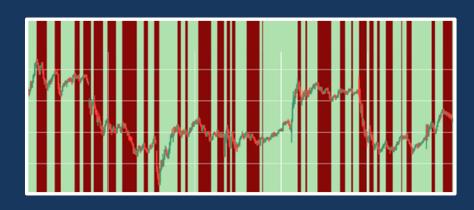
«Если ты не найдёшь способ зарабатывать деньги, пока спишь, то ты будешь работать, пока не умрешь» © Уоррен Баффет



КАК СОЗДАВАТЬ ТОРГОВЫЕ И ИНВЕСТИЦИОННЫЕ СТРАТЕГИИ ?

HOW-TO-GUIDE

# Содержание

3 Введение

### 4 ПАЙПЛАЙН

- 4 Каков порядок создания стратегий?
- 5 Тестирование гипотез
- 11 Генерация датасетов
- 18 Обучение нейронных сетей
- 20 Бек-тесты
- 25 Формирование портфеля

27 Торговля



### Введение

Команда NullPointerExeption приветствует Вас, уважаемый читатель, на страницах этого гайдбука по созданию торговых и инвестиционных стратегий. Данный гайдбук создан для хакатона Московской биржи (МОЕХ). Данный гайдбук поможет Вам с одной стороны понять каким образом в целом подходить к созданию и использованию торговых и инвестиционных стратегий, а с другой стороны как это можно сделать с применением разработок команды NullPointerExeption, созданных специально для хакатона.

Все разработки нашей команды вы можете скачать из репозиториев ниже:

- Микросервисы пайплайна разработки ML | AI стратегий: ML|AI: <a href="https://github.com/powerlife000/moex\_hackaton">https://github.com/powerlife000/moex\_hackaton</a>
- API back-end cepsep: https://github.com/DaniilVdovin/goalgomoex\_master\_server
- API front-end: https://github.com/Innosan/go\_algo\_web
- Торговая система: https://github.com/powerlife000/moex\_trading\_system
- Сервис формирования портфеля: https://github.com/sergeycommit/algopack\_top5\_service

Прежде чем мы перейдем к описанию сути каждого из этапов создания и использования торговых и инвестиционных стратегий, сделаем небольшую ремарку. Как вы понимаете, в рамках хакатона, с учётом недостатка времени, можно создавать только упрощённые системы в разных смыслах этого слова. В связи с этим, повествование в гайдбуке будет изложено немножко шире, и выйдет за рамки возможностей указанных выше исходных годов, чтобы вы могли создавать кастомные разработки для своих собственных целей.

Под конец этого введения, наша команда NullPointerExeption искренне желает всем Вам побед на фондовых рынках!

Удачи!

### Каков порядок создания стратегий?

Многие трейдеры и инвесторы с которыми я общался имеют весьма смутное представление о самом процессе создания торговых и инвестиционных стратегий. Как правило многие приходят и говорят: «Эй, я хочу обучить нейронные сети и получить кучу прибыли». К сожалению (а может быть к счастью), процесс создания торговых и инвестиционных алгоритмов весьма нетривиален. Он имеет свою логику и порядок, и мы остановимся на нем поподробнее.

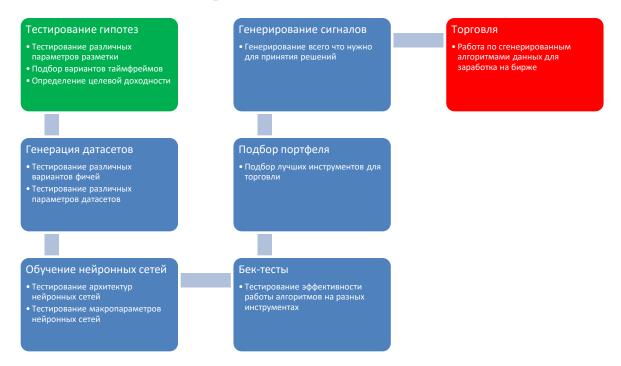


Рисунок 1 — Пайплан разработки и использования торговых и инвестиционных стратегий

Взгляните на рисунок 1, на котором изображена последовательность разработки торговых и инвестиционных стратегий (начинается с зелёного кубика, и заканчивается красным). Как вы можете видеть, в целом этапов разработки не так уж и мало. Но я вас уверяю, что каждый из этапов имеет достаточно большую глубину и множество подэтапов, которые Вам придётся пройти, чтобы вы могли создать свою доходную стратегию. И чтобы вы смогли это сделать, мы достаточно подробно остановимся на каждом из них.

