

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра математической кибернетики и компьютерных наук

**ПРОГНОЗИРОВАНИЕ РЫНКА КРИПТОВАЛЮТ С ПОМОЩЬЮ  
НЕЙРОННЫХ СЕТЕЙ**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 411 группы

направления 02.03.02 — Фундаментальная информатика и информационные  
технологии

факультета КНиИТ

Патрикеева Максима Владимировича

Научный руководитель

доцент, к. ф.-м. н.

\_\_\_\_\_

А. С. Иванова

Заведующий кафедрой

к. ф.-м. н., доцент

\_\_\_\_\_

С. В. Миронов

Саратов 2023

## **Введение**

Выпускная квалификационная работа посвящена теме прогнозирования цен на рынке с помощью нейронных сетей. Рынок - это сложная среда, где изменения происходят каждую секунду, очень динамическое и нестабильным место, где цены на товары и услуги постоянно колеблются. Однако прогнозирование цен на рынке имеет большое значение для предпринимателей и инвесторов, которые могут принимать обоснованные решения на основе предварительных прогнозов цен.

## **Актуальность**

Тема прогнозирования криптовалют с помощью нейронных сетей является одной из самых актуальных задач на данный момент. Благодаря своей способности анализировать и обрабатывать большие объемы данных, нейронные сети могут помочь в прогнозе цен на криптовалюты. Однако, важно помнить, что криптовалюты характеризуются высокой волатильностью и сложностями в предсказании. Нейронные сети могут быть полезным инструментом, но результаты прогнозов всегда следует анализировать с осторожностью и учитывать другие факторы, влияющие на цены криптовалют.

**Целью** моей выпускной квалификационной работы является построение модели для прогнозирования цен на рынке криптовалют с помощью нейронных сетей. Это исследование позволит понять, какие факторы влияют на цены на рынке, каким образом лучше всего применять модели, использующие нейронные сети для торговли на биржах, и какие методы прогнозирования являются наиболее эффективными в данном контексте. Данная работа поможет освоить новые знания о нейронных сетях и их применении на рынке.

Были сформированы следующие задачи:

- Сбор, анализ и обработка данных;
- Обучение моделей нейронных сетей;
- Оценка качества моделей с помощью различных метрик;

## **Структура бакалаврской работы**

Работа состоит введения, теоретической и практической частей, заключения и приложения.

- В теоретической части описывается предметная область, основные понятия в теме торговли криптовалютой. Далее обзревается существующие методы по прогнозированию рынка и применение нейронных сетей для

прогнозирования, описываются данные для анализа, а также приводится краткий обзор существующих публикаций на данную тематику. После описывается методология для построения модели: используемые алгоритмы и методы обработки данных, а также различные подходы к оценке качества.

- В практической части обзревается план исследования. Далее описывается подготовка данных для прогнозирования, включающая себя сбор и описание исходных данных, их предварительную обработку и отбор. После приводятся различные метрики для оценки качества моделей, ход исследования и анализ результатов.
- В приложении представлен исходный код программы

### **Основное содержание работы**

Одним из видов заработка на финансовом рынке является торговля на криптовалютном рынке. Торговля на криптовалютном рынке - это купля-продажа криптовалюты на специальных биржах. Криптовалюта - это вид цифровой валюты, который используется для совершения платежей и хранения активов. Главной криптовалютой на данный момент считается Биткоин, который работает на основе технологии блокчейн. сновная идея биткоина заключается в том, чтобы создать децентрализованную систему, в которой участники могут осуществлять прямые финансовые транзакции друг с другом без посредничества банков или других финансовых институтов. Это достигается благодаря технологии блокчейн, которая является распределенным реестром, записывающим все транзакции биткоина.

Для прогнозирования цен на рынке используются различные данные, включая экономические показатели, финансовые отчеты компаний, тренды в отрасли, информацию о геополитической обстановке и прочее. Также могут использоваться данные о покупательском спросе и предложении на определенные товары и услуги, инфляции, процентных ставках и валютных курсах. В целом можно выделить следующие группы данных:

- Фундаментальные показатели;
- Макроэкономические показатели;
- Технические индикаторы;
- Onchain-анализ для криптовалют;

Из анализа публикаций на тему прогнозирования рынка с помощью ней-

ронных сетей можно сделать выводы, о том, что наилучшими моделями являются сеть с эффектом памяти, которые могут работать с последовательностями данных такие, как: GRU, LSTM, CLSTM. Нельзя точно сказать какие наборы данных лучше всего подходят для предсказания цены криптовалют, хорошим способом будет создать правильно скомбинированный датасет, который включал бы в себя как макроэкономические показатели, так и технические индикаторы, так и фундаментальные признаки. Прогнозирование цен криптовалют, а также и акций с помощью нейронных сетей имеет множество факторов, изменение которых может оказать влияние на результат, что требует провести множество различных экспериментов.

