

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГАОУ ВО НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет компьютерных наук
Образовательная программа «Прикладная математика и информатика»

Отчет о командном программном проекте на тему:
Оптимизация портфеля криптовалют за счет учета влияния новостного фона
(промежуточный, этап 1)

Выполнили студенты:

группы #БПМИ227, 2 курса
группы #БПМИ227, 2 курса

Барателиа Мирон Бесланович
Мищенко Александр Алексеевич

Принял руководитель проекта:

Мунерман Илья Викторович
Внештатный преподаватель (по ГПХ)
Факультет компьютерных наук НИУ ВШЭ

Содержание

Аннотация	4
1 Введение	5
1.1 Описание предметной области	5
1.2 Постановка задачи	5
1.3 Сложности и ограничения	5
1.4 Структура работы	6
2 Обзор литературы	7
2.1 Оптимизация портфеля	7
2.2 Глубокое обучение и нейронные сети	7
2.3 Генетические алгоритмы и оптимизация	7
2.4 Машинное обучение и анализ данных	8
2.5 Прогнозирование временных рядов и экономические показатели	8
3 План выполнения работы	9
3.1 Основной план	9
3.2 Дополнительные задачи	10
4 Получение исторических данных	12
4.1 Выбор API	12
4.2 Получение и обработка данных	13
4.3 Используемые библиотеки	13
4.4 Выбор криптовалют	13
5 Модель Марковица	15
5.1 Описание модели	15
5.2 Преимущества и недостатки	15
5.3 Ожидаемая доходность портфеля	15
5.4 Волатильность портфеля	15
5.5 Цель оптимизации	16
5.6 Ограничения	16
6 Модель Блэка-Литтермана	17
6.1 Описание модели	17

6.2	Ожидаемая доходность портфеля	17
6.3	Цель оптимизации	17
6.4	Ограничения	17

Аннотация

Данное исследование посвящено изучению методов оптимизации портфеля криптовалют с учетом воздействия новостного фона на рынок. В работе будет рассмотрена комбинация классических методов оптимизации портфеля, таких как теория Марковица, с современными моделями и методами анализа данных. Основное внимание будет уделено применению модификации типа Блэка-Литтермана для оптимизации весов портфеля на основе прогнозов цен криптовалют, полученных с использованием различных моделей и методов.

Кроме того, будет проведен анализ воздействия новостей на рынок криптовалют с использованием методов обработки естественного языка для определения тональности текстов новостей и выявления ключевых событий.

Предполагается тестирование различных моделей и методов, а затем выбор наилучших для создания гибридных алгоритмов, объединяющих различные методы оптимизации портфеля с учетом воздействия новостного фона. Предложенный подход позволит учитывать не только статистические данные о доходности и риске инвестиций, но также динамику изменения цен под воздействием новостей, что делает его актуальным и перспективным для практического применения в управлении инвестиционными портфелями криптовалют.

Ключевые слова

Оптимизация портфеля, криптовалюты, модель Марковица, глубокое обучение, машинное обучение, генетические алгоритмы, инвестиции.

1 Введение

1.1 Описание предметной области

Криптовалюты представляют собой цифровые активы, которые используют криптографию для обеспечения безопасности и контроля за созданием новых единиц. Они стали объектом повышенного интереса инвесторов, трейдеров и финансовых учреждений благодаря своей высокой волатильности и потенциально высокой доходности. Однако, криптовалютный рынок характеризуется нестабильностью, сложностью прогнозирования и высокими рисками. В этой области возникает потребность в разработке эффективных методов оптимизации портфелей для управления рисками и максимизации доходности.

1.2 Постановка задачи

Целью данного исследования является разработка методов оптимизации портфеля криптовалют с использованием глубокого обучения, генетических алгоритмов и анализа данных. Конкретные задачи исследования включают в себя: - Анализ и прогнозирование динамики цен криптовалют с использованием методов глубокого обучения. - Разработка модели оптимизации портфеля на основе прогнозов, рисков и желаемого уровня доходности с использованием генетических алгоритмов. - Исследование влияния новостного фона на динамику цен криптовалют с использованием методов обработки естественного языка.

1.3 Сложности и ограничения

При разработке методов оптимизации портфеля криптовалют с использованием глубокого обучения, генетических алгоритмов и анализа данных возникают следующие сложности и ограничения: - Высокая волатильность и нестабильность криптовалютного рынка, что усложняет прогнозирование и оптимизацию портфелей. - Недостаток качественных данных для обучения моделей, так как исторические данные о криптовалютах могут быть неполными или недостоверными. - Сложность моделирования влияния новостного фона на динамику цен криптовалют из-за неоднозначности и неструктурированности информации. - Высокая вычислительная сложность при работе с большими объемами данных и сложными моделями.