### Тестирование гипотез

Любое действие, которое вы когда-либо делаете должно быть осмысленным и целесообразным. Особенно, когда оно сопряжено с риском и финансовыми затратами. Это же касается и работы на фондовом рынке. Построение торговых и инвестиционных стратегий должно начинаться с постановки и ответа на следующие вопросы:

- Что у вас есть сейчас??
- Что вы хотите??
- Каковы ваши ограничения??

Тестирование гипотез позволяет Вам подобрать такой вариант, который бы удовлетворял ваши амбиции, с одной стороны, но при этом подходил по ряду условий с другой стороны.

Сущность процесса тестирования гипотез весьма проста вне зависимости от алгоритма, который вы тестируете. Вам нужно оценить потенциальную работоспособность Вашего алгоритма на каком-то из финансовых инструментов.

В рамках хакатона Московской биржи наша команда реализует алгоритм отработки различной волатильности. Поэтому, для тестирования гипотез мы оцениваем потенциальную доходности отработки волатильности на разных инструментах и таймфреймах. Вы можете использовать для тестирования гипотез как Front-end, так и микросервис, разработанный на python: «https://github.com/powerlife000/moex\_hackaton/blob/main/data\_markup.py»'

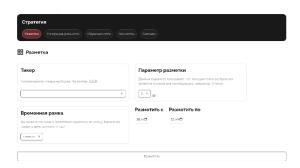


Рисунок 2 – Элемент интерфейса тестирования варинатов разметок

### Тестирование гипотез

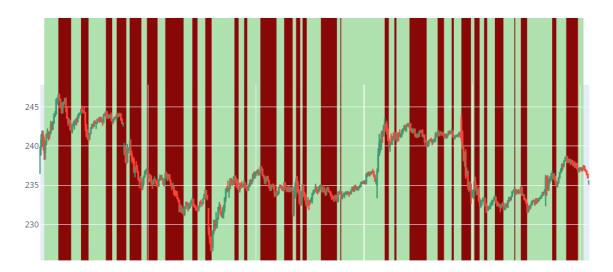


Рисунок 3 – Пример варианта тестируемой разметки

Взгляните на рисунок 3. Вне зависимости от таймфрейма и параметров периода разметки, визуально тестируемая гипотеза выглядит именно так. Зеленым показаны области покупки, красным показаны области продажи актива. Как вы понимаете, максимальная доходность достигается тогда, когда вы покупаете и продаете в наивыгоднейшей точке.

Мы оцениваем два вида потенциальной досходности:

- Теоретическая доходность (по точкам экстремума). Это идеальная недостижимая доходность.
- Теоретическая доходность при смещении точек покупки и продажи на 1 бар.

### Тестирование гипотез



Рисунок 4 — Примеры вариантов окон для расчёта доходности: A — по точкам экстремума; B — при смещении относительно точек экстремума на 1 бар



Рисунок 5 — Разница в доходности при смещении на 1 бар

Как вы понимаете, если нейронная сеть не будет успевать среагировать, то доходность будет далека от идеальной.

### Тестирование гипотез

Взгляните на рисунок 5, желтым показаны потенциальные риски по цене закрытия, которые связаны с реакцией системы, если мы опаздываем на 1 бар. Зеленым показана возможная доходность при данном опоздании. Как вы понимаете, чем выше инерционность системы, тем ниже ваша доходность.

