

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

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

Кафедра математической кибернетики и компьютерных наук

**ПОСТРОЕНИЕ ML-МОДЕЛИ ДЛЯ АВТОМАТИЗАЦИИ ТОРГОВЛИ
НА РЫНКЕ КРИПТОВАЛЮТНЫХ ФЬЮЧЕРСОВ**

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студента 2 курса 273 группы

направления 02.04.03 — Математическое обеспечение и администрирование
информационных систем

факультета КНиИТ

Севастьянова Владимира Александровича

Научный руководитель

к. ф.-м. н., доцент

С. В. Папшев

Заведующий кафедрой

к. ф.-м. н., доцент

С. В. Миронов

Саратов 2026

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 Анализ существующих решений и исследований	4
1.1 Краткое описание предметной области	4
1.1.1 Общая информация	4
1.1.2 Фьючерсы	4
1.2 Обзор и анализ существующих решений	5
1.3 Обзор и анализ научных исследований и статей	5
1.3.1 Прогнозирование ценовых значений криптовалют	5
1.3.2 Классификация трендов криптовалютного рынка	6
1.3.3 Обучение с подкреплением	6
1.3.4 Использование технического анализа и торговых признаков	6
1.3.5 Анализ	6
2 Сбор датасета и обучение ML-модели	7
2.1 Постановка задачи	7
2.2 Описание автоматизированной торговой системы	7
2.3 Подготовка датасета	7
2.4 Выбор и обучение базовой конфигурации ML-модели	8
2.5 Доработка датасета и ML-модели	9
3 Оценка эффективности модели посредством симуляции торговли	10
3.1 Разработка алгоритма симуляции торговли для тестирования модели	10
3.2 Результаты симуляции для модели	10
3.3 Доработка торговой системы для симуляции торговли	11
3.4 Результаты симуляции для торговой системы	11
ЗАКЛЮЧЕНИЕ	12

ВВЕДЕНИЕ

Криптовалютный рынок привлекает внимание инвесторов и трейдеров высокой волатильностью и потенциалом доходности, а фьючерсы на криптовалюты расширяют возможности торговли, но одновременно увеличивают риски. Высокая изменчивость и сложность этого рынка требуют постоянного анализа, мониторинга и принятия быстрых решений.

В условиях динамично меняющейся ситуации и роста количества активов ручная торговля становится менее эффективной. Автоматизация торговых процессов позволяет снизить влияние человеческого фактора, а методы машинного обучения могут использоваться для анализа данных, выявления закономерностей и оценки рыночной ситуации. Поэтому задача применения ML-модели в автоматизированной торговой системе является актуальной.

Целью данной работы является построение и обучение модели машинного обучения для определения оптимальных моментов входа в сделку на рынке криптовалютных фьючерсов с последующей интеграцией в автоматизированную торговую систему в качестве одной из подсистем алгоритма принятия решений.

Для достижения поставленной цели необходимо решить следующие задачи:

- проанализировать имеющиеся решения и изучить их устройство, применяемые методы и используемые технологии;
- провести анализ существующих исследований на тему применения методов машинного обучения при автоматизации торговли криптовалютами;
- определить набор значимых признаков и собрать датасет;
- выбрать и обосновать алгоритмы машинного обучения, наиболее подходящие для определения торговых сигналов;
- обучить модель, провести тестирование на исторических данных и оценить её эффективность;
- провести тестирование модели в реальном времени в различных рыночных ситуациях посредством симуляции торговли;
- интегрировать модель в торговую систему, провести тестирование посредством симуляции торговли и оценить изменение эффективности системы.

1 Анализ существующих решений и исследований

1.1 Краткое описание предметной области

1.1.1 Общая информация

Криптовалюта — цифровая форма денег, учёт владения которой ведётся в распределённом реестре. Торговля криптовалютами осуществляется на биржах; в данной работе рассматриваются централизованные биржи, поскольку именно они предоставляют развитую инфраструктуру для торговли фьючерсами, API-доступ к данным и механизмы исполнения ордеров.

При торговле на бирже ключевым источником информации является **биржевой стакан** — список заявок на покупку и продажу актива. Верхняя часть стакана содержит заявки на продажу (*ask*), нижняя — заявки на покупку (*bid*). Разница между лучшими ценами покупки и продажи образует **спред**, а объём заявок вблизи лучшей цены характеризует **ликвидность**. Эти параметры важны для автоматизированной торговли, поскольку прибыль от сделки должна покрывать комиссии, спред и возможное **проскальзывание**, то есть отличие ожидаемой цены исполнения от фактической.