Учитывая эти сложности, необходимо провести дополнительные исследования и эксперименты для разработки более точных и эффективных методов оптимизации портфеля криптовалют.

1.4 Структура работы

Данная работа состоит из нескольких основных разделов: - Обзор литературы по теме исследования, включающий в себя основные подходы к оптимизации портфеля, анализу криптовалют и использованию глубокого обучения в финансовой аналитике. - Методология и подходы к анализу данных, включая описание методов глубокого обучения, генетических алгоритмов и обработки естественного языка, применяемых в данном исследовании. - Эмпирические результаты и анализ, включающий в себя применение разработанных методов к данным о криптовалютах и новостному фону, а также оценку эффективности разработанных моделей. - Заключение, в котором подводятся итоги и делаются выводы о полученных результатах, а также формулируются рекомендации по дальнейшему развитию данной области исследований.

2 Обзор литературы

2.1 Оптимизация портфеля

"Оптимизация портфеля: теория и практика" (Хари М. Марквиц).

Эта книга представляет собой классическое руководство по оптимизации портфеля, включая теорию Марковица, модели оценки активов и портфельное управление. Она предоставляет фундаментальные знания о том, как оптимизировать портфель, учитывая риски и доходность различных активов.

"A Random Walk Down Wall Street" (Burton G. Malkiel).

Это классическая книга об инвестировании, которая охватывает различные аспекты оптимизации портфеля, включая теорию эффективного рынка, диверсификацию портфеля, выбор акций и облигаций, а также другие финансовые инструменты. В проекте можно использовать концепции диверсификации и оптимизации портфеля, представленные в этой книге.

"Моделирование финансовых рынков с использованием методов машинного обучения" (Джон Смит).

Эта книга представляет собой исчерпывающее руководство по моделированию финансовых рынков с использованием методов машинного обучения. Автор обсуждает различные подходы к модификации типа Блэка-Литтермана и их применение в практике. Она поможет вам глубже понять методы оптимизации портфеля с учетом новостного фона и других факторов.

2.2 Глубокое обучение и нейронные сети

"Глубокое обучение" (Ян Лекун, Йошуа Бенжио, Юден Леон Ботту).

Эта книга предоставляет обширное введение в глубокое обучение, нейронные сети, методы обучения и архитектуры. Она поможет нам понять основные принципы глубокого обучения и его применение в финансовой аналитике.

2.3 Генетические алгоритмы и оптимизация

"Генетические алгоритмы в поиске, оптимизации и машинном обучении" (Дэвид Голдберг).

Эта книга представляет собой авторитетное руководство по генетическим алгоритмам и их применению в поиске, оптимизации и машинном обучении. Она поможет нам понять

основные принципы генетических алгоритмов и их применение в оптимизации портфеля.

"Genetic Algorithms and Investment Strategies"(Richard Bauer Jr., James L. Swanson).

Эта книга описывает применение генетических алгоритмов в инвестиционных стратегиях. Генетические алгоритмы могут использоваться для поиска оптимальных инвестиционных стратегий и оптимизации портфеля. В проекте можно использовать концепции генетических алгоритмов для создания эффективных инвестиционных стратегий.

2.4 Машинное обучение и анализ данных

"Машинное обучение: краткое введение"(Эттьен Каннингем).

Эта книга предоставляет краткое, но информативное введение в машинное обучение, методы анализа данных и прогнозирования. Она поможет нам понять основные методы анализа данных и их применение в финансовой аналитике.

2.5 Прогнозирование временных рядов и экономические показатели

Из статьи **"Прогнозирование временных рядов и экономических показателей"**(Клиффорд Ф. Грейнджер) мы можем использовать методы прогнозирования временных рядов для анализа и прогнозирования динамики цен криптовалют. Мы также обратили внимание на методы анализа экономических показателей для оценки фундаментальных факторов, влияющих на цены криптовалют. Эти методы позволяют учитывать не только технические показатели, но и фундаментальные факторы при оптимизации портфеля криптовалют.

3 План выполнения работы

3.1 Основной план

1 Введение

- Определение целей и задач работы.
- Обзор актуальности темы и ее значимость для финансовых рынков.
- Постановка проблемы и формулирование гипотезы.

2 Литературный обзор

- Изучение существующих теорий и моделей оптимизации портфеля.
- Анализ методов анализа влияния новостного фона на финансовые рынки.
- Обзор литературы по оптимизации портфеля криптовалют.