**Важный момент!** Вы уже поняли, что реагировать нужно как можно быстрее. Желательно вообще в точке разворота. И весьма логичным является переход в разметке от старшего к младшему таймфрейму. То есть, если вы пытаетесь отработать волатильно на таймфрейме 1д, давайте в разметке перейдем к таймфрейму 1h. Или 1h->10m.

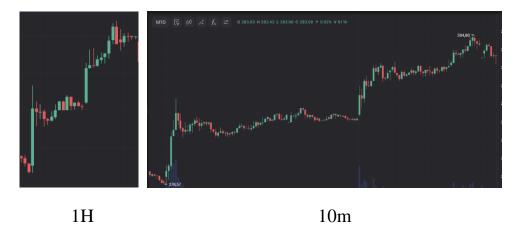


Рисунок 6 – Пример графика одного и того же окна на разных таймфреймах

Если мы увеличим разрешение данных, то вроде бы у нас должно быть теперь всё хорошо. Верно? Обучим нейронную сеть, будем зарабатывать деньги.

Но не все так просто. Обученная нейронная сеть на 10m может показывать значительно худшие результаты. И здесь как раз для вас поле R&D для подбора наиболее оптимальных параметров разметки при достижении вашего целевого уровня доходности.

### Тестирование гипотез

Хорошо. Допустим вы определились с разметкой. Что же делать дальше? Указанная разметка должна быть преобразована в сигналы, на которых будет обучаться нейронная сеть. И это очень важный момент, которому вы должны уделить самое пристальное внимание. Давайте рассмотрим его подробнее.



Рисунок 7 – Пример размечаемой волны

Предложим у нас есть некая абстрактная волна, которую мы хотим отработать (рисунок 7). Зеленым показана область покупки, а красным область продажи. Что мы должны скормить нейронной сети, чтобы она на этом обучилась? У нас есть разные варианты.

**Вариант 1.** Он же самый простой. Мы размечаем области покупки и продажи as-is. То есть, например 0 – это область продаж, 1 – область покупок. То есть у вас всего 2 сигнала: buy, sell.

**Вариант 2.** Посложнее. У вас может быть уже 3 сигнала: buy - 2, hold - 1, sell - 0.

### Тестирование гипотез



Рисунок 8 – Разметка сигналов по 2 варианту

**Вариант 3**: Еще сложнее. Здесь может быть уже 4 сигнала: buy -2, buy-hold 3, sell -0, sell-hold -1.



Рисунок 8 – Разметка сигналов по 3 варианту

Как вы понимаете, вариантов может быть множество. Причем, каждый использует что-то своё. И я скажу, что в разных стратегиях показывают эффективность разные варианты разметки.

В наших микросервисах (<a href="https://github.com/powerlife000/moex\_hackaton">https://github.com/powerlife000/moex\_hackaton</a>) мы использовали 2-ой вариант. Но как вы понимаете, вы должны исследовать различные варианты разметки для вашей стратегии. Например, какую долю от общей волны должен занимать сигнал hold? Этот ответ покажет только тестирование гипотез.

### ГЕНЕРАЦИЯ ДАТАСЕТОВ

Допустим вы определились с разметкой сигналов. Вы молодец и проделали огромную работу. Но она будет впустую без признаков, на которых ваша нейронная сеть будет обучаться. Признаки (или фичи, как их еще называют), являются сердцем вашей нейронной сети. Их генерации нужно уделить самое пристальное внимание.

Давайте остановимся на этом поподробнее.

В основе признаков датасета лежит логика принятия решений. А логика принятия решений может быть разная. Вы можете открыть интернет и посмотреть, каждый работает как хочет в меру своих предпочтений. Можно выделить различные виды анализа и способов принятия решений на их основе



Рисунок 9 – Примеры видов анализов

Как вы понимаете, признаки логики работы тех или иных видов анализов могут быть заложены в датасеты.

### ГЕНЕРАЦИЯ ДАТАСЕТОВ

В нашем микросервисе генерации датасетов (<a href="https://github.com/powerlife000/moex\_hackaton/blob/main/data\_gen.py">https://github.com/powerlife000/moex\_hackaton/blob/main/data\_gen.py</a>) мы заложили множество вариантов технического анализа.