Для анализа динамики цены также используются **японские свечи**. Каждая свеча агрегирует движение цены за фиксированный интервал времени и содержит цену открытия, цену закрытия, максимум, минимум и объём торгов. В работе используются минутные свечи, так как они являются компромиссом между доступностью данных через API бирж, объёмом хранимой информации и требуемой частотой принятия решений.

1.1.2 Фьючерсы

Фьючерс — производный финансовый инструмент, позволяющий заключать сделки на изменение цены базового актива. В работе рассматриваются беспоставочные бессрочные фьючерсы на криптовалюты, так как именно они широко используются на криптовалютных биржах и позволяют открывать позиции как в сторону роста цены (**лонг**), так и в сторону её снижения (**шорт**).

Важной особенностью фьючерсов является возможность использования **кредитного плеча**, при котором объём позиции превышает размер собственных средств трейдера. Это повышает потенциальную доходность, но одновременно увеличивает риск убытков и **ликвидации** — принудительного закрытия позиции при недостаточном уровне маржи. Поэтому при автоматизации торговли

фьючерсами особенно важны фильтрация слабых сигналов, контроль риска и учёт рыночной ликвидности.

При работе с криптовалютами необходимо учитывать высокую волатильность и новостной фон: на изменение цен влияют макроэкономические факторы, решения регуляторов, новости об инфраструктуре рынка, взломы бирж и протоколов, обновления блокчейнов, а также события вокруг отдельных активов. Такие факторы могут вызывать резкие разрывы в ценах и снижать качество моделей, основанных только на исторических ценах и объёмах.

1.2 Обзор и анализ существующих решений

Обзор существующих решений проводился в период с сентября по декабрь 2024 года. В ходе него были рассмотрены такие платформы, как Cryptohopper, 3Commas, Pionex, Trality, Mudrex, Intellectia, Octobot, CryptoHero и Freqtrade. Многие решения, как среди рассмотренных, так и среди прочих, заявляют об использовании искусственного интеллекта в целях привлечения внимания клиентов, однако используют его либо крайне ограниченно, либо в совершенно другом аспекте. Интересными для целей текущего исследования являются следующие способы применения методов машинного обучения: поставка торговых сигналов, интеграция с техническим анализом, оптимизация стратегий или анализ новостей, отчётов и событий в мире криптовалют. Однако сделать полноценные выводы, определить конкретные методы и провести сравнительный анализ аналогов не представляется возможным ввиду коммерческой тайны устройства большинства систем и значительной степени влияния пользователя на успешность торговли.

1.3 Обзор и анализ научных исследований и статей

1.3.1 Прогнозирование ценовых значений криптовалют

В ряде рассмотренных работ исследуется прямое прогнозирование цены или её направления. В таких исследованиях для Bitcoin сравниваются классические алгоритмы машинного обучения, включая Random Forest, Support Vector Regression и XGBoost, а также нейросетевые подходы, использующие исторические рыночные и блокчейн-данные. Общий вывод этих работ состоит в том, что криптовалютные временные ряды отличаются высокой волатильностью и нелинейностью, поэтому модели должны быть устойчивы к шуму и резким изменениям рыночной ситуации.

1.3.2 Классификация трендов криптовалютного рынка

Более близкими к теме работы являются исследования, рассматривающие прогнозирование как задачу классификации. В них применяются методы, учитывающие дисбаланс классов, связи между криптовалютами, а также рыночные, блокчейн- и социальные признаки. Такие работы показывают, что переход от численного прогноза цены к определению направления движения лучше согласуется с задачами торговых систем.

1.3.3 Обучение с подкреплением

Отдельное направление связано с обучением с подкреплением. В рассмотренных работах такие методы применяются для построения торговых агентов и управления портфелем, однако требуют более сложной среды симуляции, больших вычислительных ресурсов и тщательного контроля переобучения.

1.3.4 Использование технического анализа и торговых признаков

Ряд исследований использует признаки технического анализа и свечные данные. В них рассматриваются распознавание паттернов японских свечей, применение технических индикаторов для генерации торговых сигналов и постановка задачи торговли криптовалютами как классификации направления движения с последующей проверкой на исторических данных.