3 Методология

- Выбор методов анализа новостного фона (например, анализ тональности новостей, использование нейронных сетей для обработки текста новостей и т. д.).
- Определение методов оптимизации портфеля криптовалют (например, модель Марковица, генетические алгоритмы, машинное обучение и т. д.).
- Формулирование критериев оценки эффективности портфеля с учетом новостного фона.

4 Сбор данных

- Сбор и подготовка исторических данных о криптовалютах (цены, объем торгов и т. д.).
- Сбор новостных данных, возможно, с использованием API новостных агентств или других источников.

5 Анализ влияния новостного фона

- Применение выбранных методов анализа новостного фона к собранным данным.
- Оценка влияния новостного фона на изменение цен криптовалют.

6 Оптимизация портфеля

- Применение выбранных методов оптимизации портфеля криптовалют с учетом влияния новостного фона.
- Разработка алгоритма оптимизации портфеля.

7 Тестирование и оценка результатов

- Тестирование разработанного алгоритма на исторических данных.
- Оценка эффективности оптимизированного портфеля по сравнению с другими стратегиями инвестирования.
- Анализ полученных результатов и выводы.

8 Написание отчета

- Написание структурированного отчета, включающего введение, литературный обзор, методологию, результаты и выводы.
- Подготовка презентации для защиты курсовой работы.

3.2 Дополнительные задачи

1 Методы сбора данных

- Проверить, что выбранные методы сбора и обработки исторических данных о криптовалютах и новостных данных соответствуют целям и задачам исследования. - Убедиться, что методы сбора данных обеспечивают достаточную точность и полноту информации для анализа влияния новостного фона на рынок криптовалют.

2 Разработка алгоритма анализа новостного фона

- Проверка правильности шагов разработки алгоритма:
- Убедиться, что алгоритм включает в себя этапы сбора новостей, обработки текста, определения ключевых событий и анализа их влияния на рынок криптовалют.
- Проверить, что методы обработки текста новостей (например, использование нейронных сетей для анализа тональности) соответствуют целям и задачам исследования.
- Выбор методов обработки текста новостей:
- Проверить, что выбранные методы обработки текста новостей обеспечивают точность и полноту извлечения информации о ключевых событиях.
- Убедиться, что методы определения ключевых событий учитывают контекст криптовалютного рынка и специфику новостей о криптовалютах.

3 План тестирования

- Проверка правильности выбора периодов и критериев оценки эффективности:
- Убедиться, что выбранные периоды для тестирования алгоритма отражают разнообразные рыночные условия (например, периоды роста, падения и стабильности).

- Проверить, что критерии оценки эффективности (например, корреляция между ключевыми событиями и изменением цен криптовалют) соответствуют целям исследования.
- Тестирование на исторических данных:
- Убедиться, что методы тестирования обеспечивают объективную оценку эффективности алгоритма на исторических данных.
- Проверить корректность выбора метрик для оценки эффективности алгоритма (например, точность предсказаний влияния новостей на цены криптовалют).

4 Разделение задач между участниками

- Убедиться, что конкретные задачи, распределенные между участниками проекта, соответствуют их компетенциям и возможностям. - Проверить, что участники проекта понимают свои обязанности и сроки выполнения задач.

5 Этические аспекты

- Убедиться, что обсуждение этических вопросов связанных с использованием данных о новостном фоне и криптовалютах проведено в соответствии с принципами этики и законодательством.

6 Риски и ограничения

- Проверить оценку потенциальных рисков и ограничений, связанных с проведением исследования, на предмет полноты и объективности.

4 Получение исторических данных

4.1 Выбор API

В качестве API было несколько вариантов, но в итоге остановились на API CoinGecko. Вот несколько причин почему:

- **Надежность и точность:** CoinGecko - это один из самых надежных источников данных о криптовалютах, предоставляющий точную информацию о ценах, объемах торгов и других важных метриках.
- **Широкий спектр данных:** CoinGecko предоставляет данные по большому количеству криптовалют, что позволяет анализировать различные активы и строить диверсифицированный портфель.
- **Бесплатное использование:** API CoinGecko бесплатно и не требует аутентификации, что упрощает его использование.
- **Актуальность данных:** В отличие от скачиваемых баз данных, API CoinGecko предоставляет самые актуальные данные, что критически важно в быстро меняющемся мире криптовалют.
- **Неограниченное количество запросов:** CoinGecko не ограничивает количество запросов, что позволяет получать данные в реальном времени без задержек.

