

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ

ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра теории функций и стохастического анализа

**РАЗРАБОТКА СИСТЕМЫ ПРОГНОЗИРОВАНИЯ  
БИРЖЕВЫХ КОТИРОВОК НА ОСНОВЕ МАШИННОГО  
ОБУЧЕНИЯ**

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студента 2 курса 248 группы

направления 09.04.03 — Прикладная информатика

механико-математического факультета

Смятского Дмитрия Александровича

Научный руководитель

доцент, к. ф.-м. н., доцент

\_\_\_\_\_

Л. В. Борисова

Заведующий кафедрой

д. ф.-м. н., доцент

\_\_\_\_\_

С. П. Сидоров

Саратов 2024

## Введение

Инвестирование набирает популярность с каждым годом. По итогам четвертого квартала 2022 года 29,2 миллиона человек или 38% экономически активного населения в России находятся на брокерском обслуживании, что в 1,45 раза больше, чем годом ранее. К существенным причинам роста популярности можно отнести доступность инструментов инвестирования, возможность открытия индивидуальных инвестиционных (ИИС) и брокерских счетов. Сегодня многие российские банки предлагают своим клиентам инвестиционные продукты, мобильные приложения для управления инвестированием, а также проводят широкую рекламную кампанию, предлагая всевозможные учебные курсы по управлению личным капиталом.

Дополнительной значимой причиной для роста популярности инвестирования являются налоговые льготы, которые могут получить владельцы индивидуальных инвестиционных счетов. Так, например, с помощью ИИС за год можно получить налоговый вычет в размере 52 тысяч рублей, если соблюдены все необходимые условия. Таким образом, государство поощряет граждан, которые инвестируют личные средства в ценные бумаги.

Люди обращают внимание на рынок ценных бумаг в поисках большей прибыли: он предлагает более гибкие и доходные возможности управления деньгами. Здесь можно выбрать ценные бумаги, а также инвестиционные фонды, драгоценные металлы и валюту в зависимости от желаемого уровня доходности и приемлемого уровня риска.

В такой ситуации актуальность приобретает вопрос о механизме принятия решений о входе и выходе с рынка, ведь данный механизм является залогом получения прибыли при торговле.

Существует два основных подхода к решению данной проблемы — фундаментальный анализ, который строится на субъективном восприятии рыночных и производственных факторов предприятия, и технический анализ, который строится на основе поиска математических закономерностей поведения цены актива в прошлом и позволяет на их основе прогнозировать цены в будущем.

Однако, технический анализ также не лишен субъективности, так как разные эксперты могут прийти к разным выводам при одних и тех же исход-

ных данных. Поскольку в основе технического анализа лежит именно прогнозирование, а прогноз никогда не бывает абсолютно верным, то и ошибка может быть значительной, что непременно ведет к риску получить убытки при торговле.

В связи с этим актуальна задача разработки моделей на основе машинного обучения, которые были бы способны обрабатывать большой объем исторических цен актива и строить на его основе максимально возможный по точности прогноз на будущий период торгов, минимизируя ошибки и исключая субъективное человеческое восприятие.

Целью магистерской работы является создание системы (програмного обеспечения), позволяющего прогнозировать будущие котировки ценных бумаг, используя методы машинного обучения. В рамках задач данной работы предстоит:

1. Изучить особенности трейдинга на рынке ценных бумаг и смежные с этим процессом понятия.
2. Исследовать машинное обучение как метод прогнозирования цены акций.
3. Реализовать программу на языке Python, позволяющую строить прогнозы, используя машинное обучение.
4. Разработать графический интерфейс для более удобного взаимодействия с модулями программы.
5. Построить прогнозы по нескольким различным ценным бумагам с использованием двух моделей машинного обучения и провести сравнительный анализ результатов.

Результатом работы станет приложение, позволяющее трейдеру загружать и визуализировать котировки ценных бумаг, настраивать параметры и строить модели машинного обучения, а так же сравнивать результативность ранее построенных моделей.

