

Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)

Хорошо



Михаил Шардин ★

личный блог



25 августа 2026, 04:39

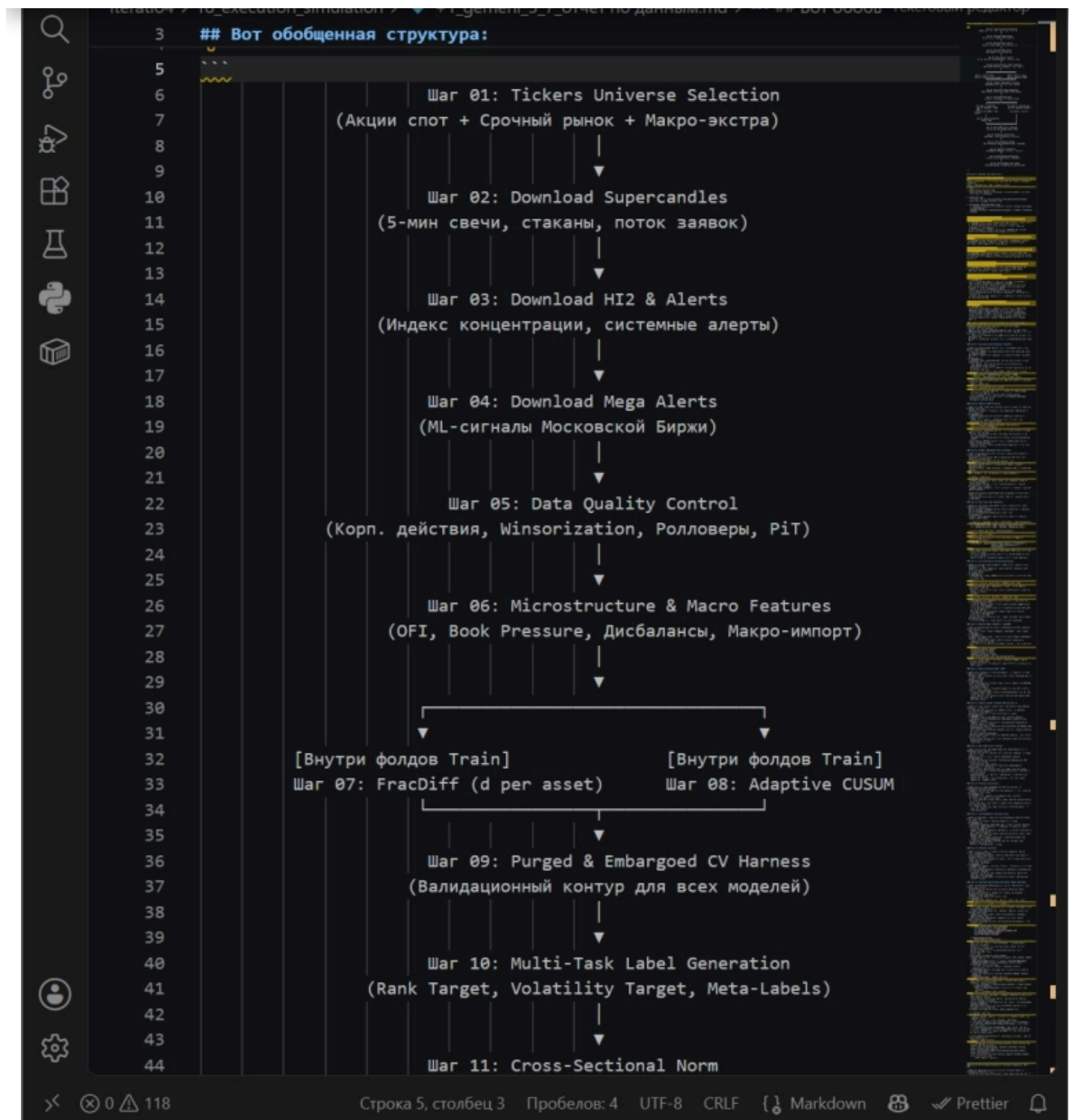
+ Подписаться

Как я провёл лето: строил ML-пайплайн для торговли на Мосбирже

Лето подходит к концу и я хочу рассказать о своём увлечении последних месяцев — машинном обучении (Machine Learning) на данных Московской биржи. Эта тема уже давно не даёт мне покоя и в этот раз я решил подойти более основательно — активно использовал новейшие языковые модели, типа Claude Sonnet 5, Gemeni 3.7 Flash не только для написания кода, но и для работы с результатами этих скриптов. При этом не использовал полностью автономных агентов — все языковые модели работали как консультанты в чате — все решения принимал я, в том числе какую исходную информацию давать конкретной модели на каждом из этапов.