Существует несколько алгоритмов и моделей нейронных сетей, которые могут быть использованы для предсказания цен на криптовалюты. Некоторые из них включают в себя:

1. LSTM (Long Short-Term Memory) - это вид рекуррентной нейронной сети (RNN), который разработан для эффективной обработки последовательностей данных, сохраняя долгосрочные зависимости. Он отличается от обычных RNN своей способностью запоминать информацию на протяжении длительного временного интервала и устранять проблему затухания и взрывного градиента.

2. GRU (Gated Recurrent Unit) - является типом рекуррентной нейронной сети (RNN), который разработан для эффективной обработки последовательных данных. GRU отличается от стандартных RNN и LSTM своей более простой структурой и способностью моделировать долгосрочные зависимости в данных.

3. BLSTM (Bidirectional Long Short-Term Memory) является вариантом рекуррентной нейронной сети (RNN), который комбинирует две LSTM (Long Short-Term Memory) сети для обработки последовательных данных. BLSTM применяется для учета контекста как в прошлом, так и в будущем и обеспечения более полного понимания последовательных данных.

4. Conv-LSTM (Convolutional Long Short-Term Memory) - это модификация стандартной LSTM (Long Short-Term Memory) архитектуры, которая включает сверточные слои. Conv-LSTM предназначена для обработки и анализа последовательных данных с пространственной структурой, таких как изображения или видео.

Для предсказания цены криптовалют на данный момент предоставлено множество показателей, поэтому очень важно произвести грамотный отбор при-

знаков. Существует множество методов по их отбор. Наиболее эффективные и популярные из них следующие:

- RFE - это метод отбора признаков, который используется в машинном обучении для определения наиболее важных признаков в наборе данных. Процесс RFE начинается с обучения модели машинного обучения на исходном наборе признаков. Затем модель оценивает важность каждого признака и удаляет наименее важный признак. После удаления признака модель повторно обучается на уменьшенном наборе признаков и процесс повторяется до тех пор, пока не будет достигнуто желаемое количество признаков;
- PCA - метод анализа данных и снижения размерности, который используется для выделения наиболее значимых признаков из многомерного набора данных. PCA выполняет линейное преобразование данных, чтобы получить новые некоррелированные переменные, называемые главными компонентами.

Основная идея PCA заключается в том, чтобы найти новые оси или направления в пространстве данных, по которым разброс данных максимален. Первая главная компонента (PC1) направлена вдоль оси наибольшего разброса данных, вторая главная компонента (PC2) - вдоль второй по величине оси разброса, и так далее. Главные компоненты являются ортогональными друг другу, что означает, что они не коррелированы;

- VIF - мера, используемая для оценки мультиколлинеарности между предикторами (независимыми переменными) в множественной линейной регрессии. Мультиколлинеарность возникает, когда существует высокая корреляция между предикторами, что может затруднить интерпретацию и оценку влияния каждого предиктора на зависимую переменную.

VIF вычисляется для каждого предиктора путем оценки соотношения между дисперсией коэффициента регрессии этого предиктора и дисперсией коэффициента регрессии, которая получается, если бы этот предиктор был независим от остальных предикторов;

Существует несколько подходов к оценке качества прогнозов на цены криптовалют:

1. Сравнение фактических и прогнозируемых значений: Данный подход подразумевает сравнение фактических значений цен с прогнозируемыми моде-

лю. По сравнению ошибок можно судить о качестве прогнозов.

2. Использование статистических метрик: Статистические метрики (например, среднее абсолютное отклонение, средняя квадратичная ошибка и коэффициент корреляции) могут быть использованы для оценки точности прогнозов.

3. Анализ доходности: Данный подход основывается на анализе доходности портфеля, сформированного на основе прогнозов цен. Если доходность превышает уровень доходности инвестиций в активы с низкой риском, тогда можно говорить об успешности прогнозов.