### **Трейдинг на рынке ценных бумаг**

Предметная область данной работы затрагивает сферу трейдинга на рынке ценных бумаг и включает в себя изучение базовых понятий и принципов, связанных с этой деятельностью.

Рынок ценных бумаг представляет собой совокупность экономических

отношений по поводу их выпуска и обращения между участниками. Фактически рынок ценных бумаг можно разделить на два подвида: первичный и вторичный. Под первичным рынком принято понимать отношения между эмитентом и первым покупателем с целью привлечения инвестиционных ресурсов. Соответственно на вторичном рынке представлены ранее выпущенные ценные бумаги или бумаги, представленные для гашения. Фондовый рынок — это место, где происходит торговля акциями, облигациями, валютами и прочими активами.

Важной задачей вторичного рынка ценных бумаг остается предоставление инвестору возможности возврата вложенных в ценные бумаги денежных средств путем их продажи без существенной потери в цене. Без развитого вторичного рынка, обеспечивающего постоянную ликвидность и распределение фондовых рисков, не может эффективно существовать первичный рынок ценных бумаг. Одна из главных функций вторичного рынка — установление реальной рыночной цены отдельных ценных бумаг.

Трейдинг представляет собой процесс торговли акциями, облигациями и другими ценными бумагами на фондовом рынке с целью извлечения прибыли. Основная задача трейдера заключается в анализе рыночной ситуации, прогнозировании динамики цен на активы и принятии решений о покупке или продаже ценных бумаг.

Для успешного трейдинга важно обладать знаниями о структуре и особенностях рынка ценных бумаг, понимать различия между акциями и облигациями, разбираться в основных финансовых инструментах и их влиянии на стоимость активов. Также необходимо изучить основы технического и фундаментального анализа, чтобы иметь возможность принимать обоснованные решения о входе и выходе из сделок.

### **Акции**

Акция — это эмиссионная ценная бумага, закрепляющая права ее владельца (акционера) на получение части прибыли акционерного общества в виде дивидендов, а также на участие в управлении акционерным обществом и на часть имущества, остающегося после ликвидации акционерного общества.

Фактически, акции представляют собой право собственности — свиде-

тельство о владении частью компании эмитента с правом участия в управлении и получение прибыли. Немаловажным фактом является то, что владение акциями сопряжено с высоким уровнем риска потери инвестиций в полном объеме в случае финансового краха компании. Основными источниками доходов от владения акциями являются дивиденды.

### **Облигации**

Облигация — это ценная бумага, удостоверяющая право владельца требовать ее погашения, то есть выплаты номинальной стоимости или номинальной стоимости и процентов, в установленные сроки. При этом условия и сроки погашения оговариваются в решении о выпуске облигаций.

Номинальной стоимостью облигации (номинал облигации) является стоимость облигации, установленная при ее выпуске. Датой погашения облигации является календарный день, когда владельцу облигации будут возвращены денежные средства в размере номинальной стоимости облигации, умноженной на количество приобретенных облигаций.

Инвестирование в облигации является альтернативой банковским вкладам и покупке акций. При умеренных рисках покупатель ценных бумаг получает стабильный источник дохода и относительно высокую процентную ставку. Это сочетание объясняет популярность облигаций с выплатами в рублях и иностранной валюте.

### **Анализ динамики цен**

На сегодняшний день существует три основных метода анализа динамики цен на рынке: интуитивный подход, фундаментальный анализ и технический анализ. Интуитивно-понятный метод походит для профессиональных биржевых трейдеров с большим опытом. Фундаментальный анализ дает возможности изучения ценовых движений на макроэкономическом уровне. Технический анализ предоставляет методы прогнозирования цены ценных бумаг, основанный на математических расчетах.

Особое внимание инвесторов, как правило, уделяется именно техническому анализу. Одним из преимуществ метода является возможность его использования практически в любых условиях. Принципы технического анализа можно применять при торговле любым активом. В то время как фундаментальные аналитики специализируются на одном товаре. Классическими

инструментами технического анализа являются графики и индикаторы.