1.3.5 Анализ

Проведённый обзор показывает, что для задач автоматизированной торговли важны корректная подготовка данных, учёт дисбаланса классов, борьба с переобучением и проверка модели на исторических данных. Наиболее близкими к данной работе являются подходы, где модель не прогнозирует точную будущую цену, а классифицирует направление движения, торговый сигнал или рыночную ситуацию. Это соответствует выбранной постановке: модель должна выступать дополнительной подсистемой алгоритма принятия решений и оценивать потенциальный вход в сделку.

Также следует учитывать, что значительная часть исследований сфокусирована на Bitcoin. В настоящей работе рассматривается множество криптовалютных фьючерсов, поэтому модель должна учитывать различия в ликвидности, волатильности и характере движения разных контрактов.

2 Сбор датасета и обучение ML-модели

2.1 Постановка задачи

ML-модель разрабатывалась как дополнительная подсистема для уже существующей автоматизированной торговой системы. Её задача заключается не в самостоятельном управлении всеми этапами сделки, а в определении моментов, потенциально подходящих для входа в позицию на рынке криптовалютных фьючерсов. Выход модели представлен тремя классами: открытие в лонг, открытие в шорт и воздержание от открытия сделки. Главным критерием качества является не только значение классификационных метрик, но и влияние модели на результат торговой системы при последующей симуляции.

2.2 Описание автоматизированной торговой системы

Система реализована на Python и использует асинхронное взаимодействие с биржами и базой данных. Для хранения исторических данных и результатов работы применяется PostgreSQL. Архитектура включает backend, frontend, СУБД, алгоритмический сервис и сервис-прослойку для унифицированной работы с API бирж. Основным интересом представляет алгоритмический сервис: он выполняет парсинг рыночных данных, расчёт торговых метрик, поиск потенциальных сделок, принятие решений об открытии и закрытии позиций, а также оркестрацию торговых процессов. Вердикт ML-модели встраивается в существующий алгоритм как одна из независимых метрик качества потенциальной сделки.

2.3 Подготовка датасета

Для обучения требовалось собрать данные, отражающие как текущее состояние конкретного актива, так и общий рыночный контекст. В качестве исходных источников рассматривались фьючерсные курсы, биржевые стаканы, японские свечи и данные спотового рынка. После анализа ограничений API и доступных вычислительных ресурсов было принято решение собирать фьючерсные bid/ask значения с частотой 1 секунда и минутные японские свечи. Данные собирались по 150 наиболее ликвидным контрактам к USDT с бирж Binance, Bybit, HTX и Bitget.

Сбор стаканов и открытого интереса был исключён из-за ограничений API по частоте запросов и глубине доступного стакана. Парсинг актуальных курсов уже был реализован в системе, для свечных данных были добавлены асинхронные методы получения и записи в базу. Для каждой свечи сохранялись

время открытия, цены открытия и закрытия, максимум, минимум, объём торгов в активе и объём торгов в USDT. В результате был накоплен валидный набор данных более чем за год, с мая 2024 по октябрь 2025 года, общим объёмом более 1 Тб.

Разметка выполнялась по трём классам. Точка относилась к классу открытия в лонг или шорт, если в течение окна 15–30 минут после неё цена устойчиво смещалась в соответствующую сторону сильнее заданного порога, например, больше чем на 2%. Если такое движение не подтверждалось, точка относилась к классу воздержания. Чтобы исключить флуктуации курса, проверялось, что отклонение происходит в некоторой окрестности — полминуты.

В итоговый датасет вошли текущие bid/ask значения, минутные свечи за последний час, статистические признаки на скользящих окнах, RSI, отношение текущего объёма к среднему, а также аналогичные агрегированные признаки по BTC/USDT и ETH/USDT как индикаторы общего состояния рынка. Для сопоставимости активов абсолютные ценовые значения преобразовывались в относительные изменения, свечные признаки рассчитывались относительно соседних свечей или цены открытия текущей свечи, а объёмы нормализовались через отношение к среднему объёму за окно.

2.4 Выбор и обучение базовой конфигурации ML-модели

Данные обладают рядом особенностей: высокой волатильностью, шумом, дисбалансом классов и отсутствием строгой стационарности. Поэтому при выборе модели рассматривались как простые базовые алгоритмы, так и модели, способные учитывать нелинейные зависимости между признаками. Были проверены логистическая регрессия, дерево решений, случайный лес, XGBoost и LightGBM.

