



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

```

{
    "ZAO NT": {
        "tinkoff_data": {
            "dadata_info": {
                "name_full": "МИКРОФИНАНСОВАЯ КОМПАНИЯ ЗАЙРЕН",
                "name_short": "МФК ЗАЙРЕН",
                "address": "г. Новосибирск, Октябрьская улица, д 57",
                "management": "Макаров Роман Сергеевич"
            },
            "moex_data": {
                "seed_id": "ZILL",
                "shortname": "ЗМФ",
                "emitter_title": "Губернское акционерное общество \\"Завод имени И.А. Лихачева\\"",
                "emitter_inn": "7725043886",
                "isin": "RU0000088193",
                "futures_seed_id": null
            }
        },
        "tinkoff_data": {
            "name": "ЗМФ",
            "value": "a257d925-305e-43e2-a9d5-1174c1190dee",
            "fifig": "8800000000000000",
            "class_code": "TQTR"
        },
        "dadata_info": {
            "name_full": "ЗАВОД ИМЕНИ И.А. ЛИХАЧЕВА",
            "name_short": "ЗМФ",
            "address": "г. Москва, ул. Автозаводская, д 23А к 2, помещ 201",
            "management": "Ключев Алексей Александрович"
        }
    },
    "ZMEZ": {
        "moex_data": {
            "seed_id": "ZVEZ",
            "shortname": "ЗБЕЗДА",
            "emitter_title": "Публичное акционерное общество \\"ЗБЕЗДА\\"",
            "emitter_inn": "7731038760",
            "isin": "RU0000091300",
            "futures_seed_id": null
        },
        "tinkoff_data": 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.

**Пикабу Игры**  
 +1000 бесплатных онлайн игр

**Пикаджамп**  
 Аркады, Казуальные, На ловкость

[Играть](#)

### Топ прошлой недели

Animalrescueed  
41 пост

a.lisik  
3 поста

Oskanov  
8 постов

[Посмотреть весь топ](#)

**Лучшие посты недели**

Рассылка Пикабу:  
отправляем самые  
рейтинговые материалы за 7  
дней 🔥

Укажите  
**Подписаться**

Нажимая «Подписаться»,  
я даю согласие на [обработку](#)  
[данных](#) и [условия почтовых](#)  
[рассылок](#).

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.

Помощь	Правила
Кодекс Пикабу	соцсети
Команда	О
Пикабу	рекомендация
Моб.	Х
приложение	О компании

[Промокоды Биг Гик](#)  
[Промокоды Lamoda](#)  
[Промокоды МВидео](#)  
[Промокоды Яндекс Маркет](#)  
[Промокоды Пятерочка](#)  
[Промокоды Aroma Butik](#)  
[Промокоды Яндекс](#)  
[Путешествия](#)  
[Промокоды Яндекс Еда](#)  
[Постила](#)  
[Футбол сегодня](#)



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:

- А: порог выше.
- В: порог ниже.
- С: 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 выбрать по стопу.

Книга Маркос Лопез де Прадо «[Машинное обучение: алгоритмы для бизнеса](#)»

Для установки цели я использовал тройной барьер по де Прадо:

- Верхний барьер (Take Profit): Цена + NxATR
- Нижний барьер (Stop Loss): Цена - MxATR
- 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.

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
```

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 Граала тут очень далеко. Почему? Даже небольшая ошибка в вероятностях, умноженная на комиссии и шум, быстро съедает весь перевес.

Чтобы понимать, что 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 „покупай“, 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 (4H, 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.

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.



1 😊



12



РЕКЛАМА · 16+

**RESTAFF**

restaff.work

Платите самозанятым и принимайте НДС к вычету - ReStaff

⭐+ Подробная консультация и демо бесплатно!

Получить консультацию

РЕКЛАМА · 16+

**SourceCraft**

sourcecraft.dev

Грант на 6 000 ₽ для деплоя в облако сайта или бота!

Узнать больше



Программирование на python  
942 поста · 12K подписчика

Добавить пост

Подписаться

...

Правила сообщества

Публиковать могут пользователи с любым рейтингом. Однако!

Приветствуется:...

Подробнее ▾

[Все комментарии](#)   [Автора](#)

Раскрыть 12 комментариев

Чтобы оставить комментарий, необходимо [зарегистрироваться](#) или [войти](#)