Индикаторы технического анализа представляют собой большую категорию инструментов, каждый из которых имеет свои преимущества и недостатки. Каждый тип индикатора имеет свои особенности: например, трендовые индикаторы часто запаздывают, но их сигналы более надежны, чем сигналы осцилляторов, которые считаются ключевыми. У большинства индикаторов есть ключевое преимущество — благодаря автоматическому расчету они значительно упрощают задачу трейдерам, давая четкие и расплывчатые сигналы. Некоторые индикаторы вообще освобождают трейдера от задач аналитика и все, что от него требуется, это вовремя нажать «купить» или «продать».

Альтернативой классическим индикаторам может служить анализ динамики цен с использованием нейронных сетей и ARIMA моделей.

Существуют различные типы нейронных сетей, которые могут использоваться для выполнения технического анализа. Например, искусственные нейронные сети используются для прогнозирования цен на активы и определения оптимальных моментов для входа и выхода из сделок. Рекуррентные нейронные сети полезны для анализа временных рядов и могут использоваться для обнаружения трендов и циклов в данных.

ARIMA модель позволяет анализировать временные ряды и прогнозировать будущие значения. Она состоит из трех основных компонентов: авторегрессии, интегрирования и скользящего среднего. Авторегрессия учитывает зависимость текущего значения от предыдущих, интегрирование позволяет учитывать тренд и сезонные эффекты, а скользящее среднее используется для сглаживания данных. ARIMA может быть полезна для прогнозирования цен на акции, валютные пары и другие финансовые инструменты.

Однако, важно помнить, что технический анализ не является заменой фундаментального анализа и должен использоваться в сочетании с другими инструментами для принятия инвестиционных решений.

### **Машинное обучение**

Машинное обучение — это область, в которой используются алгоритмы и методы, позволяющие компьютерным системам учиться на данных и делать прогнозы или выдавать решения. Машинное обучение — это разновидность

искусственного интеллекта.

### **Глубокое обучение**

Глубокое обучение (англ. «Deep Learning») — это форма машинного обучения, ориентированная на многослойные (глубокие) нейронные сети, вдохновленные архитектурой человеческого мозга, в частности, биологическими нейронами, используемыми для обработки информации. Глубокие нейронные сети изучают желаемую задачу самостоятельно, без необходимости программисту заранее указывать какие-либо необходимые вводные данные.

Решаемая задача может быть кластеризацией, классификацией или регрессией. Классификация предполагает использование наблюдений в качестве входных данных и отнесение их к определенному классу из проблемной области. По сути, нейронная сеть глубокого обучения — это математическая функция, которая принимает вектор в качестве входных данных (наблюдения) и выдает вектор в качестве выходных данных.

### **Модель авторегрессии скользящего среднего (ARIMA)**

ARIMA модели могут быть полезны при прогнозировании цен на биржевые инструменты, поскольку они позволяют учитывать автокорреляцию и сезонные эффекты во временных рядах цен. Однако для получения точных прогнозов необходимо использовать качественные данные и правильно выбрать параметры модели. В машинном обучении ARIMA обычно относится к классу статистических моделей, которые предоставляют выходные данные, линейно зависящие от их предыдущих значений с добавлением случайных факторов.

ARIMA является мощным инструментом для прогнозирования котировок ценных бумаг, поскольку она способна учитывать различные характеристики временного ряда. Это позволяет моделировать сложные зависимости между текущими и прошлыми значениями и предсказывать будущие котировки с высокой степенью точности. Предварительная обработка данных и настройка параметров модели, таких как порядки авторегрессии, интегрирования и скользящего среднего, играют критическую роль в получении достоверных прогнозов с использованием ARIMA.

### **Программная реализация системы**

#### **Используемые технологии**

Разработка системы будет вестись с помощью языка программирования Python и следующего набора библиотек:

- pandas,
- numpy,
- matplotlib,
- sklearn,
- tensorflow,
- keras,
- statsmodels,
- tinkoff-investments,
- pysimplegui.