Введите текст комментария

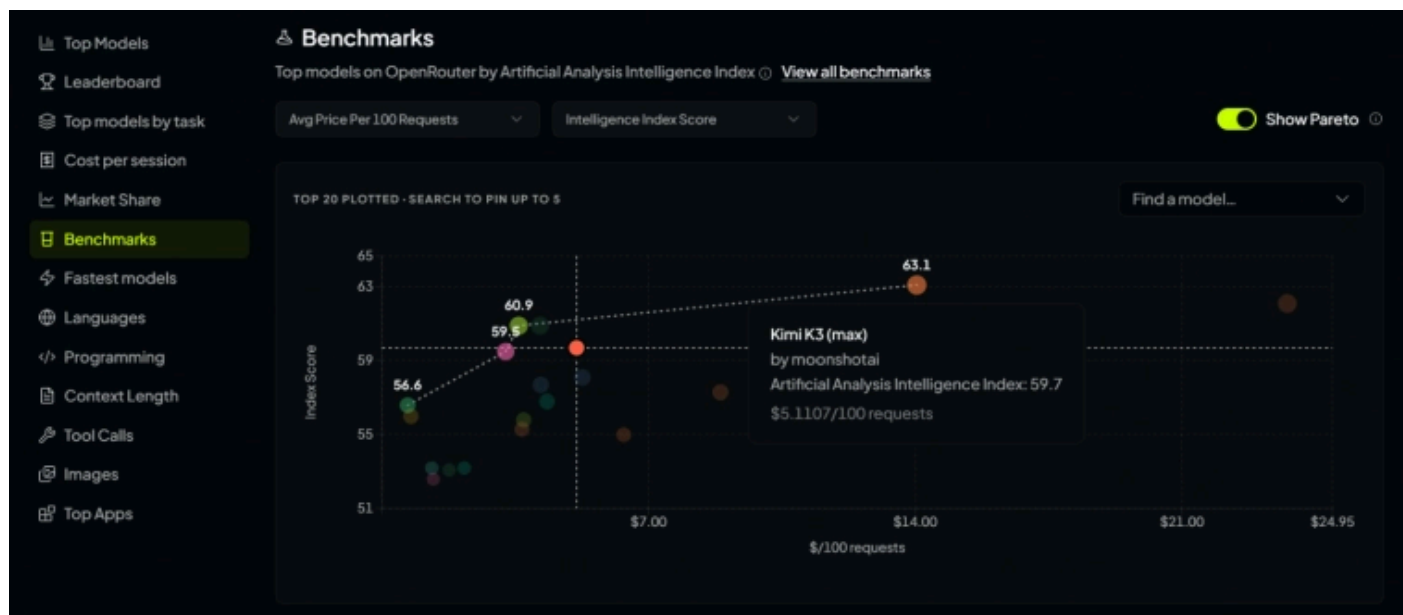
Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)



Предвидя будущие вопросы я сразу отвечу что это мой эксперимент и он не только не зарабатывает, но и пока даже непонятно чем закончится — сейчас это всё в процессе. Эту статью решил написать не чтобы «засветить трейдерским граалем» — которых я кстати считаю что не существует, потому что вся информация общедоступна. Эта статья написана

Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)

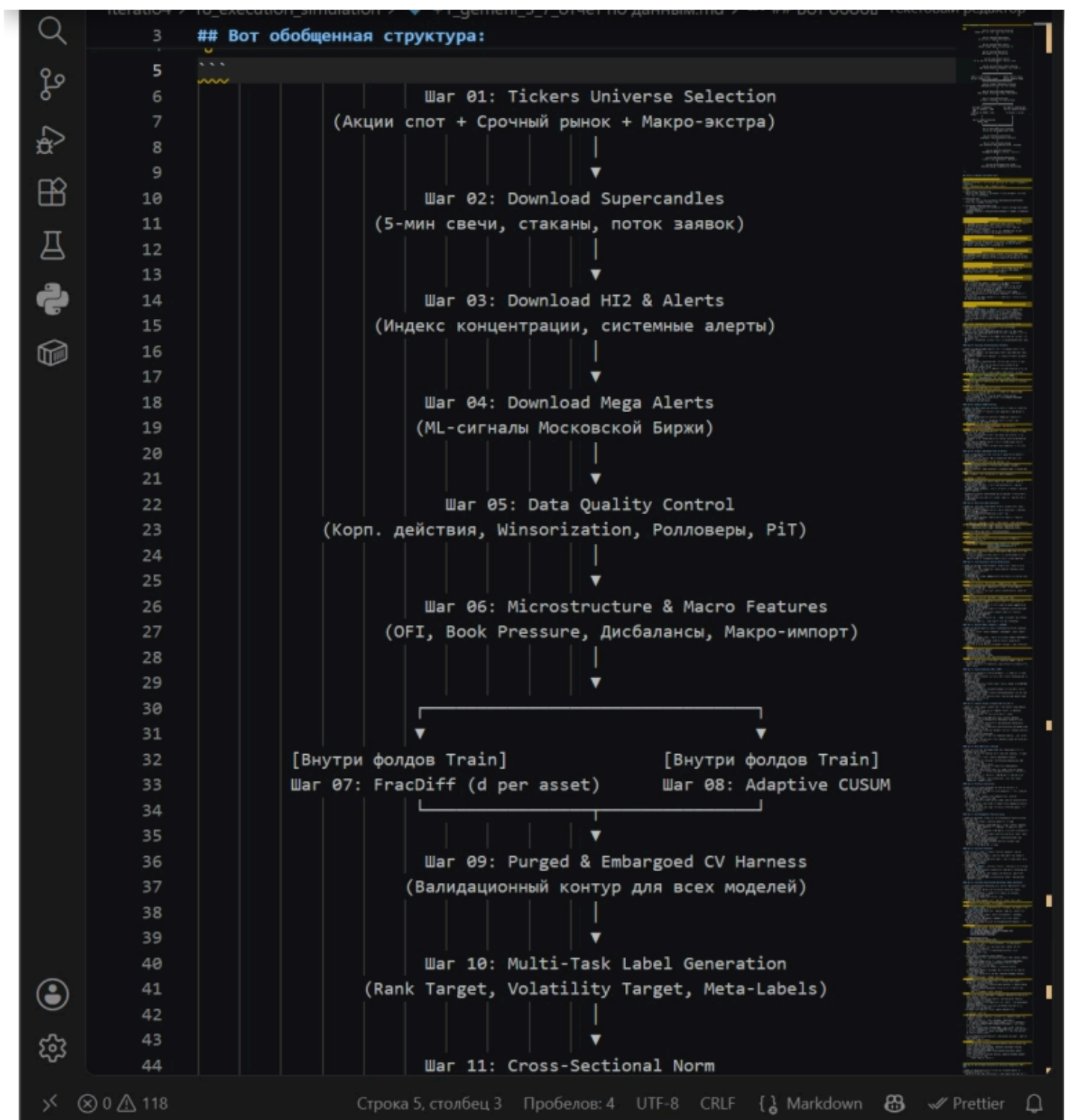
ты и можешь сразу найти понравившиеся интересные модели, которые профессионально занимаются темой — так уже бывало.