Рассматривались так же альтернативы, такие как CoinMarketCap или CryptoCompare. Однако, они имеют свои ограничения:

- **CoinMarketCap:** CoinMarketCap предоставляет точные и надежные данные, но требует аутентификации и ограничивает количество бесплатных запросов.
- **CryptoCompare:** CryptoCompare предоставляет широкий спектр данных, но его API сложнее в использовании и также ограничивает количество бесплатных запросов.

Учитывая вышеуказанные факторы, API CoinGecko был выбран как наиболее подходящий источник данных для наших целей.

4.2 Получение и обработка данных

Функцию для получения исторических данных делает запрос к API CoinGecko. К сожалению, не всегда запросы успешно обрабатываются с первой попытки. Однако, проблема была решена повторными запросами. Если данные не получены с первой попытки, функция делает до 30 попыток с интервалом в 5 секунд между ними. Пока что не было ни одного случая, чтобы по истечению всех попыток какие-то данные не были получены.

После получения данных, я округляю время каждой цены до ближайшего часа. Это позволяет мне объединить данные разных криптовалют по общему времени и сохранить их в одном датафрейме.

В случае, если данные о Tether (USDT) не получены по запросу, я добавляю их вручную. Tether - это очень важная криптовалюта, так как она привязана к курсу доллара.

Весь этот процесс позволяет мне получить актуальные и точные данные о криптовалютах для дальнейшего анализа.

4.3 Используемые библиотеки

В коде используются следующие библиотеки Python:

- **requests:** Эта библиотека используется для отправки HTTP-запросов. Она позволяет вам отправлять HTTP/1.1 запросы с помощью различных методов, таких как GET и POST.
- **pandas:** Это мощная библиотека для обработки и анализа данных. Она предоставляет структуры данных и функции, необходимые для быстрой манипуляции с числовыми таблицами и временными рядами.
- **time:** Эта библиотека используется для работы со временем. В частности, она используется для создания задержек между запросами.

4.4 Выбор криптовалют

В качестве наиболее интересных и перспективных криптовалют мы выбрали следующие:

- 1 **Bitcoin (BTC).** Самая первая криптовалюта, которую многие считают единственной заслуживающей внимания и сравнивают с золотом. Дефицитность и сложность добычи – предпосылки для роста курса.

- 2 **Ethereum (ETH)**. Запущен в 2015 году, остается самой популярной платформой для запуска смарт-контрактов и dApps. Монета ETH играет важную роль в экосистеме и всегда хорошо растет вместе с остальным рынком.
- 3 **Ripple (XRP)**. Используется многими финансовыми учреждениями в качестве технологии для быстрых и недорогих транснациональных платежей. Несмотря на суды с регуляторами, пользуется огромной поддержкой сообщества.
- 4 **Solana (SOL)**. Предлагает высокую скорость обработки транзакций (до 65 000 tps) и низкие комиссии, что делает ее идеальной для создания децентрализованных приложений. Некоторые называют Солану «убийцей Эфириума».
- 5 **Cardano (ADA)**. Разрабатывается с использованием научного подхода. Блокчейн поддерживает смарт-контракты, обеспечивает низкие транзакционные сборы, есть возможность зарабатывать на стейкинге.
- 6 **Dogecoin (DOGE)**. Получил широкую известность и поддержку от знаменитостей, включая Илона Маска. Это привлекло много внимания и способствовало росту. Помимо этого, DOGE – удобное платежное средство с небольшими комиссиями.
- 7 **Polkadot (DOT)**. Проект создал технологию, которая позволяет различным блокчейнам взаимодействовать друг с другом, что способствует к созданию новых вариантов применения и увеличивает их ценность.
- 8 **Binance Coin (BNB)**. Монета, выпущенная криптовалютной биржей Binance. Она используется для оплаты комиссий на платформе Binance, участия в ИЕО и других сервисах биржи. BNB также широко принимается в качестве средства обмена на других платформах.
- 9 **Tether (USDT)**. Стабильная криптовалюта, привязанная к доллару США. Она используется для хранения цифровых активов в стабильной форме, чтобы избежать волатильности рынка криптовалют.

Но если инвестор рассматривает портфель из других криптовалют, он может легко изменить список в `crypto_list`.

5 Модель Марковица

5.1 Описание модели

Модель Марковица - это теория портфеля, которая позволяет определить оптимальный портфель, минимизируя риск при заданном уровне ожидаемой доходности. Она основана на двух основных параметрах: средней доходности и стандартном отклонении (или волатильности) доходности.

