

РЕКЛАМА

**Хабр Карьера**  
Актуальные навыки для рынка  
Учитесь тому, за что готовы больше платить

**Хабр** Курсы — инвестиция в себя

еppenoso 13 янв в 05:22

**ML на Мосбирже — почему мой грааль не работает?**

Средний 7 мин 11K

Open source\*, Python\*, Финансы в IT, Машинное обучение\*, Искусственный интеллект

Кейс Сезон ИИ в разработке

Время после нового года решил провести с пользой и окунуться в машинное обучение. Заняться Machine Learning — и посмотреть получится что-то или нет с российским рынком акций на Московской бирже.

Моей целью было построить такую систему, которая будет учиться на истории и в перспективе торговать лучше чем случайное блуждание 50/50. Но из-за комиссий и спреда подобные блуждания изначально отрицательны — чтобы выйти в плюс надо как минимум покрывать комиссии.

Если говорить о результатах очень кратко, то технически всё работает, но вот финансовый результат на грани безубыточности.

Если Вы только интересуетесь этой темой Вы можете посмотреть какие-то шаги в моей статье, а если Вы уже опытный разработчик подобных систем, то можете подсказать что-нибудь в комментариях.

Причём вся эта работа выглядит совершенно не так как показывается в фильмах про уолл-стрит: фактически это написание скриптов и монотонный запуск и всё происходит полностью локально на компьютере.

Скриншот VS Code открытым деревом проекта

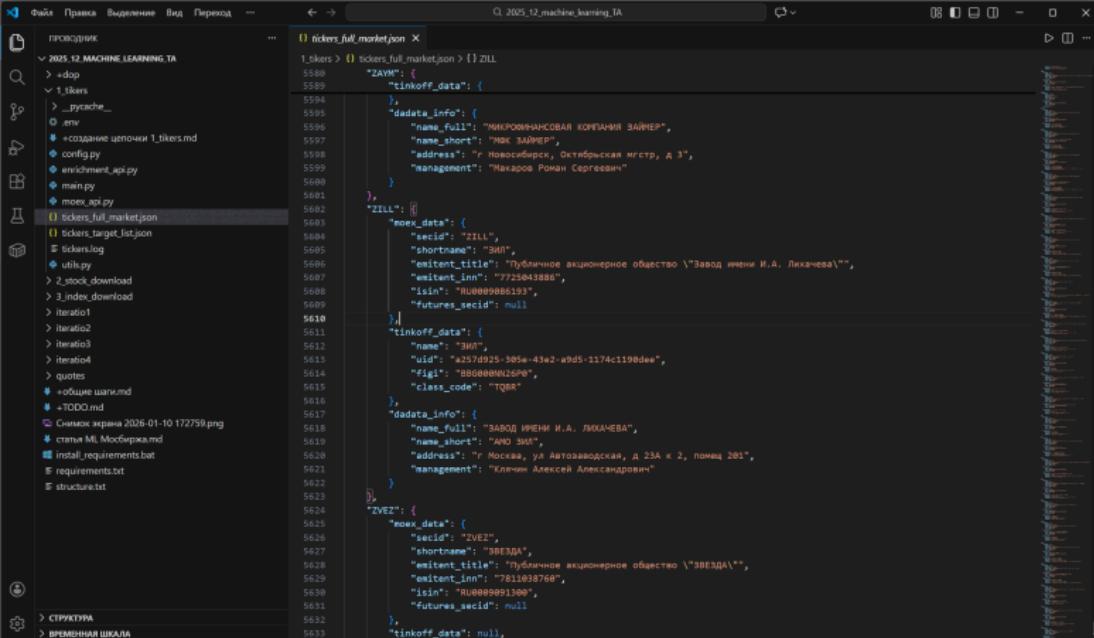
## Охота за данными

Вообще данные очень важны. Иначе получается что мусор на входе просто перетекает в мусор на выходе.

Для отбора бумаг я руководствовался следующей логикой: сначала ищу общий список всех акций, торгуемых на Московской бирже и делаю выборку — в неё входят только те, у которых есть фьючерсы. Дальше оставляю только тикеры акций, которые имеют фьючерсы.

Потом беру два API — одно от брокера, а второе API, предоставляющее информацию по всем юридическим лицам России — это API DaData. У каждой акции ведь есть ИНН компании.