Python — высокоуровневый язык программирования общего назначения с динамической строгой типизацией и автоматическим управлением памятью, ориентированный на повышение производительности разработчика, читаемости кода и его качества, а также на обеспечение переносимости написанных на нём программ. Синтаксис ядра языка минималистичен, за счёт чего на практике редко возникает необходимость обращаться к документации. Сам же язык известен как интерпретируемый и используется в том числе для написания скриптов.

### **Входной набор данных**

Для построения моделей машинного обучения необходим входной набор данных, в данном случае — исторические цены биржевых активов. Источником такой информации могут служить различные торговые интернет-площадки, одна из которых — сервис инвестиций от банка Тинькофф. API Тинькофф Инвестиций способен предоставлять информацию о стоимости активов за указанный период бесплатно. Для использования этого сервиса требуется быть клиентом Тинькофф Банка и иметь активный брокерский счет, после выполнения этих условий можно обращаться к API сервиса с помощью библиотеки tinkoff-investments, используя для авторизации приватный токен пользователя.

Для получения набора данных был реализован метод, позволяющий запрашивать исторические цены актива за указанный период у API Тинькофф Инвестиций и сохранять полученный ответ в CSV файл, который можно мно-



гократно использовать при построении моделей машинного обучения.

### **Нейронная сеть LSTM**

Для построения нейронной сети LSTM был реализован одноименный класс, который содержит все необходимые методы для конфигурации, обучения и тестирования нейросетевой модели. Также были реализованы методы, позволяющие получать прогнозные цены акции на будущий период.

### **ARIMA**

Для построения ARIMA моделей был реализован специальный класс, содержащий необходимые методы для построения, обучения и тестирования модели. Дополнительно был реализован сеточный поиск наилучших параметров модели. Аналогично нейросетевым LSTM моделям, ARIMA позволяет получать прогнозные котировки ценных бумаг на будущий период.

### **Разработка графического интерфейса**

Для упрощения процесса работы с разными модулями программы был реализован графический интерфейс с помощью библиотеки PySimpleGUI. Библиотека использует механизм разметки (англ. «layout»), чтобы определить и структурировать пользовательский интерфейс. Набор вкладок для единственного окна разрабатываемого приложения описан в файле

`layouts.py` и содержит:

1. `layout_datasets` — вкладка с функционалом получения котировок ценных бумаг в формате CSV;
2. `layout_lstm` — вкладка предоставляет возможность задавать параметры и строить модель LSTM;
3. `layout_arima` — вкладка предоставляет возможность задавать параметры и строить модель ARIMA;
4. `layout_results` — вкладка для анализа и сравнения результатов, полученных при построении моделей.

На рисунке 1 представлен прототип графического интерфейса на примере вкладки «LSTM».