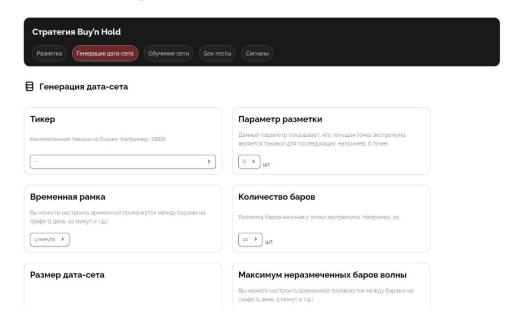


Рисунок 10 – Генерация датасетов на фронте

Если вы захотите использовать наши сервисы как с Front-end (рисунок 10), так и напрямую, то всего будет сгенерировано порядка 2600 признаков, которые учитывают тренды, волатильность, объемы, дивергенции, свечи, индикаторы и многое другое.

Но здесь есть несколько важных моментов:

- 1. У нас датасет сгенерирован по стандартным параметрам индикаторов. И как вы понимаете, в вашем случае может быть такое, что требуется обогащение или дополнительная настройка индикаторов. Для этого могут быть взяты исходные коды.
- 2. Наш датасет учитывает только технический анализ. Если рассматривать инвестиционные стратегии, то важно расширить датасетм и учесть фундаментальные показатели компаний.

### ГЕНЕРАЦИЯ ДАТАСЕТОВ

Теперь давайте перейдем к общим, но очень важным вещам, которые помогут Вам создать свои собственные эффективные датасеты.

#### Важный момент №1.

Генерируемые вами признаки должны усиливать и развивать логику принятия решений нейронной сетью. Например, если вы понимаете, что в принятии инвестиционных решений трейдеры и инвестиционные фонды руководствуются, в том числе, фундаментальными показателями компаний, то признаки данного вида аналитики должны быть включены в датасеты.

Но при обогащении датасетов нельзя руководствоваться логикой, мол давайте как в «помойку» добавим множество признаков, авось прокатит и сделает лучше. Следует понимать, что с одной стороны признаки могут зашумлять данные и ухудшать качество модели. С другой стороны, если у вас много признаков, то вы вынуждены обучать модель на больших данных.

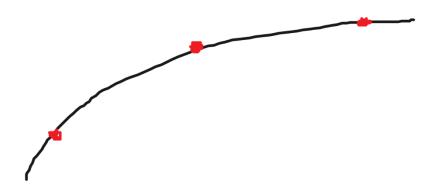


Рисунок 11 – Абстрактная кривая эластичности одного признака от другого

Допустим у вас есть понимание того, что в вашем датасете один признак зависит от другого и вам нужно обучить модель, чтобы она построила кривую хотя бы по трем точкам.

### ГЕНЕРАЦИЯ ДАТАСЕТОВ

В таком случае матрица экспериментов для N признаков бы выглядела так Таблица 1 – Пример абстрактной матрицы планирования экспериментов

		1
Точка 1 признака 3	Точка 1 признака 2	Точка 1 признака 1
		Точка 2 признака 1
		Точка 3 признака 1
	Точка 2 признака 2	Точка 1 признака 1
		Точка 2 признака 1
		Точка 3 признака 1
	Точка 3 признака 2	Точка 1 признака 1
		Точка 2 признака 1
		Точка 3 признака 1
Точка 2 признака 3	Точка 1 признака 2	Точка 1 признака 1
		Точка 2 признака 1
		Точка 3 признака 1
	Точка 2 признака 2	Точка 1 признака 1
		Точка 2 признака 1
		Точка 3 признака 1
	Точка 3 признака 2	Точка 1 признака 1
		Точка 2 признака 1
		Точка 3 признака 1

Взгляните на таблицу 1, которая представляет собой не полную патрицу планирования экспериментов для 3-х признаков. А представьте то у вас их 1000?? 10 000?? Вы можете себе представить сколько вам нужно данных для качественного обучения модели?