Используя оба этих API — оба бесплатных кстати — обогащаю каждую запись дополнительными сведениями.



```

{
  "ZAVM": {
    "tinkoff_data": {
      "name": "НИКРОФИНАНСОВАЯ КОМПАНИЯ ЗАЙМЕР",
      "shortname": "НОК ЗАЙМЕР",
      "address": "г. Новосибирск, Октябрьская магистр, А 3",
      "management": "Макаров Роман Сергеевич"
    },
    "daidata_info": {
      "name_full": "Публичное акционерное общество \"Завод имени И.А. Лихачева\"",
      "name_short": "ПАО ЗМП",
      "existant_title": "7725043886",
      "existant_inn": "110000001919",
      "isin": "RU0000001260",
      "futures_secid": null
    }
  },
  "ZILL": {
    "tinkoff_data": {
      "name": "ЗМП",
      "shortname": "ЗМП",
      "uid": "a37d095-085e-43e2-a9d5-1174c1190d0e",
      "figi": "BBG000001260",
      "class_code": "TQR"
    },
    "daidata_info": {
      "name_full": "ЗАВОД ИМЕНИ И.А. ЛИХАЧЕВА",
      "name_short": "ЗАО ЗМП",
      "address": "г. Москва, ул. АвтоАвтодорская, д 23А к 2, помещ 201",
      "management": "Кличкин Алексей Александрович"
    }
  },
  "ZVEZ": {
    "tinkoff_data": {
      "name": "ЗВЕЗДА",
      "shortname": "ЗВЕЗДА",
      "existant_title": "Публичное акционерное общество \"ЗВЕЗДА\"",
      "existant_inn": "7811038760",
      "isin": "RU0000001300",
      "futures_secid": null
    },
    "daidata_info": {
      "name_full": "Публичное акционерное общество \"ЗВЕЗДА\"",
      "name_short": "ЗАО ЗВЕЗДА",
      "existant_title": "7811038760",
      "existant_inn": "7811038760",
      "isin": "RU0000001300",
      "futures_secid": null
    }
  }
}

```

Фрагмент файла `tickers_full_market.json` с общим списком тикеров

После этого скачиваю котировки акций с Московской биржи. И загружаю три индекса: IMOEX, IMOEX2, RTSI, RGBI.

Для этого скачиваю минутки — они готовые и сразу в архивах через API брокера — можно очень быстро скачать полностью все бумаги.

С индексами сложнее. Брокерский API не отдавал историю по IMOEX и RGBI, видимо, из-за ограничений лицензии биржи (привет, MOEX AlgoPack). Пришлось писать парсер для прямого API Московской биржи (ISS MOEX) — скорость гораздо медленнее, но я скачиваю тоже минутки.

Скачать все доступные минутки с IMOEX, IMOEX2, RTSI, RGBI занимает около 20 минут.

Котировки индекса

## Работа со временем

Я начал работу с того, что выделил три интересующих меня таймфрейма. 15 минут, 1 час и 1 день, основной 1 час.

Написал скрипт который преобразует архивы с минутками от API брокера в файлы с котировками нужных таймфреймов.

### Итерация 1,3,4: временные бары (обычные)

Сначала я просто агрегировал минутки в 15 минут, 1 час и 1 день через скрипт, но рынок неравномерен. Возможно для ML-модели временные свечи могут быть очень шумные, так как содержат разное количество информации.

Скрипт normalization\_quotes\_stock.py который читает 1-минутные архивы (ZIP) из quotes/1min\_zip и агрегирует данные в 15min, 1h, 1d

## Итерация 2: свечи рублевого объема

Потом я временно перешёл к событийным барам. Новая свеча формируется не по времени, а когда проходит фиксированный объем в рублях, например, 50 млн руб., но для разных акций этот порог разный, потому что рассчитывается динамически. Разделил их по условным классам:

- A: порог выше.
- B: порог ниже.
- C: отбрасываются.

Скрипт который создаёт из минутных котировок свечи рублевого объема и классифицирует бумаги по классам

## Учитель для робота. Разметка

Как объяснить машине, что такое «хорошая сделка»? Рынок может продолжать рост, но перед этим выбить по стопу.

Книга Маркос Лопез де Прадо «Машинное обучение: алгоритмы для бизнеса»

#### Книга «Машинное обучение: алгоритмы для бизнеса»

Привет, Хаброжители! Маркос Лопез де Прадо делится тем, что обычно скрывают, — самыми прибыльными ал...

[habr.com](https://habr.com)



Для установки цели я использовал тройной барьер по де Прадо:

- Верхний барьер (Take Profit): Цена + N×ATR

- Нижний барьер (Stop Loss): Цена -  $M \times ATR$
- Вертикальный барьер (Time Limit): Если прошло 100 баров, а цена никуда не пришла — выходим.