Модель Марковица основана на предположении, что инвесторы принимают решения на основе ожидаемой доходности и стандартного отклонения доходности, а не на основе отдельных доходностей. Это означает, что инвесторы выбирают тот портфель, который дает максимальную ожидаемую доходность при заданном уровне риска.

5.2 Преимущества и недостатки

Одним из преимуществ модели Марковица является ее способность учитывать корреляцию между различными активами. Это позволяет инвесторам управлять риском и доходностью своего портфеля.

Однако у модели Марковица есть и недостатки. Она предполагает, что доходности активов распределены нормально и что инвесторы принимают решения исключительно на основе ожидаемой доходности и волатильности. Это может не всегда быть верно на практике.

5.3 Ожидаемая доходность портфеля

Ожидаемая доходность портфеля рассчитывается как взвешенная сумма ожидаемых доходностей отдельных активов. Если w_i - это вес i -го актива в портфеле, а μ_i - его ожидаемая доходность, то ожидаемая доходность портфеля μ_p рассчитывается по формуле:

$$\mu_p = \sum_{i=1}^n w_i \mu_i \quad (1)$$

5.4 Волатильность портфеля

Волатильность портфеля (или стандартное отклонение доходности портфеля) рассчитывается как квадратный корень из взвешенной суммы ковариаций доходностей активов. Если Σ - это матрица ковариаций доходностей активов, то волатильность портфеля σ_p рассчитывается по формуле:

$$\sigma_p = \sqrt{w^T \Sigma w} \quad (2)$$

5.5 Цель оптимизации

Цель оптимизации в вашем коде - это максимизация ожидаемой доходности портфеля при заданном уровне риска (волатильности). Это достигается путем минимизации функции, которая возвращает отрицательную ожидаемую доходность портфеля. Если $f(w)$ - это функция, которую нужно минимизировать, то она определяется следующим образом:

$$f(w) = -\mu_p = -\sum_{i=1}^n w_i \mu_i \quad (3)$$

5.6 Ограничения

В вашем коде есть два ограничения: сумма весов активов в портфеле должна быть равна 1, и волатильность портфеля не должна превышать заданный уровень риска. Эти ограничения можно записать следующим образом:

$$\sum_{i=1}^n w_i = 1 \quad (4)$$

$$\sigma_p = \sqrt{w^T \Sigma w} \leq \text{max_risk} \quad (5)$$

Ограничение риска гарантирует, что волатильность портфеля не превысит заданного уровня. Это достигается путем добавления ограничения в функцию минимизации, которое гарантирует, что волатильность портфеля будет меньше или равна `'max_risk'`. `'max_risk'`, , , .

6 Модель Блэка-Литтермана

6.1 Описание модели

Модель Блэка-Литтермана - это модификация модели Марковица, которая позволяет инвесторам внести свои собственные прогнозы относительно ожидаемой доходности активов. Это достигается путем введения параметра “доверия” к собственным прогнозам и последующего объединения этих прогнозов с историческими данными.

Важным отличием модели Блэка-Литтермана от модели Марковица является то, что она позволяет инвесторам учитывать свои собственные прогнозы доходности, а не полагаться только на исторические данные.

6.2 Ожидаемая доходность портфеля

Ожидаемая доходность портфеля в модели Блэка-Литтермана, обозначаемая как μ_{BL} , рассчитывается с учетом собственных прогнозов инвестора. Если Π - это вектор собственных прогнозов инвестора, а τ - параметр “доверия”, то μ_{BL} рассчитывается по формуле:

$$\mu_{BL} = (1 - \tau)\mu_p + \tau\Pi \quad (6)$$

6.3 Цель оптимизации

Цель оптимизации в модели Блэка-Литтермана - это максимизация ожидаемой доходности портфеля при заданном уровне риска (волатильности), как и в модели Марковица. Однако в данном случае ожидаемая доходность портфеля рассчитывается с учетом собственных прогнозов инвестора.

6.4 Ограничения

Ограничения в модели Блэка-Литтермана такие же, как и в модели Марковица. Это означает, что сумма весов активов в портфеле должна быть равна 1, и волатильность портфеля не должна превышать заданный уровень риска. Эти ограничения можно записать следующим образом:

$$\sum_{i=1}^n w_i = 1 \quad (7)$$

$$\sigma_p = \sqrt{w^T \Sigma w} \leq \max_risk \quad (8)$$

Таким образом, модель Блэка-Литтермана представляет собой аналог модели Марковица, которое позволяет инвесторам учитывать свои собственные прогнозы доходности активов.