### **План исследования**

Прогнозирование рынка криптовалют с помощью нейронных сетей имеет огромное поле для исследований. В зависимости от конкретных целей и возможностей можно придумать множество подходов и путей решения. Основываясь на изученных материалах и исследованиях, описанных в теоретической части данной работы, были выдвинуты следующие тезисы, определяющие план исследования:

- Целевое значение - В зависимости от целевой переменной, модели будут иметь разные архитектуры, критерии оценки качества, а также разную обработку данных. Данная работа включает в себя предсказание двух меток: цена закрытия следующей свечи и бинарное предсказание, на то, что закроется ли следующая свеча выше ее открытия или ниже(задача бинарной классификации);
- Прогнозирование цены может происходить в разных разрезах времени. В зависимости от выбранного таймфрейма данные о ценах имеют разное количество исторических данных и разный характер изменения. Для исследования зависимости качества прогнозирования от интервала времени в работе используются данные в 3-х разных интервалах: 1 день, 1 час, 15 минут.
- Виды нейронных сетей - на данный момент множество видов нейронных сетей и их модификаций. Из анализа изученных публикаций и работ на тему использования нейронных сетей для прогнозирования рынка было выявлено, что наилучшим видом нейронных сетей для данной задачи являются рекуррентные нейронные сети (RNN) из-за учета последовательности данных, что позволяет им эффективнее работать с временными

рядами. В данном исследовании будут использоваться следующие слои RNN: LSTM, GRU, Bidirectional, Conv-LSTM, а также комбинации разных слоев.

- Задачи предсказания цены или вида свечи, относятся к задачам регрессии и бинарной классификации соответственно, поэтому для оценки качества модели были использованы разные виды метрик относительно этих задач. Несмотря на это, такие метрики не способны дать объективную оценку о прибыльности модели, вследствие чего было необходимо создать дополнительные метрики оценки способности модели приносить прибыль для пользователя.

Структура проекта будет содержать 6 файлов в зависимости от цели предсказания и таймфрейма данных:

- Предсказание цены биткоина - 15 минут;
- Предсказание цены биткоина - 1 час;
- Предсказание цены биткоина - 1 день;
- Предсказание направления цены биткоина - 15 минут;
- Предсказание направления цены - 1 час;
- Предсказание направления цены - 1 день;

### **Подготовка данных для прогнозирования**

Первым шагом практической части работы стало определение источника данных. На данный момент существует множество API-решение, предоставляющие самые различные данные и метрики для анализа и прогнозирования в реальном времени. В исследовании было использовано бесплатное API-решение биржи Cryptocompare. Данное API имеет гибкий и простой для понимания функционал для работы с ним. Для подключения к API биржи Cryptocompare необходимо зарегистрироваться на их сайте и сформировать API-ключ

Исходный набор данных для временного интервала в 15 минут, 1 час и 1 день имеет следующие признаки:

- time - время, по которой закрылась торговля в данный период времени;
- close - это цена, по которой закрылась торговля в данный период времени;
- low - наименьшая цена, зафиксированная на рынке в течение данного периода времени;
- high - наибольшая цена, зафиксированная на рынке в течение данного периода времени;

- open - это цена, по которой началась торговля в данный период времени;
- volumefrom - представляет собой количество базовой валюты, которое было продано за определенный период времени. Например, если на бирже было продано 100 BTC за день, то VolumeFrom за этот день будет равен 100 BTC;
- volumeto - с другой стороны, представляет собой количество вторичной валюты, которое было использовано для покупки базовой валюты. Например, если 100 BTC были проданы за 1 000 000 USD, то VolumeTo за этот день будет равен 1 000 000 USD;

Важный шаг перед обучением моделей - это обработка данных. Для успешного прогнозирования необходимо из имеющихся данных вычлениить наиболее важную информацию, с помощью которой нейронные сети могли бы найти зависимости между целевой переменной и другими признаками. Исходный набор данных для интервала 15 минут и 1 час содержит несколько основных признаков, между которыми моделям будет сложно найти зависимости. Основной задачей в шаге обработки данных было предоставить моделям такую же картину, которую видят пользователи смотря на график изменения цены криптовалюты. К исходному набору данных были добавлены следующие признаки:

- close\_diff - разница между ценами закрытия соседних свечей;
- low\_diff - разница между ценами низших точек соседних свечей;
- high\_diff - разница между ценами высших точек соседних свечей;
- open\_diff - разница между ценами открытия соседних свечей;
- color\_of\_candle - цвет свечи(красный или зеленый, красный цвет - цена закрытия свечи ниже цены открытия, зеленый цвет - цена закрытия свечи выше цены открытия);
- number\_candle\_trend - порядковый номер свечи в тренде(обозначает порядковый номер свечи в последовательности свечей одинакового цвета);
- width\_open\_close - разница между ценами открытия и закрытия свечи;
- width\_open\_high - разница между ценами открытия и высшей точкой свечи;
- width\_open\_low - разница между ценами открытия и низшей точкой свечи;
- SMA\_200 - скользящее среднее цены криптовалюты за последние 200 периодов времени;
- EMA\_200 - экспоненциальное скользящее среднее цены криптовалюты за