#### Важный момент №2.

Отбор признаков.

### ГЕНЕРАЦИЯ ДАТАСЕТОВ

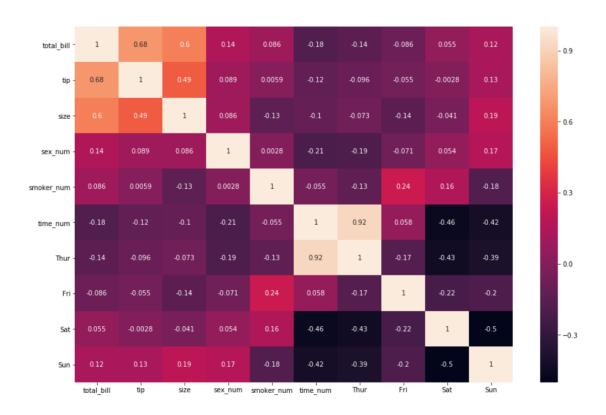


Рисунок 12 – Матрица кореляций признаков (абстрактные признаки)

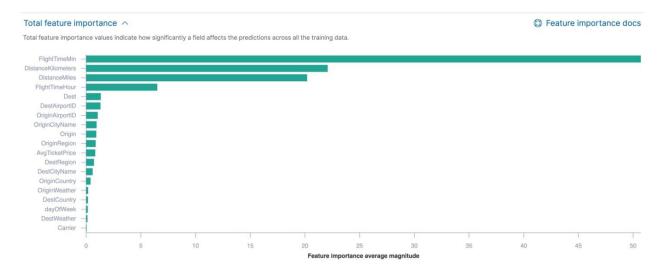


Рисунок 13 – Feature impotance (абстрактные признаки)

### ГЕНЕРАЦИЯ ДАТАСЕТОВ

Для повышения качества датасеты вы должны брать те, которые не имеют высокую степень корреляции между собой (матрица корреляций, рисунок 12) и имеют наибольшее влияние на расчитываемые выходные сигналы (feature impotance, рисунок 13).

#### Важный момент №3

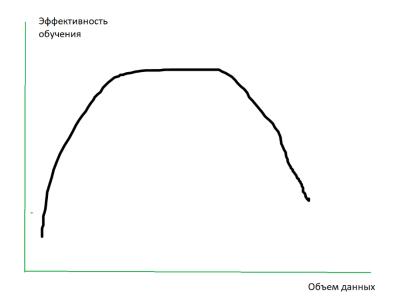


Рисунок 14 – Абстрактная кривая эффективности обучения от объема данных

Нельзя руководствоваться логикой, что чем больше данных тем лучше. Есть некая оптимальная точка на кривой, когда модель обучается лучше всего. И данный параметр нужно подбирать опытным путём.

#### Важный момент №4

Достаточность данных.

Ваших данных должно быть достаточно для получения обобщенной модели.

### ГЕНЕРАЦИЯ ДАТАСЕТОВ

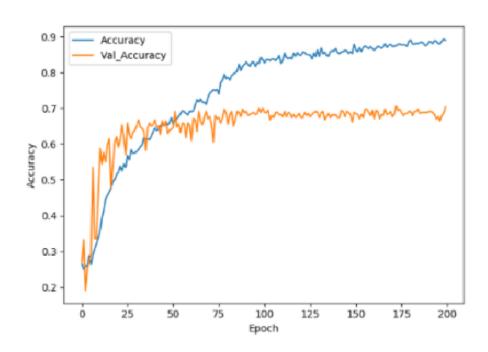


Рисунок 15 – Абстрактные примеры точности модели на тренировочных и тестовых данных

Взгляните на расхождение оранжевой (вариационная выборка) и синей линий (тренировочная выборка). Когда у вас будет мало данных, то ваша модель может адаптироваться к тренировочной выборке, и в итоге она не станет обобщенной. То есть, она плохо будет работать вне условий, где её обучали.