Скрипт расставляет метки:

Метка 1: сработал Take Profit.

Метка 0: вышло время, выход в ноль (минус комиссия).

Метка -1: сработал Stop Loss, потеря денег.

Фрагмент в Visual Studio Code

## Глаза модели: инженерия признаков

Я использовал [CatBoost \(Categorical Boosting\)](#) это библиотека машинного обучения с открытым исходным кодом от «Яндекса», основанная на градиентном бустинге над деревьями решений. Я не подаю сырье цены (OHLCV: Open (цена открытия), High (максимальная цена), Low (минимальная цена), Close (цена закрытия) и Volume (объем торгов)), так как они не стационарны, потому что цена 100 ₽ в 2010 и 100 ₽ в 2024 — это разные сущности, а использую только относительные величины.

Я старался подавать именно Log Returns (логарифмические доходности), потому что Log returns аддитивны и симметричны: падение на 50% и рост на 100% имеют одинаковый масштаб в логарифмах.

Сами признаки, в разных итерациях по разному было:

- Качество импульса: не просто «цена выросла», а как она выросла. Использую автокорреляцию и эффективность тренда.
- Микроструктура: что происходило внутри этой свечи рублевого объема? Какая концентрация объема?

- Межрыночные связи: как актив ведет себя относительно индекса Мосбиржи и индекса гособлигаций RGBI.
- Классика: RSI (нормализованный через Z-score), расстояния до скользящих средних.

Файл для генерации признаков и разметки

## Моделирование

Для того чтобы заниматься ML все инструменты у меня были: компьютер 32 Гб оперативки с видеокартой GPU 16 Гб, Python как основной язык, Docker чтобы не зависеть от капризов драйверов, Numba для ускорения расчётов, Linux для администрирования Docker контейнеров.

Нейросети как консультант: как я нашел и заказал ПК без подсветки для локальной работ...

Внимание, знатоки! Если вы профессионально разбираетесь в компьютерном железе, тонкостях сборки ИЛИ ...

habr.com



Для моей задачи как будто даже избыточная конфигурация — потому что все расчёты протекают очень быстро.

Dockerfile

Делаю это из под Ubuntu, хотя всю разработку веду из-под Windows.

01\_build.sh

Ещё использую измененную версию пошагового тестирования (Walk-Forward Optimization), которая используется в трейдинге для поиска и проверки торговых стратегий, но добавляя «очистку» (purging) данных: она избегает перекрытия обучающих и тестовых периодов, чтобы предотвратить подгонку стратегии под шум истории, делая результаты более реалистичными и устойчивыми к будущим изменениям рынка.

### Результаты и боль

При просмотре результатов магия машинного обучения быстро испаряется.

Я учитываю комиссии:

```
COMMISSION_PCT = 0.04 / 100 # комиссия брокера
SLIPPAGE_PCT    = 0.02 / 100 # проскальзывание
```

Объяснить код с  SourceCraft

Каждая сделка автоматически теряет 0,06% или 0,12% на круг для акций. Кажется мелочью, но при сотнях сделок именно эти десятые доли процента превращают модель в убыточную. Модели нужно предсказывать движения >0,3-0,5%, чтобы быть в плюсе.

По качеству прогнозов мой ML стабильно показывает AUC 0,54–0,55. Формально это лучше случайного угадывания (0,50), но до Грааля тут очень далеко. Почему? Даже небольшая ошибка в вероятностях, умноженная на комиссии и шум, быстро съедает весь перевес.

Чтобы понимать,  то именно я меряю, важно разобраться в трёх ключевых метриках.

AUC (Area Under the Curve) — это мера того, насколько хорошо модель умеет отличать «хорошие» сделки от «плохих». Если AUC = 0,5, модель — это монетка. Если 0,55 — она угадывает чуть чаще, чем случайность. В вакууме это звучит неплохо, но в трейдинге такого преимущества часто недостаточно, чтобы перекрыть издержки.

Fold — это один из прогонов в Walk-Forward Optimization. История рынка режется на последовательные отрезки: на одном модель учится, на следующем тестируется. Каждый такой отрезок — отдельный fold. Это имитация реальности: мы всегда торгуем на будущем, которого модель «не видела». Поэтому один fold может быть прибыльным, а следующий — убыточным, просто потому что режим рынка поменялся.

Тесты

Precision (точность) — это ответ на вопрос: «если модель сказала „купуй“, как часто она оказывается права?». Это критично для торговли, потому что даже модель с неплохим AUC может генерировать кучу ложных сигналов, которые будут съедать депозит комиссиями и стоп-лоссами.