Лучшие модели по индексу искусственного интеллекта по версии <https://openrouter.ai/benchmarks>. Использую этот график (он меняется, а не статичен) чтобы выбирать модели — какие вообще модели использовать, с кем переписываться. Можно выбирать как по умности, так и по способности корректно писать код — это разные категории

Перед тем как начинать этот эксперимент я переписывался с Claude Fable 5 в чате — это самая умная и самая дорогая публично доступная большая языковая модель — у меня были некоторые свои мысли, а также источники в виде книг и статей и мне хотелось чтобы модель оформила все в пошаговый план действий. Консультация через чат-бота стоила примерно \$20.

Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)



Мой план действий, оформленный в текстовую схему в Visual Studio Code, слегка обрезался на скриншоте, шагов итого 20

В итоге я остановился на такой архитектуре, когда каждый шаг — это независимый набор Python скриптов, который запускается вручную и работают только с конкретным шагом

Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)

американским рынком) и шортить акции на Московской бирже — дорого и не всегда есть возможность — особенно для частного лица. Из-за этого я выбрал открывать длинную позицию в акциях и одновременно короткую — во фьючерсе.

```

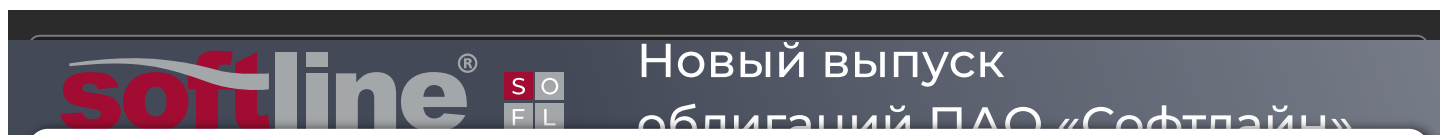
1  """
2  =====
3  НАЗНАЧЕНИЕ:
4      Главный оркестратор
5      Собирает акции с фьючерсами и дополнительные фьючерсы,
6      формируя данные точно по запрошенной структуре. Добавлен вывод статистики.
7  =====
8  """
9
10 import time
11 import os
12 from dadata import Dadata
13
14 from config import (
15     DADATA_TOKEN, STOCKS_TO_FUTURES_MAP,
16     START_YEAR_HISTORY, EXTRA_FUTURES_ASSETS,
17     EXCLUDED_NON_STOCK_TICKERS
18 )
19 import utils
20 import moex_api
21 import enrichment_api
22
23 def main():
24     start_time = time.time()
25     script_dir = os.path.dirname(os.path.abspath(__file__))
26     utils.setup_logging(os.path.join(script_dir, 'tickers.log'))
27
28     print("🔥 ЗАПУСК: Сбор фьючерсов и акций\n")
29
30     moex = moex_api.MoexClient()
31     dadata_client = Dadata(DADATA_TOKEN) if DADATA_TOKEN else None
32
33     # Словари для накопления статистики
34     stats_extra = {}
35     stats_stocks = {}
36
37     # --- ЭТАП 1: Дополнительные фьючерсы (Индексы, Товары, Ставки) ---
38     print("--- ЭТАП 1: Сбор дополнительных фьючерсов ---")
39     extra_dataset = {}
40
41     for name, asset_code in EXTRA_FUTURES_ASSETS.items():