- последние 200 периодов;
- SMA\_50 - скользящее среднее цены криптовалюты за последние 50 периодов времени;
  - EMA\_50 - экспоненциальное скользящее среднее цены криптовалюты за последние 50 периодов;
  - RSI - технический индикатор, показывающий относительную силу цены, вычисляющийся за последние 14 периодов времени;
  - MACD - линия MACD, является результатом взятия более долгосрочной ЕМА и вычитания ее из более краткосрочной ЕМА. Наиболее часто используемые значения - 26 дней для долгосрочной ЕМА и 12 дней для краткосрочной ЕМА;
  - Signal - Сигнальная линия представляет собой ЕМА линии MACD, описанной в MACD. Трейдер может выбрать, какую ЕМА длины периода использовать для сигнальной линии, однако 9 является наиболее распространенной;
  - Histogram - С течением времени разница между линией MACD и сигнальной линией будет постоянно меняться. Гистограмма MACD учитывает это различие и отображает его в виде легко читаемой гистограммы. Разница между двумя линиями колеблется вокруг нулевой линии.;
  - close\_next/direction - цена закрытия следующей свечи/направление следующие свечи;

Набор данных для временного интервала в 1 день помимо признаков, которые присутствуют в датасетах с интервалом 15 минут и 1 час также имеет следующие данные.

Набор данных для временного интервала в 1 день разбит на 3 части. Первая набор данных имеет те же признаки, что и наборы данных для пятнадцати минут и одного часа. Второй набор данных включает в себя различные параметры, которые отражают состояние блокчейна криптовалюты. Он имеет в себе следующие признаки:

- **unique\_addresses\_all\_time** - это общее количество уникальных адресов в блокчейне криптовалюты за всю историю его существования;
- **new\_addresses** - это количество новых адресов, которые были созданы в блокчейне криптовалюты за определенный период времени;
- **active\_addresses** - это количество активных адресов в блокчейне крип-

товалюты за определенный период времени. Активные адреса - это те, которые были использованы в сети криптовалюты для отправки или получения средств;

- **transaction\_count** - это количество транзакций, которые были проведены в блокчейне криптовалюты за определенный период времени;
- **transaction\_count\_all\_time** - это общее количество транзакций, которые были проведены в блокчейне криптовалюты за всю его историю;
- **large\_transaction\_count** - это количество крупных транзакций в блокчейне криптовалюты за определенный период времени;
- **average\_transaction\_value** - это среднее значение размера транзакции в блокчейне криптовалюты за определенный период времени;
- **block\_height** - это номер последнего блока в блокчейне криптовалюты;
- **hashrate** - это скорость, с которой майнеры в блокчейне криптовалюты могут вычислять хэши;
- **difficulty** - это уровень сложности майнинга в блокчейне криптовалюты. Он определяет, насколько трудно майнерам вычислить хэш блока;
- **block\_time** - это время, которое требуется для создания нового блока в блокчейне криптовалюты;
- **block\_size** - это размер блока в блокчейне криптовалюты;
- **current\_supply** - это общее количество криптовалютных монет, которое уже было выпущено в блокчейне криптовалюты;

Третий набор данных содержит информацию о распределении баланса криптовалюты между кошельками пользователей разных размеров. Он имеет в себе следующие признаки:

- **data** - время, по которой закрылась торговля в разрезе дня;
- **totalVolume** - общий объем криптовалюты, находящийся на всех адресах, указанных в диапазоне;
- **addressesCount** - количество адресов, входящих в диапазон, у которых есть баланс криптовалюты;
- **range** - диапазон баланса криптовалюты на адресах. Например, если **range** равен 100-1000, то это означает, что баланс криптовалюты на адресах в этом диапазоне составляет от 100 до 1000 единиц;

Для отбора признаков были применены следующие методы: PCA, RFE и VIF.

Следующий важный этап в подготовке данных - это нормализация данных. Чаще всего в задачах регрессии датасет разбивается на тренировочный, тестовый и валидационный набор данных, после чего каждый из них отдельно нормализуется в определенном диапазоне и не подвергается нормализации целевая переменная. Т.к. в данной работе в качестве данных используются временные ряды, то важно сохранить структуру последовательности данных для обучения. Таким образом целевая переменная в задаче предсказания цены также будет нормализовываться.