И вот здесь появляется самая болезненная часть. На одном из фолдов я получаю красивую Equity Curve, где капитал растёт.

График fold\_2\_best\_equity

На другом — та же самая модель превращает счёт в “пилу”: заработали на тренде, потом долго и мучительно всё отдали на боковике.

График fold\_3\_best\_equity

Когда я писал этот текст мне в голову пришло, а что если модель обучать на максимизацию финансового показателя, такого как Коэффициент Шарпа?

### Заключение: вопросы к залу

Я проделал некоторую работу, но результаты пока выглядят не особо приятными.

Мои гипотезы:

1. Ошибка в методологии?
2. Мало данных?
3. Предсказывать не направление, а волатильность?
4. Перейти на более высокие таймфреймы (4Н, 1D), где комиссия съедает меньшую долю движения?
5. Нужно использовать данные из стакана (Order Book)? С получением истории стакана для частного лица большие проблемы. Бесплатно доступен лишь очень ограниченный набор инструментов.
6. CatBoost слишком прост, нужны трансформеры?

Я занимаюсь Machine Learning (ML), когда система учится на таблицах. Но есть ведь ещё Deep learning (глубокое обучение) когда идёт анализ больших объёмов данных и выявления сложных закономерностей автономно. Но боюсь для моей задачи слишком мало данных. Наверное только на истории стаканов OrderBook будет работать.

Google Maps для биржевого стакана: пишем визуализатор данных Московской биржи на R...

Обычный трейдер смотрит на свечной график, но свеча — это уже тень прошлого, постфактум. Между тем н...

habr.com



Приглашаю в комментарии: кто реально запускал Machine Learning на Мосбирже в плюс? Или просто опытных людей. Где я свернул не туда?

Не сдерживайте себя — напишите комментарий — буду рад любой критике в комментариях.

**Автор:** Михаил Шардин

[Моя онлайн-визитка](#)

[Telegram «Умный Дом Инвестора»](#)

13 января 2026 г.

**Теги:** [machine learning](#), [deep learning](#), [московская биржа](#), [мосбиржа](#), [аис](#), [Сезон ИИ в разработке](#), [алгоритмическая торговля](#)

**Хабы:** [Open source](#), [Python](#), [Финансы в IT](#), [Машинное обучение](#), [Искусственный интеллект](#)

### Редакторский дайджест

Присыпаем лучшие статьи раз в месяц

Электропочта

Подписаться

Оставляя почту, я принимаю [Политику конфиденциальности](#) и даю согласие на получение рассылок



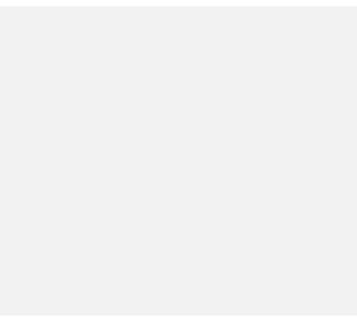
256

92.1

Карма    Общий рейтинг

**Михаил Шардин** @empenoso

Автоматизация / Data &amp; ML / Финансы / Smart Home

[Подписаться](#)[Сайт](#) [Сайт](#) [GitHub](#)[Комментарии 31](#)

## Публикации

[ЛУЧШИЕ ЗА СУТКИ](#) [ПОХОЖИЕ](#)**David\_Osipov**

19 часов назад

### Как «серый» бизнес Femida Search строит зомби-ферму внутри Хабра для обхода Песочницы

[Простой](#)[4 мин](#)[11К](#)[Репортаж](#)[+77](#)[16](#)[51](#)**monobogdan**

8 часов назад

### Как я на КПК оперативную память увеличивал [Длиннопост про железо]

[Простой](#)[6 мин](#)[6.6К](#)[Ретроспектива](#)[+34](#)[6](#)[5](#)**AlexeyPolunin**

20 часов назад

### Как я научился без скандалов выходить из зомби-проектов систем автоматизации

[9 мин](#)[8.7К](#)[+32](#)[31](#)[6](#)

 shadowform  
4 часа назад

## История о том, как я пытался подключиться к Starlink в России. История полная приключений

 Средний  16 мин  4K

Из песочницы

 +31 21 18 Alekseyl  
22 часа назад

## Как мы сократили объем данных в 10 раз, не повредив пользовательскому опыту, или переезд Postgres → ClickHouse

 Средний  13 мин  10K

Кейс

 +29 39 7 HappyTalkie  
9 часов назад

## Как мы продавали компьютеры в 90-х. Часть #02

 6 мин  4.4K +20 1 5 DRoman0v  
3 часа назад

## Умное освещение в доме на базе Philips Hue. Личный опыт

 7 мин  2.5K +15 7 1 Laborant\_Code  
22 часа назад

## От пустоты к идее: как я создал свою первую доску вдохновения

 Простой  27 мин  6.9K

Обзор

 +14 11 0 beeline\_cloud  
19 часов назад

## «Галя, у нас замена»? Почему хайп со сменой программистов на системы ИИ — далеко не первая «паническая атака» в отрасли