То есть, если вы возьмете финансовый инструмент, который модель не видела, то скорее всего вы получите низкие метрики. Это связано с процессом переобучения, в связи с малым количеством данных.

Объем датасета должен быть достаточен для качественного процесса обучения.

### Обучение нейронных сетей

Вот мы и подошли к самому важному моменту, построению модели, которая и будет вам зарабатывать деньги. Как бы вы хорошо не поработали на предыдущих этапах, если вы качественно не обучите модель, то ничего не получите.



Рисунок 16 – Пример Front-end сервиса обучения нейронных сетей

Следует понимать очень важную вещь. Хотя мы в нашем микросервисе (<a href="https://github.com/powerlife000/moex\_hackaton/blob/main/edu\_neural.py">https://github.com/powerlife000/moex\_hackaton/blob/main/edu\_neural.py</a>) обучения нейронных сетей даём одну архитектуру, нужно понимать, что она кастомизируется для различных стратегий. Могут быть использованы различные слои, нейроные, параметры регулязации и скорости обучения. И всё это может изменяться в зависимости от данных, с которыми вы работаете.

Подбор оптимальной структуры и параметров обучения весьма сложный процесс.

Если вы исследуете различные варианты архитектуры, то вы можете воспользоваться Keras Tuner (<a href="https://keras.io/keras\_tuner/">https://keras.io/keras\_tuner/</a>) и протестировать различные варианты на небольших данных.

Однако следует помнить одну важную вещь! Разработанная вами модель на НЕБОЛЬШИХ ДАННЫХ, может, а скорее всего и будет, работать подругому НА БОЛЬШИХ ДАННЫХ.

### Обучение нейронных сетей

В связи с вышесказанным, процесс разработки архитектуры можно представить следующими этапами:

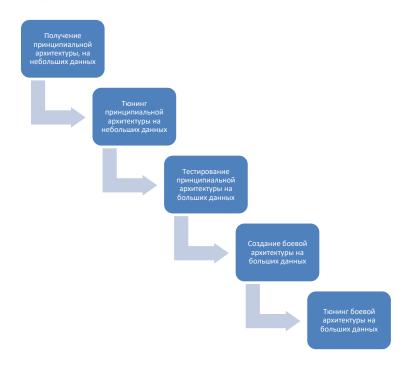


Рисунок 17 – Последовательность разработки нейронных сетей

Сначала вы должны создать то, что работает на маленьких данных. И только после того, как вы оптимизировали модель, которая показывает хорошую эффективность в малом, вы должны переходить к большим данных.

На каждом из этапов разработки вы должны оценивать эффективность работы ваших нейронных сетей.

Оценивать вы их можете технически и бизнесово:

- Технически: всевозможные парамметры точности: f1, roc-auc и пр.
- Бизнесово: profit & loss и др.

#### Бек-тесты

Если вы выполнили предыдущие этапы, то вы можете собой гордится. Вы уже практически на финишной прямой, и вам осталось убедиться в том, что ваша нейронная сеть работает эффективно прежде чем начать торговлю.

Оценка работоспособности нейронной сети — это важнейший этап, без него вы не можете выпускать нейронную сеть в бой. Давайте подробнее остановимся на некоторых очень важных нюнсах процесса тестирования.



Рисунок 18 – Разбитие датасета не тренировочную и тестовую выборки

Когда вы обучаете вашу модель, то она обучается в неких граничных условиях на тренировочных данных. Взгляните на рисунок 18. Ваша нейронная сеть видит только те данные, которые вы её показали. И этими граничными условиями могут быть время, сами финансовые инструменты и другое.

Важно! Вы не должны тестировать вашу нейронную сеть на данных, которые были в тренировочной выборке.

Если вы обучали вашу нейронную сеть с данных с 2010 по 2020 годы, то тестировать вы можете на 2021 и позже.

Если вы обучали вашу нейронную сеть на данных Сбербанка, то тестировать вы её можете на других финансовых инструментах.