```

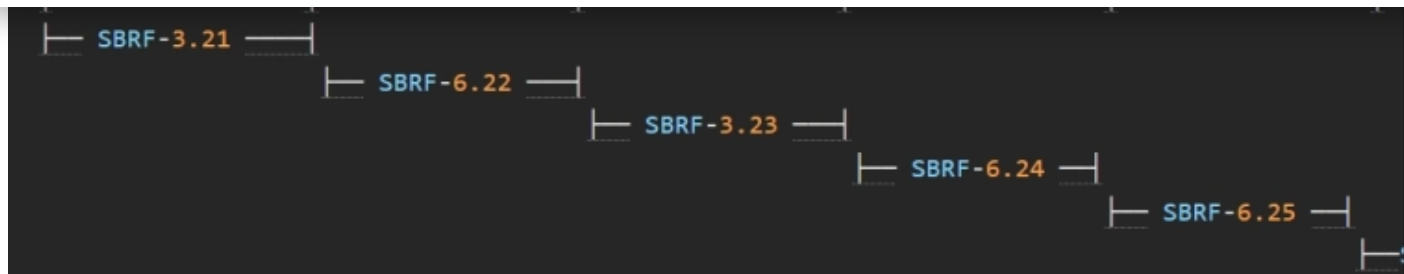
Скрипт поиска связок акция + фьючерс в Visual Studio Code

Через API Московской биржи (ISS) можно бесплатно скачать данные о торгуемых сейчас акциях (если кто-то выбыл со временем, то информации об этом нет — Мосбиржа выдаёт только текущий момент), а затем я сравниваю этот список со списком базовых активов, для которых на срочной секции существуют фьючерсы. В результате получаю именно те акции, для которых одновременно существует и акция и фьючерс.

Дальше мне нужен не просто текущий фьючерс, а история контрактов начиная с 2021 года — фьючерсы имеют ограниченный срок жизни. В отличие от акции, у обычного фьючерсного контракта есть дата экспирации. Например, контракт LKU6 относится к конкретному сроку поставки/исполнения и после окончания обращения перестаёт торговаться. Поэтому история фьючерса — это не один бесконечный временной ряд, а последовательность отдельных контрактов с перекрытием по датам.



Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)



Поэтому для каждого базового актива я отдельно запрашиваю историю и отбираю контракты. А для каждого найденного контракта дополнительно запрашивается расширенная информация через MOEX ISS — его название, группа, дата начала торгов, последняя дата торгов и дата исполнения. В итоге получается не просто список тикеров, а небольшая база всех исторических контрактов, с которой дальше уже можно работать программно.

Дополнительно для акций я получаю через ISS Мосбиржи ИНН эмитента и через поставщика данных DaData подтягиваю название компании, адрес и руководителя. Это не нужно непосредственно для модели, но такой задел на будущее.

Ранее я уже проверял связь российских инструментов с мировыми ценами и увидел, что многие из них действительно быстро реагируют на изменения связанных глобальных рынков. Поэтому в текущем эксперименте я решил добавить несколько внешних факторов: нефть Brent, золото, натуральный газ, опосредованные индексы SPY ETF Trust, QQQ ETF Trust, ну и курсы CNY/RUB и Si, индекса МосБиржи, индекс RGBI (гособлигации).

Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)

```

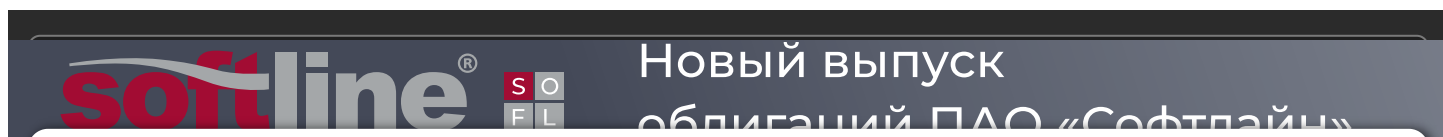
5  # Центр управления настройками: все ключевые параметры, списки активов
6  # и API-ключи находятся вверху файла. Tinkoff полностью удален.
7  """
8  import os
9  from dotenv import load_dotenv
10
11  load_dotenv()
12
13  DADATA_TOKEN = os.getenv('DADATA_TOKEN')
14  |
15  # Год, начиная с которого собираем историю фьючерсных контрактов
16  START_YEAR_HISTORY = 2021
17
18  # Список дополнительных активов (Сырье, Валюта, Индексы, Ставки)
19  # Здесь указаны именно ASSETCODE (коды базового актива на FORTS)
20  # Все символы "/" заменены на "-" для предотвращения ошибок с путями папок
21  EXTRA_FUTURES_ASSETS = {
22      'Brent': 'BR',
23      'Si (USD-RUB)': 'SI',
24      'CNY-RUB': 'CR',
25      'MOEX Index': 'MIX',
26      'RTS Index': 'RI',
27      'RBI Index': 'RBI',
28      'SPY ETF Trust': 'SPYF',
29      'QQQ ETF Trust': 'NASD',
30      'GOLD': 'GOLD',
31      'Natural Gas': 'NG',
32      'Copper': 'CE'
33  }
34
35  # 'WTI': 'WTI', - почти нет данных
36  # 'RUONIA': 'RUON', - почти нет данных
37
38  # МAPPING тикеров фьючерсов (ASSETCODE) на тикеры акций (TQBR)
39  FUTURES_TO_STOCKS_MAP = {
40      'BELU': 'BELU',
41      'GAZR': 'GAZP',
42      'ISKJ': 'ABIO',

```

Дополнительные факторы через фьючерсы Мосбиржи

Весь результат сохраняется в два текстовых JSON-файла: отдельно акции, у которых есть фьючерсы, и отдельно дополнительные фьючерсные активы.