Логистическая регрессия использовалась как базовая интерпретируемая модель, но она недостаточно хорошо описывает нелинейные зависимости финансовых данных. Одиночное дерево решений склонно к переобучению, Random Forest оказался устойчивее, но хуже выделял редкие классы открытия сделки. XGBoost и LightGBM показали лучшие результаты среди рассмотренных вариантов; при этом LightGBM оказался немного эффективнее по совокупности F1-score, скорости обучения и доли ложных входов.

В базовой конфигурации LightGBM показал accuracy 0.84, macro-precision 0.75, macro-recall 0.73 и macro F1-score 0.73. Доля ложных входов среди предска-

заний классов открытия сделки составила 34%. Для данной задачи это важный показатель, так как ошибочный вход может приводить к убыточной сделке, тогда как пропуск потенциально удачного входа означает только упущенную возможность. Поэтому дальнейшая доработка была направлена не только на рост F1-score, но и на снижение числа ложных торговых сигналов.

2.5 Доработка датасета и ML-модели

Анализ базовых результатов показал, что качество модели может снижаться из-за участков данных, на которые сильно влиял новостной фон. Для учёта таких периодов использовался сервис CryptoPanic: из него извлекались важные новости, после чего в датасет добавлялся бинарный флаг, активный в течение двух часов после появления значимой новости. Такой подход позволяет учитывать нестандартные рыночные периоды, не добавляя в датасет информацию из будущего.

Другим изменением датасета является добавление времени суток и дня недели, что создаёт некоторую цикличность в данных и позволяет модели установить связь активности и специфики торгов с днём недели и временем торговых сессий в различных регионах, например, азиатской и американской сессиями.

Подбор гиперпараметров LightGBM выполнялся с помощью Optuna. После доработки датасета и настройки модели качество классификации улучшилось, результаты представлены в таблице 1.

Таблица 1 – Метрики LightGBM после доработки

Класс	Precision	Recall	F1-score	Accuracy
Открытие в лонг	0.74	0.70	0.72	—
Открытие в шорт	0.76	0.72	0.74	—
Воздержание от открытия	0.95	0.96	0.95	—
Avg	0.82	0.79	0.80	0.88

После доработки средний F1-score достиг 0.80, ассурасу составила 0.88, а средний процент ошибочных входов в сделку сократился с 34% до 25%. Однако классификационные метрики не полностью отражают результат модели, поскольку в реальных условиях прибыльность зависит от многих других факторов, поэтому итоговая оценка модели выполнялась посредством симуляции торговли.

3 Оценка эффективности модели посредством симуляции торговли

3.1 Разработка алгоритма симуляции торговли для тестирования модели

Для оценки модели была реализована симуляция торговли. В начале каждой минуты выполнялся парсинг свечей, в начале каждой секунды обновлялись актуальные bid/ask значения, после чего пересчитывались признаки в скользящих окнах и запускалось предсказание модели. Если модель предлагала открыть сделку, информация о ней сохранялась в базе данных. Отдельные фоновые процессы отвечали за проверку условий закрытия сделки и за обработку важных новостей, при появлении которых симуляция временно приостанавливалась.

Такой подход позволил учесть задержки получения данных и избежать ошибки look-ahead bias: модель принимала решение только на основании информации, которая была доступна к моменту очередной итерации симуляции. При этом симуляция оставалась упрощённой: не учитывались частичное исполнение ордеров, изменение ликвидности в стакане между принятием решения и исполнением, проскальзывание, фандинг, ликвидации и сбои API. Поэтому полученные результаты следует рассматривать как оценку полезности сигнала модели, а не как гарантированную доходность торговой стратегии.

Поскольку задача модели сводится только к выбору моментов входа в сделку, успешность сделки оценивалась по следующему правилу: если в период от 15 до 30 минут после открытия существовал промежуток не менее 30 секунд, в течение которого сделку можно было закрыть с прибылью, предсказание считалось успешным. Размер каждой симулируемой сделки задавался как одна сотая текущего депозита, комиссии учитывались, а параметры размера позиции и закрытия сделки не оптимизировались под выбранный период.

3.2 Результаты симуляции для модели

В результате самостоятельной симуляции на протяжении 21 дня в декабре 2025 года модель увеличила начальный депозит на 14%. Доля ошибочных сделок составила примерно четверть от общего числа, а средняя убыточность таких сделок оказалась выше средней прибыльности успешных входов. Это подтверждает, что для данной задачи важны не только стандартные метрики классификации, но и снижение числа ложных входов в сделку.