#### Бек-тесты

Теперь когда мы разобрались на каких данных мы должны проводить бектесты, давайте подробнее рассмотрим сами тесты. В наших микросервисах (<a href="https://github.com/powerlife000/moex\_hackaton/blob/main/calc\_profit.py">https://github.com/powerlife000/moex\_hackaton/blob/main/calc\_profit.py</a>) мы тестируем работу модели в двух направлениях: техническое и бизнесовое.

#### Техническое направление

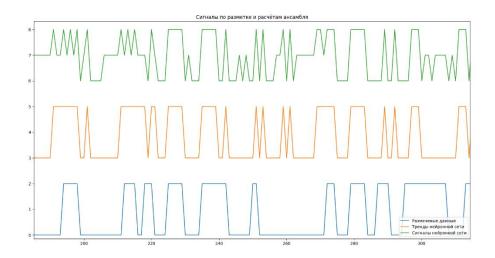


Рисунок 19 – Пример генерации сигналов нейронной сетью

В данном направлении мы технически оцениваем, что из себя представляет генерация сигналов нейронной сетью и на сколько качественно она это делает.

Качество работы технически мы оцениваем по следующим показателям:

Accuracy, roc-auc, precision, recall, f1, logloss.

Понятно, что чем выше показатели точности, тем выше показатели доходности. Однако итоговую эффективность мы оцениваем бизнесово.

### Бек-тесты

#### Бизнесовое направление

При оценки качества работы нейронной сети с точки зрения бизнеса мы оцениваем экономические показатели эффективности её работы и сравниваем с идеально теоретически недостижимыми.

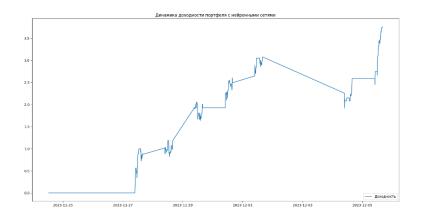


Рисунок 20 — Пример динамики портфеля на бек-тесте при торговле нейронными сетями

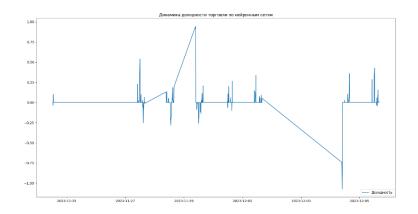


Рисунок 21 – Пример трейдов нейронной сети

### Бек-тесты

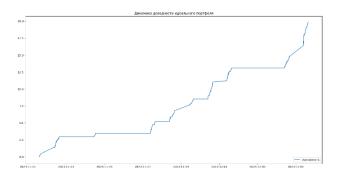


Рисунок 22 – Пример идеального портфеля

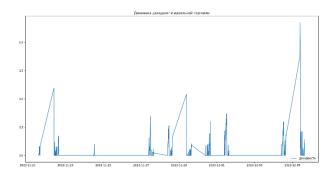


Рисунок 23 – Пример идеальных трейдов

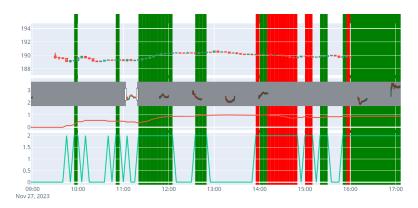


Рисунок 24 — Пример трейдов нейронной сети: **зеленое** — закрыто с прибылью, **красное** — закрыто с убытком

#### Бек-тесты

В ходе бек-теста работы нейронной сети анализируются следующие показатели: волатильность, максимальная просадка, доходность, коэффициент Шарпа, число сделок. Вы можете использовать иные показатели оценки эффективности стратегии.

Мы оцениваем эффективность работы нейронной сети по двум направлениям:

- В сравнении со стратегией buy&hold
- В сравнении с идеальной торговлей

В бою важно оценивать вашу стратегию по сравнению с бейслайном. То есть, у вас УЖЕ могут быть какие-то стратегии которые могут работать на различной логике, и вам нужно оценить сравнительную эффективность вашей новой стратегии по сравнению со старой и в зависимости от этого принимать решения.