Получается примерно такая схема: сначала Московская биржа отвечает на вопрос «что сейчас вообще торгуется», затем мой код определяет пересечение акций и фьючерсов, после этого восстанавливает историю контрактов и обогащает полученный набор дополнительной информацией.



Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)

```

8940 "moex_data": {
8941   "secid": "WUSH",
8942   "shortname": "ИВУШХолдинг",
8943   "eminent_inn": "9703103060"
8944 },
8945 "daddata_info": {
8946   "name_full": "ВУШ ХОЛДИНГ",
8947   "name_short": "ВУШ ХОЛДИНГ",
8948   "address": "г Москва, ул Долгоруковская, д 21 стр 1",
8949   "management": "Чуйко Дмитрий Владимирович"
8950 },
8951 "contracts": [
8952   {
8953     "ticker": "WUH4",
8954     "name": "WUSH-3.24: Фьючерсный контракт на обыкновенные акции ПАО \"ВУШ Холдинг\"",
8955     "asset_code": "WUSH",
8956     "group": "Акции",
8957     "FRSTTRADE": "2023-06-29",
8958     "LSTTRADE": "2024-03-21",
8959     "LSTDEDATE": "2024-03-22"
8960   },
8961   {
8962     "ticker": "WUH5",
8963     "name": "WUSH-3.25: Фьючерсный контракт на обыкновенные акции ПАО \"ВУШ Холдинг\"",
8964     "asset_code": "WUSH",
8965     "group": "Акции",
8966     "FRSTTRADE": "2024-03-06",
8967     "LSTTRADE": "2025-03-20",
8968     "LSTDEDATE": "2025-03-21"
8969   },
8970   {
8971     "ticker": "WUH6",
8972     "name": "WUSH-3.26: Фьючерсный контракт на обыкновенные акции ПАО \"ВУШ Холдинг\"",
8973     "asset_code": "WUSH",
8974     "group": "Акции",
8975     "FRSTTRADE": "2025-09-04",
8976     "LSTTRADE": "2026-03-19",
8977   }
8978 ]

```

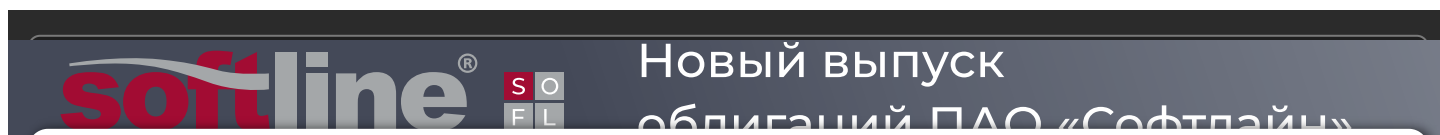
Результат работы скрипта поиска в Visual Studio Code, файл на 9 тысяч строк

Причём всё это сделано отдельным независимым шагом. Если завтра я поменяю список активов или захочу добавить ещё один внешний фактор, мне не приходится переделывать весь последующий ML-пайплайн — достаточно заново сформировать этот начальный слой данных.

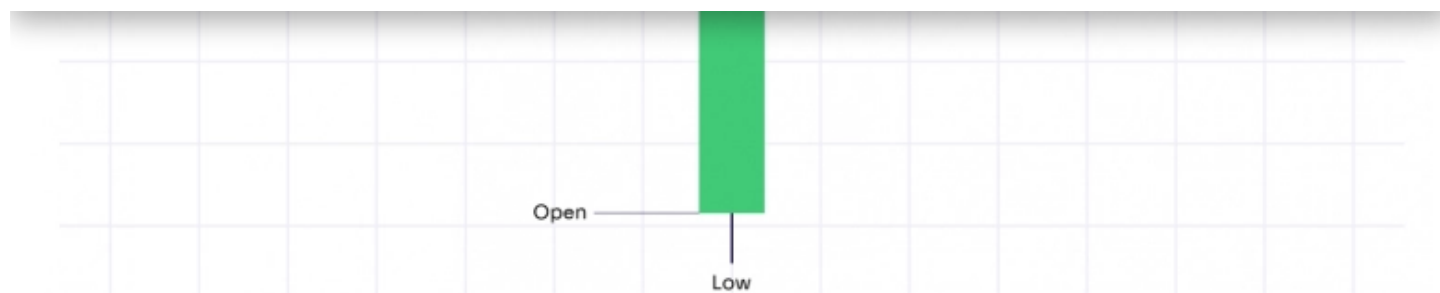
Шаг 2. Погружение в микроструктуру

Вообще чтобы что-то делать с машинным обучением нужны не только список бумаг, но и данные по ним.

Все привыкли к обычным свечам когда есть цена открытия, максимальная цена, минимальная цена, цена закрытия и объём (обозначаются часто как OHLC). Все эти данные нарезаны по времени. Это такой общепринятый формат — по ним рисуется графики и в общем-то всё понятно, но для моего эксперимента с машинным обучением одного OHLCV недостаточно.



Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)

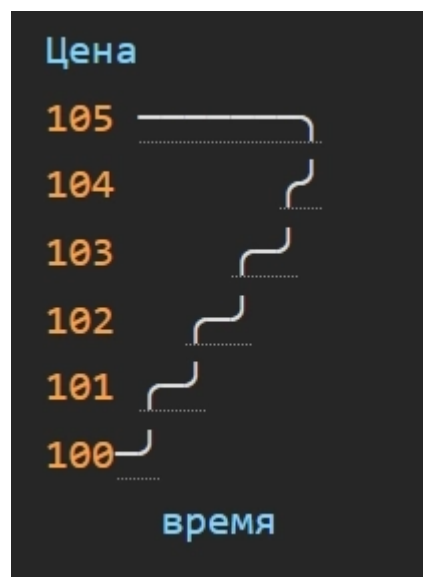


Растущая свеча в вакууме

Потому что внутри одной свечи может происходить много интересного: кто-то совершает агрессивную покупку — то есть исполняет заявку по доступной цене продавца. Кто-то выставляет большое количество лимитных заявок. Кто-то снимает их, не дожидаясь исполнения. В какой-то момент стакан может быть почти пустым, а через несколько миллисекунд в нём появится большой объём. Цена при этом может практически не измениться.

Уточнение: лимитная заявка — это условие «купить или продать не хуже указанной мной цены». Такие заявки образуют стакан — список доступных объёмов на разных ценовых уровнях. Например, путь цены:

Вариант А:



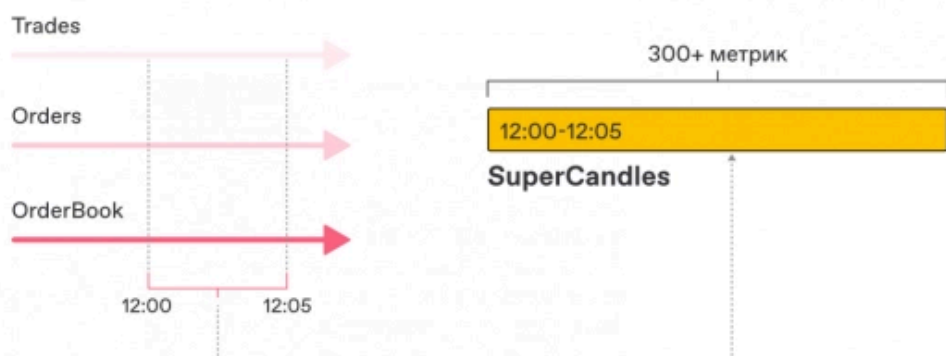
Вариант В:

Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)



То есть две свечи с одинаковыми OHLC могут быть получены совершенно разным поведением участников рынка.

Московская биржа через Algorack предоставляет уже агрегированную информацию о том, что происходило внутри временного интервала 5 минут:



- на потоке сделок «Trades» (средневзвешенные цены, доля покупок и продаж ...)
- на потоке заявок «Orders» (кол-во и объемы новых и снятых заявок)
- на стакане «OrderBook» (дисбаланс покупок и продаж, ликвидность в стакане, спреды)

❗ ПРИМЕЧАНИЕ

Super Candles публикуются каждые 5 минут. Данные публикуются в течение 10-20 секунд. Например, данные за интервал 10:00 - 10:05 доступны в 10:05:12

Исторические данные Super Candles доступны с 2020 года

Algorack работает только на 5 минутках

Внутри алгопака есть:

TradeStats это сделки: это не просто объем, а профиль агрессии. Кто бил по рынку — покупатели или продавцы? Объемы разделены на vol b (покупки) и vol s (продажи),

Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)

OBStats это срез стакана (Order Book): глубина, спреды на 10-м уровне `spread_lv10`, эффективный спред на объем в 1 млн руб `spread_1mio`, а также дисбаланс объемов по всему стакану `imbalance_vol`.

Для модели это потенциально гораздо более интересный источник признаков, чем просто OHLC.

Биржевой стакан:

ПРОДАЖИ	
Цена	Объём
100.04	800
100.03	250
100.02	700
100.01	150
100.00	300
99.99	900
99.98	500
99.97	120
ПОКУПКИ	

← лучшая покупка

Но нужен ли вообще алгопак для рыночной микроструктуры на Московской бирже?

Для моего эксперимента пятиминутная агрегация — довольно крупный интервал. Если я захочу перейти к более быстрому анализу, возникает вопрос: можно ли получить те же показатели самостоятельно из потока сделок и стакана?

И у меня даже произошел спор с товарищем — можно ли собрать такие данные как в Алгопаке самому? И например учить на 5 минутках истории Алгопака, а для реалтайма собирать эти же данные самостоятельно, но из других доступных для частного лица

Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)

И действительно — в значительной степени это возможно. Если у тебя постоянно пишется поток сделок, из него можно восстановить обычную статистику TradeStats: OHLC, объём, оборот, количество сделок, VWAP, объёмы покупок и продаж и другие производные признаки.

Со стаканом ситуация тоже похожая. Если постоянно поддерживать его актуальное состояние, можно периодически, но очень быстро, например 50 миллисекунд, делать снимок и самостоятельно рассчитывать спред, количество уровней, суммарный объём, дисбаланс и многие другие показатели.

Но есть нюанс: стакан — это ещё не история всех заявок. Представим, что на цене 100 рублей в стакане стояло 1000 лотов. Через несколько мгновений осталось 500. Что произошло?

Первый вариант — кто-то снял заявку на 500 лотов.

Второй — кто-то агрессивно купил эти 500 лотов.

Третий — заявка была каким-то образом изменена.

