. A Competition Strategy to Cost-Sensitive Decision Trees / Fan Min, William Zhu // International Conference on Rough Sets and Knowledge Technology – 7th International Conference, RSKT 2012, Chengdu, China, August 17-20, 2012. Proceedings: Rough Sets and Knowledge Technology. – Springer-Verlag Berlin Heidelberg 2012. – pp. 359-368.
23. Silva, I.N. Artificial Neural Networks: A Practicle course / Ivan Nunes da Silva, Danilo Hernane Spatti, Rogerio Andrade Flauzino, Luisa Helena Bartocci

Liboni, Silas Franco dos Reis Alves. – Springer International Publishing, 2017 – 307 p.

24. OpenMP 4.0 API C/C++ Syntax Quick Reference Card [Электронный ресурс]: Режим доступа: <http://www.openmp.org>

25. OpenMP Application Program Interface. Examples. Version 4.0.1 February 2014 [Электронный ресурс]: Режим доступа: <http://www.openmp.org>