 7 мин  23K

Аналитика

 +12 32 81



iMonin

18 часов назад

## Как вихревая трубка Ранка-Хилша превращается в «Вихревой вакууматор»

11 мин

6.8K

+11

4

2

## Подход к облаку, где сложность остаётся на стороне провайдера

Турбо

Показать еще

### ИСТОРИИ



Год облака в подарок: 2 VM, БД, 100 ГБ



Годнота из блогов компаний



Через 3, 2, 1...



Подборка статей и советов о психологических проблемах в работе



Смотри в оба



Ошибки разума

Статьи о когнитивных искажениях

### ВАКАНСИИ

Deep Learning Engineer (GigaChat Prod)

от 350 000 ₽ · Сбер · Москва

Senior ML-инженер

от 400 000 до 600 000 ₽ · DataLouna · Можно удаленно

Python Backend Developer

от 250 000 до 500 000 ₽ · Hard Bootstrapping LLC · Санкт-Петербург

Middle/Senior Data Scientist (команда ИИ и аналитика)

от 200 000 до 350 000 ₽ · Сбер · Москва

ML Data engineer (LLM Gigachat)

от 200 000 ₽ · Сбер · Москва

Больше вакансий на Хабр Карьере

### МИНУТОЧКУ ВНИМАНИЯ



Как я запустил ComfyUI и взял AI-графику под контроль

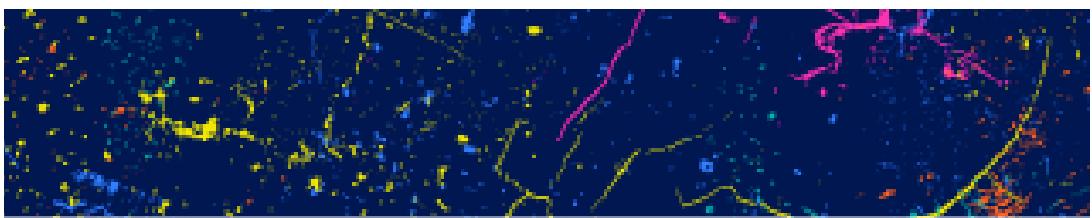


Облако, где инфраструктура становится невидимой

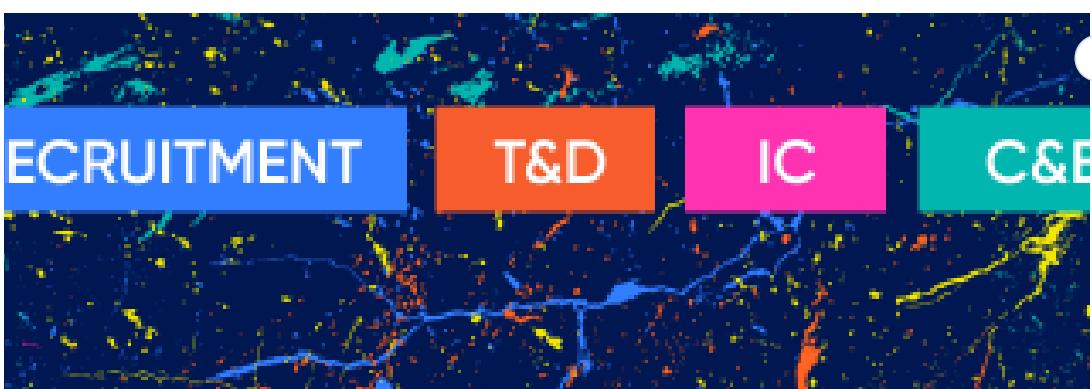


Как меняется работа в data: разобрали по модели айсберга

#### БЛИЖАЙШИЕ СОБЫТИЯ



## ПРАКТИКА HR 2026



5 марта

Конференция «ПРАКТИКА HR 2026» — 4 события по ключевым HR-направлениям

Москва

Другое

Больше событий в календаре

Хабр



Настройка языка

Техническая поддержка

© 2006–2026, Habr