Нормализация данных проводилась с помощью функции `MinMaxScaler` и `StandardScaler` из библиотеки `Sklearn`. Нормализация данных с помощью функций `MinMaxScaler` проводилась в диапазоне от 0 до 1, если признаки содержали только положительные значения и от -1 до 1, если признаки имели и отрицательные значения.

После нормализации данные было необходимо преобразовать в последовательности. Таким образом, чтобы одному предсказываемому значению соответствовала серия предыдущих данных за какое-то определенное количество шагов. Ниже приведена функция преобразования данных в такие последовательности.

### **Оценка качества прогнозов**

Для оценки качества прогнозирования моделей использовались, как абсолютные, так и процентные метрики, а также были созданы дополнительные метрики, оценивающие эффективность моделей с точки зрения их доходности.

Для прогнозирования цены закрытия свечи были использованы, следующие метрики:

- Метрики для задач регрессии: `RMSE`, `MAE`, `MAD`, `R2`, `SMAPE` и др.;
- Метрики для задач классификации: `Accuracy`, `recall`, `presicion`, `f1-score`;
- Дополнительные метрики: Средний процент предсказания направления, прибыльность и др.;

### **Ход исследования**

Первым этапом было определение набора предсказания для предсказания, вторым этапом - определение длины последовательностей, на основании которой модель предсказывала бы следующее значение, иначе говоря количество интервалов. Третьим этапом было определение архитектуры рекуррентной нейронной сети. Для каждой вида рекуррентной сети подбирались лучшая ар-

хитектура, потом они сравнивались между собой по метрикам и выбирался лучший слой рекуррентной сети из набора: LSTM, GRU, Bidirectional(LSTM), Bidirectional(GRU), CONV-LSTM) и лучшая модель.

### **Анализ результатов**

По итогам проделанной работы были получены следующие результаты:

На этапе отбора признаков лучшим способом оказался отбор признаков с помощью метода RFE. Для каждого метода определялось оптимальное количество признаков для обучения. Выбор производился на основе не только RMSE, но R2, SMAPE и принесенной прибыли по метрике count\_money.

На этапе определения количество элементов последовательности не было выявлено оптимального значения для всех случаев. Но в большей части модель показывала наиболее лучшие метрики, обучаясь на данных с размером последовательностей в 3 единицы времени. Так например, для задачи предсказания цены на 15-минутном таймфрейме модель показывает значительно более лучшие показатели метрик с размером последовательности в 3.

На этапе сравнения разных слоев RNN в среднем лучшими показателями обладали BiGRU, BiLSTM и Conv-LSTM сеть. Вероятно это связано с их более сложной структурой по сравнению с сетями LSTM и GRU, что может способствовать их более глубокому и гибкому обучению.

В плане сравнения обучения моделей на разных таймфреймах и применения нейронных сетей в задачах предсказания цены и бинарной классификации можно сделать следующие выводы:

- Наиболее лучшими для предсказания цены оказались данные с таймфреймом в 15 минут;
- В задачах бинарной классификации нейронные сети не показали высоких результатов, максимальные показатели accuracy достигали 55 процентов;
- Прогнозирование рынка на данных с таймфреймом в 1 день показало худшие результаты по сравнению с остальными. Несмотря на более расширенный выбор данных для построения модели, объем данных слишком мал, в следствии чего сложно дать объективную оценку обученным моделям;

### **Заключение**

Прогнозирование рынка криптовалют с помощью нейронных сетей представляет собой захватывающую и перспективную область исследований. В ходе

исследования были различные методы и модели прогнозирования, основанные на нейронных сетях, и их применимость к анализу криптовалютных рынков.

Результаты исследования показывают, что использование рекуррентных нейронных сетей в сочетании с тщательным выбором и обработкой данных может быть эффективным инструментом для прогнозирования рынка. В ходе исследования было выяснено, что лучшая показатели имела модель, обучающаяся на данных с 15-минутным таймфреймом, использующим в качестве отбора метод RFE и архитектуру нейронной сети с одним слоем CONV-LSTM сети.

Данную модель можно эффективно использовать как помощник при ручной торговле на бирже, так и в торговле роботом. Однако, следует отметить, что прогнозирование рынка криптовалют является сложной задачей, и нейронные сети не являются универсальным решением для всех сценариев. Успех прогнозирования зависит от качества данных, выбранных признаков и архитектуры нейронной сети. Также, важно учитывать быстро меняющуюся природу криптовалютных рынков и возможность появления непредсказуемых событий, которые могут оказать значительное влияние на результаты прогнозирования.