Прошу обратить внимание на важный нюанс. Если ваша нейронная сеть в какой-то промеждуток времени показала большую или меньшую доходность по сравнению с бейслайном, то это может быть как случайностью, так и закономерностью.

Для качественного принятия решений относительно выбора той или иной нейронной сети или алгоритма вам нужно провести качественные A|B тесты, основанные на аппарате математической статистики. Об этом вы можете прочитать в открытых источниках.

### Формирование портфеля

После того как вы обучили вашу стратегию, вы встаете перед вопросом: «На чём торговать и куда инвестировать?». С точки зрения диверсификации рисков вам выгодно сформировать портфель, но что в него должно войти?

Один из вариантов формирования портфеля это потенциальная доходность того или иного финансового инструмента. Вы можете проскорить каждый финансовый инструмент по разным параметрам, например по прогнозу доходности на какой-то период в будущем, как например это делали мы в нашем сервисе (<a href="https://github.com/sergeycommit/AlgoPack\_time\_series">https://github.com/sergeycommit/AlgoPack\_time\_series</a>).

Скорить финансовые инструменты вы можете по различным показателям:

- средняя доходность (например, по ЕМА);
- максимальная доходность;
- коэффициент прибыли к риску;
- др.



Рисунок 25 – Каков прогноз??

В основе скоринга лежит оценка потенциальной доходности инструмента будущих периодов.

### Формирование портфеля

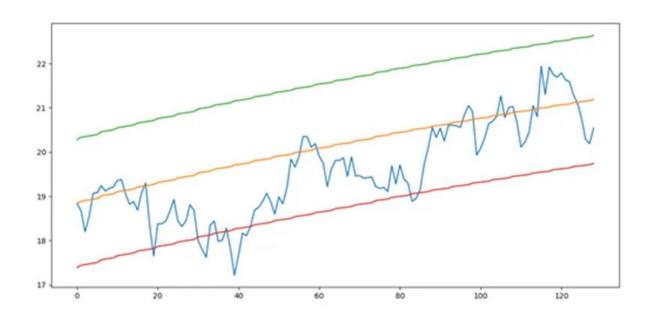


Рисунок 26 – Пример прогнозирования канала будущих периодов

Суть работы такой системы заключается в том, чтобы спрогнозировать движение канала котировок какого-то финансового инструмента в будущем с достаточно высокой вероятностью.

Зная вектор движения канала мы можем определить множество различных параметров, относительно которых можем проскорится.

### ТОРГОВЛЯ

После того как вы выполнили все этапы по разработке стратегии, вы можете приступать к её тестированию в бою и непосредственной торговле на бирже. Мы настоятельно рекомендуем сначала протестировать работу вашей стратеги на демо счетах.

В рамках данного хакатона мы используем MetaTrader5 и написанную для него торговую платформу (<a href="https://github.com/powerlife000/moex\_trading\_system">https://github.com/powerlife000/moex\_trading\_system</a>), которая использует сигналы нейронной сети.

```
C:\Windows\System32\cmd.exe - python MetaTrader5_robot.py - \( \text{ \t
```

Рисунок 27 – Пример работы торговой платформы



Рисунок 28 – Пример MetaTrader 5

### ТОРГОВЛЯ

Здесь я остановлюсь на одном важном моменте. Генерация сигналов для торговли и их использование это не одно и тоже. Нейронная сеть не говорит вам по какой цене покупать и продавать. Она не говорить как входить и выходить из позиции. Она говорит о том, когда, по её мнению, следует входить и выходить.

Пример даже на это вопросе, когда есть очень важный нюанс. Вы не можете принимать решение на текущем баре. Сигнал, который вы обрабатываете это всегда последний закрытый бар.

Разработка бизнесовой логики входа и выхода из позиции уже на вас. Здесь отдельный огромный пласт R&D и оптимизации, в том числе с учётом объемов.

А мы на это заканчиваем наш гайд-бук и желаем вам успехов в торговле и инвестициях.

С любовью и уважением,

ваши NullPointerExeption!