Полученный результат не следует интерпретировать как устойчивую ожи-

даемую доходность: период тестирования был ограничен, рынок оставался относительно спокойным, а несколько удачных или неудачных сделок могли заметно повлиять на итоговую динамику депозита. Тем не менее положительная самостоятельная симуляция показала, что модель способна формировать последовательность торговых сигналов, не приводящую к устойчивому снижению баланса при заданных правилах закрытия сделок.

3.3 Доработка торговой системы для симуляции торговли

Для оценки вклада модели в существующий алгоритм принятия решений торговая система была доработана так, чтобы на одном и том же потоке данных одновременно проверялись две версии: исходная и с ML-моделью. На каждой итерации решение об открытии сделки принималось дважды, а результаты сохранялись в две одинаковые по структуре таблицы симулируемых сделок. В версии с моделью её вердикт добавлялся как ещё одна метрика, поэтому сделка открывалась только при выполнении более строгого условия подтверждения.

Дополнительно сохранялись вердикты всех метрик, включая ML-модель. Это позволило после завершения симуляции отделить ситуации, когда сделка не была открыта именно из-за решения модели, от случаев, где результат определялся другими условиями в системе. Для корректного сравнения перед принятием решений фиксировался снимок всех необходимых данных, поскольку асинхронный парсинг мог обновить значения во время обработки очередной итерации.

3.4 Результаты симуляции для торговой системы

В результате симуляции на протяжении 21 дня в апреле 2026 года торговая система с учётом ML-модели открыла на 22% меньше сделок, чем исходная версия, и увеличила начальный депозит на 37% против 28% у исходной системы. Доля ошибочных сделок составила около 20%, при этом почти в трети случаев вердикт модели не влиял на итоговое решение алгоритма.

Сравнение двух вариантов системы показывает, что основной эффект от внедрения модели заключается в фильтрации части слабых входов. Система с моделью открывала меньше сделок, но итоговый результат оказался выше, что согласуется с выбранной постановкой: ML-модель не заменяет алгоритм принятия решений, а выступает дополнительной подсистемой оценки потенциальной сделки.

ЗАКЛЮЧЕНИЕ

В рамках данной работы был проведён обзор существующих решений для автоматизации торговли на рынке криптовалют и криптовалютных фьючерсов, а также проанализированы исследования, посвящённые применению методов машинного обучения в данной предметной области. На основе проведённого анализа и ввиду особенностей системы, для интеграции в которую разрабатывалась модель, был выбран подход, при котором ML-модель используется не для прямого прогнозирования будущей цены, а для классификации рыночной ситуации и определения моментов, потенциально подходящих для входа в сделку.

В ходе практической части работы был определён набор значимых признаков, организован сбор исторических данных и выполнена разметка датасета для трёх классов. Были рассмотрены такие модели машинного обучения, как логистическая регрессия, дерево решений, случайный лес, XGBoost и LightGBM, последняя была выбрана в качестве основной. После доработки датасета и подбора гиперпараметров итоговая модель показала средний F1-score 80%, а средний процент ошибочных входов в сделки сократился с 34% до 25%.

Для оценки модели в условиях, приближенных к реальным, была разработана симуляция торговли. В симуляции, где модель принимала решения самостоятельно, за 21 день начальный депозит увеличился на 14%. После интеграции в автоматизированную торговую систему модель сократила количество открываемых в системе сделок на 22% и повысила её доходность с 28% до 37%. Это показывает, что модель может быть полезна как дополнительный фильтр, позволяющий сократить число неудачных входов в сделки и повысить качество решений существующего алгоритма.

При этом полученные результаты следует рассматривать с рядом ограничений, так как симуляция не полностью отражает реальные условия: на практике итоговую доходность могут снизить проскальзывания, задержки исполнения, изменение ликвидности, ликвидации. Кроме того, исследуемые периоды были относительно спокойными для рынка, что могло положительно сказаться на результате. Тем не менее цель работы была достигнута: построена, обучена и протестирована ML-модель, пригодная для интеграции в автоматизированную торговую систему в качестве одной из подсистем алгоритма принятия решений.

Результаты работы были представлены на студенческой научной конференции факультета КНиИТ в 2026 году.