Если у нас есть только агрегированное состояние стакана и лента сделок, однозначно ответить на этот вопрос нельзя.

Именно здесь появляется принципиальная разница между состоянием стакана и историей жизненного цикла отдельных заявок.

В L2-стакане мы видим объём, агрегированный по ценовым уровням. Но у нас нет идентификаторов конкретных заявок и полного жизненного цикла каждой из них:

```
<code class="1c">10.00 → 500
10.01 → 300
10.02 → 700</code>
```

Поэтому самостоятельно получить точное значение `cancel_orders` простым анализом изменений агрегированного стакана нельзя.

История событий с номерами заявок:

```
<code class="1c">10:00:001 ADD id=123 300 @ 10.00
10:00:017 ADD id=124 200 @ 10.00
10:00:024 TRADE id=123 100
10:00:031 CANCEL id=124 200
```

Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)

Номера заявок вроде как можно видеть через QUIK или TRADEAQ, но это не точно — я сам так не пробовал и даже не исследовал вопрос.

И конечно, у Мосбиржи есть полная история всех заявок (Full Order Book, но это терабайты данных (Big Data), с которыми частному лицу работать невероятно сложно и дорого.

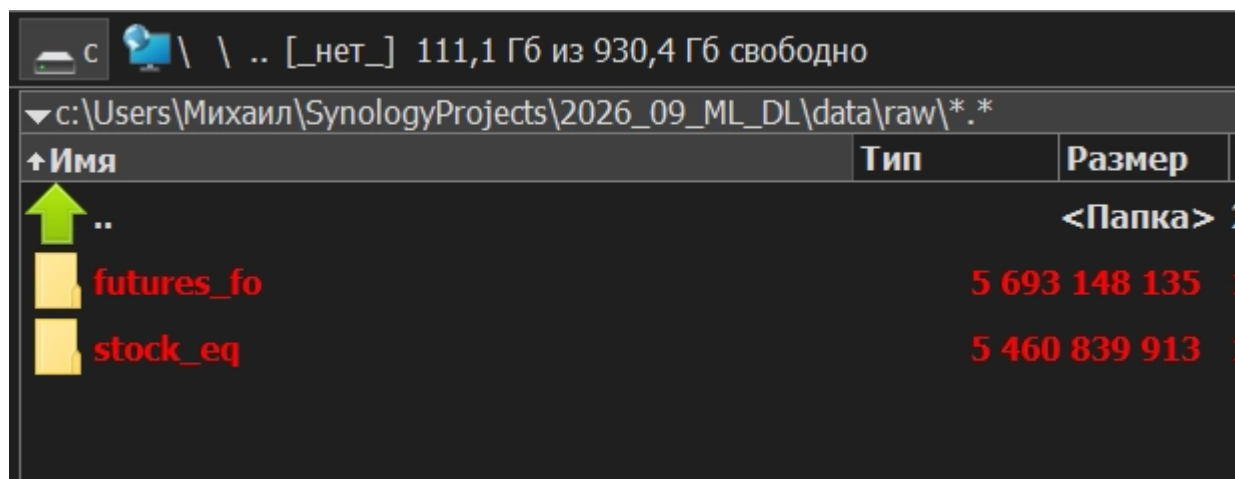
Поэтому здесь мой спор с товарищем закончился: он прав — значительную часть статистики можно посчитать самостоятельно, но это не то же самое, что иметь биржевую информацию о жизненном цикле каждой заявки.

Почему я всё-таки пока выбрал Algorack

Но есть и более прозаичная причина. Чтобы самостоятельно накапливать такую историю, нужно с сегодняшнего дня постоянно записывать рыночные данные. Не просто запустить один скрипт, а обеспечить непрерывную работу инфраструктуры, следить за соединениями, хранить поток данных и проверять, что ничего не потерялось. И если начать это делать сегодня, то завтра у тебя будет один день истории. Через неделю — неделя. А мне нужна история начиная с 2021 года.

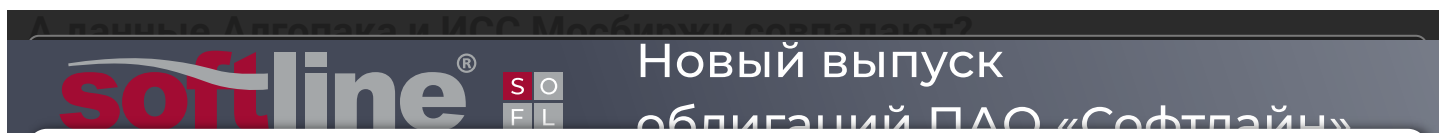
Можно, конечно, построить собственную инфраструктуру и ждать несколько лет. Но для исследовательского проекта это довольно странный способ провести лето.

Algorack в этом смысле решает главную проблему сразу: исторические данные уже накоплены за прошлые годы и доступны для загрузки. Цена вопроса — деньги и определённые ограничения самого продукта.



Имя	Тип	Размер
..	<Папка>	
futures_fo		5 693 148 135
stock_eq		5 460 839 913

данных в формате Parquet с Алгопака Мосбиржи, скачивается кстати многие часы, потому что по кусочкам



Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)

Поэтому я написал отдельный валидатор и решил сравнить пятиминутные данные Algorack с обычными свечами, которые можно бесплатно получить через ISS Московской биржи.