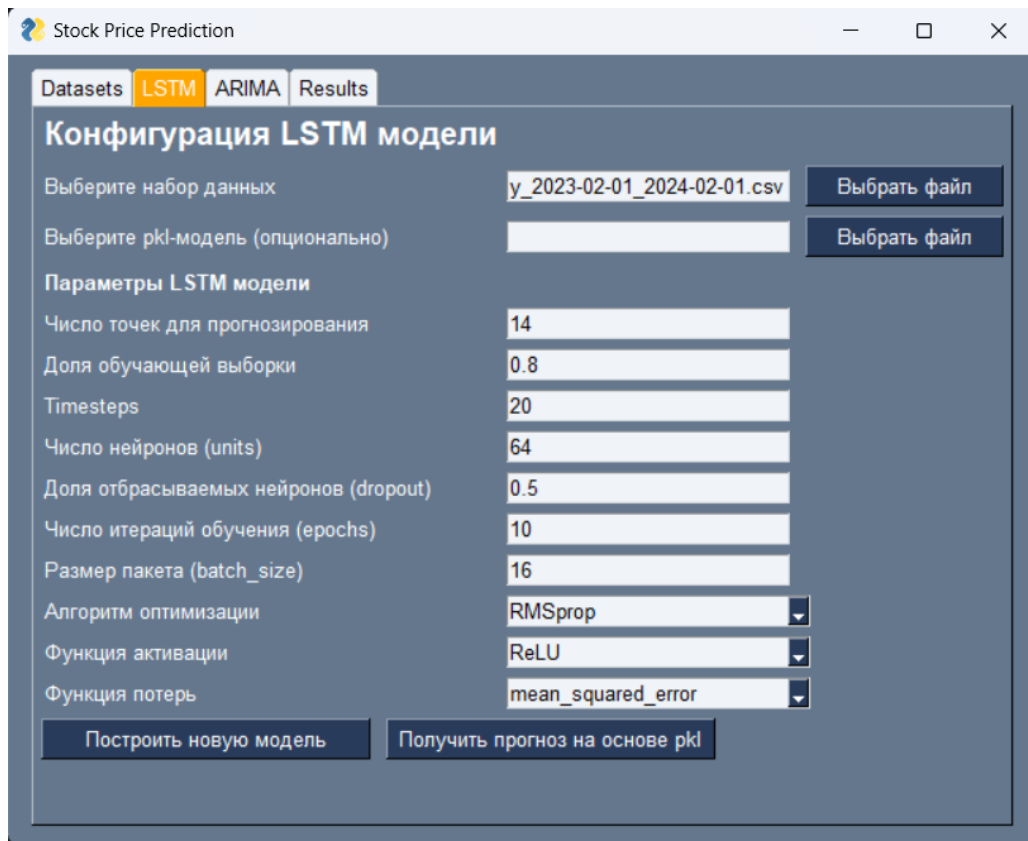


Рисунок 1 – Вкладка «LSTM»

## Сравнительный анализ LSTM и ARIMA прогнозов

После реализации графического интерфейса системы был проведен сравнительный анализ точности прогнозов, полученных с помощью ARIMA и LSTM моделей. Для построения прогнозов были выбраны 5 акций крупных российских компаний:

1. Газпром (GAZP),
2. Лукойл (LKOH),
3. Ростелеком (RTKM),
4. Сбербанк (SBER),
5. Яндекс (YNDX).

Для моделирования использованы исторические цены выбранных акций за каждый день в период с 1 февраля 2023 по 1 февраля 2024 года. Для каждого финансового инструмента были подобраны оптимальные параметры модели, выполнена ее тренировка и тестирование с целью получения результатов. Для сравнительной оценки качества моделей использован показатель RMSE (среднеквадратичная ошибка).

В результате, средняя величина RMSE для моделей ARIMA и LSTM оказалась сопоставимой, что свидетельствует о примерно одинаковой точности и эффективности прогнозирования цен на акции с использованием обеих методик. Однако, учитывая специфику применения каждой из моделей, могут быть ситуации, в которых одна из них будет более предпочтительной. Например, ARIMA может быть более подходящей для анализа временных рядов с ярко выраженными сезонными эффектами, в то время как LSTM может обеспечить более точные прогнозы в условиях сложной нелинейной динамики цен. Таким образом, для принятия окончательного решения о выборе модели необходимо учитывать специфические особенности каждого конкретного случая.

### **Заключение**

В ходе выполнения работы было разработано приложение, позволяющее строить модели машинного обучения на основе нейронных сетей и ARIMA моделей. Помимо основного функционала, был реализован и дополнительный — приложение предоставляет графический интерфейс, с использованием которого пользователь может загрузить котировки ценных бумаг, визуализировать их, задать параметры и построить модели машинного обучения, после получения прогноза все результаты сохраняются на отдельную вкладку, где они доступны для будущего анализа и сравнения.

С использованием разработанного приложения был произведен анализ динамики цен по 5 различным российским ценным бумагам. На основе полученных результатов, была дана сравнительная характеристика для нейронных сетей и ARIMA моделей по точности их прогнозов.

Работа прошла апробацию на различных конференциях, в частности, на ежегодной студенческой конференции «Актуальные проблемы математики и механики», которую проводил механико-математический факультет СГУ в апреле 2024 года в секции «Анализ данных», в XII Международной научно-практической конференции «Математическое и компьютерное моделирование в экономике, страховании и управлении рисками», ноябрь 2023 года.

Таким образом, все поставленные задачи выполнены, следовательно, цель данной работы успешно достигнута.