Скришот <https://smart-lab.ru/q/futures/LKU6/>, но на минутках уже не перемотать на июль, потому что у TradingView ограничение по истории: глубина на разных диапазонах разная. Например, для фьючерса LKU6 за 2 июля 2026 года (когда я проводил подробный анализ) только за один день набралось около десяти случаев расхождения.

На первый взгляд некоторые различия были совсем небольшими: например, в 10:05, 13:10, 13:20, 15:25, 17:20 и 17:50 цена закрытия отличалась на один пункт, а объём — всего на два-три контракта.

Ещё оказалось, что при сравнении нужно очень внимательно смотреть на границы интервалов: в Algorack время 09:05 означает фактически интервал с 09:00:00 до 09:04:59; а в стандартных свечах ISS время относится к началу минутного интервала.

А вот большие расхождения уже интереснее. Были и случаи, которые нельзя объяснить одним только пограничным timestamp.

Например, для того же LKU6 в 11:45 один источник показывал объём 135 контрактов, а другой — 229. Разница — 94 контракта. В 12:50 было 22 против 39.

Это уже не ситуация одна сделка попала в соседнюю пятиминутку, хотя писал им в

Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)

задавать модели гораздо более содержательные вопросы:

- «Что происходило с агрессивными покупками?»
- «Увеличивался ли дисбаланс стакана?»
- «Сколько заявок появилось и сколько было снято?»
- «Насколько дорогим было бы исполнение крупной заявки?»
- «Что происходило с ликвидностью непосредственно перед движением цены?»

Пока я не знаю, какие из этих признаков действительно окажутся полезными. И возможно, значительная их часть окажется бесполезной.

Итоги первой части

На этом этапе у меня ещё нет ни модели, ни торгового сигнала, ни ответа на главный вопрос — можно ли вообще заработать на этой информации.

Зато есть потраченное время и главное: исторический набор рыночных данных, собранный в единую структуру.

Я восстановил историю акций и фьючерсов с 2021 года, добавил туда внешние факторы и получили не только цены и объёмы, но и микроструктуру рынка: агрессивные покупки и продажи, заявки, снятия, спреды, глубину стакана и дисбаланс ликвидности.

При этом пришлось отдельно проверять данные Алгопака и искать причины расхождений с ISS.

Теперь данные подготовлены. И следующий вопрос гораздо интереснее: существуют ли в этих данных Московский биржи вообще информация, которую машинное обучение способно превратить в устойчивое торговое преимущество?

Автор: Михаил Шардин

 [Моя онлайн-визитка](#)

 Канал «Умный Дом Инвестора» в [TG](#) или [MAX](#)

25 августа 2026 года

Данная публикация является личным мнением автора. Мнение владельца сайта может не совпадать с мнением автора.

softline® 

Новый выпуск

облигаций ПАО «Софтлайн»

Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)



Михаил Шар...

Пермь

569 6 562

с 23 января 2019

+ Подписаться

114 КОММЕНТАРИЕВ

Сначала старые ▾



Anest

25 августа 2026, 05:07

Садьдо/сантимент по рыночным исполненным заявкам покупки/продажи Вы можете сами с легкостью распарсить информацию из таблицы — «Таблица обезличенных сделок», причем на лету и за любой период, но только в рамках торговой сессии. Со стаканом сложнее, надо будет самому собирать/записывать данные и уже анализировать.

— Показать 6 ответов



AlexShul

25 августа 2026, 05:10

Картинки все рисуете, так и жизнь пройдет мимо))

— Показать 1 ответ



columbus

25 августа 2026, 05:26

„Пока я не знаю, какие из этих признаков действительно окажутся полезными. И возможно, значительная их часть окажется бесполезной. “

Вы и сами знаете, что это не сработает.

— Показать 1 ответ



softline®



Новый выпуск

облигаций ПАО «Софтлайн»

Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)

вещами — брали по очереди в руки свечку и измеряли глубину стакана

— Показать 7 ответов



Ещё 20 комментариев

Напишите комментарий...



ОТПРАВИТЬ

Читайте на SMART-LAB:



Кто не присылает своих клиентов и сотрудников на конференцию смартлаба останется позади индустрии

Наша конфа до сих лучшая, потому что на ней максимально представлены все компании российского рынка . Именно нашу...



13.10.2026



Что изменится в ОСАГО в ближайшее время

Главное ожидаемое нововведение — автоматическая проверка ОСАГО дорожными камерами. Точная дата запуска пока не...



Ренессанс страхование

13:30



Как сделать альфу с помощью искусственного интеллекта

Роботы захватили фондовые рынки. Сейчас нейросети торгуют лучше инвесторов. Как не остаться в стороне? Одни из лучших

softline®



Новый выпуск

облигаций ПАО «Софтлайн»

Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)



Рынок продолжает торговаться в стиле «туда-сюда», но мы то знаем, что Вы подписаны на мозговик чтобы забрать денежки н...



Mozgovik

12.10.2026

Установите приложение Смартлаба:



RuStore



AppGallery



App Store



[О смартлабе](#)

[Реклама](#)

[Полная версия](#)

Источник: [ПАО Московская Биржа](#)

softline®



Новый выпуск

облигаций ПАО «Софтлайн»