

МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ

ФИЗТЕХ-ШКОЛА ПРИКЛАДНОЙ МАТЕМАТИКИ И  
ИНФОРМАТИКИ

КАФЕДРА ДИСКРЕТНОЙ МАТЕМАТИКИ

---

# Применение моделей глубокого обучения для построения торговой стратегии

---

*Автор:*  
Зелезецкий Даниил  
Владимирович

*Научный руководитель:*  
Куликов Александр  
Владимирович

3 апреля 2024 г.

## 1 Аннотация

В рамках данной работы производится проверка эффективности применения моделей глубокого обучения для прогнозирования финансовых временных рядов с дальнейшим поиском оптимального портфеля посредством решения задачи оптимизации. В качестве данных в работе используются цены акций, входящих в состав Индекса Московской Биржи на момент проведения эксперимента, а также 57 технических индикаторов, которые применяются для признакового описания временных рядов цен данных активов. Все эксперименты производятся с использованием языка программирования Python и его библиотек Pandas, Numpy, PyTorch, CVXPY.

## 2 Введение

Анализ временных рядов представляет собой развитую теорию, имеющую длинную историю, начиная с появления первых статистических подходов к анализу временных данных и заканчивая современными методиками, такими как применение глубокого обучения. В теории временных рядов существует множество задач и приложений, а их прогнозирование играет особую роль.

Среди всех возможных задач из данной области, особое внимание привлекают задачи прогнозирования цен финансовых активов по причине некоторых особенностей, среди которых можно выделить две основные:

1. Временной ряд цен актива, как правило, не имеет чётких и явно выраженных характеристик, таких как тренд, сезонная и циклическая компоненты, а также постоянная дисперсия и среднее значение. Отсутствие подобных черт делает ряд более непредсказуемым и сложно прогнозируемым[1].

2. Если рассматривать цену актива с точки зрения финансов как науки, то на её значение напрямую влияют объёмы спроса и предложения, а также механизмы работы биржи. Если принципы работы механизмов доподлинно известны, то факторы принятия решений игроков о том, покупать или продавать актив скрыты от глаз исследователя, а это значит, что не существует явно выраженных и общепринятых признаков, объясняющих поведение игроков и которые можно использовать для описания

ряда в задаче его прогнозирования.

Так или иначе, приведённые выше особенности не подавляют интерес исследователей по всему миру к данной задаче. В конце XIX века и начале XX века математики и статистики начали активно исследовать методы анализа временных рядов для прогнозирования экономических и финансовых показателей. С развитием компьютерных технологий во второй половине XX века стало возможным проводить более сложные вычисления и анализировать большие объёмы данных. Это способствовало развитию более точных методов анализа временных рядов, таких как авторегрессионные модели, скользящие средние и спектральный анализ. Пару десятков лет назад, задача анализа временных рядов приобрела новое измерение с появлением классических методов машинного обучения, однако наибольший прирост в точности прогноза появился с началом развития моделей глубокого обучения, поскольку именно они продемонстрировали, что способны находить сложные паттерны во входящих последовательностях, которые были недоступны лаконичным, но от того и ограниченным классическим моделям. Различные методы глубокого обучения, такие как алгоритмы полносвязных, рекуррентных и сверточных нейронных сетей стали широко применяться для анализа и прогнозирования временных данных в различных областях, включая финансы, экономику, медицину, климатологию и другие. Следует также отметить, что популярность нейросетевых архитектур начала набирать обороты по мере развития вычислительных мощностей компьютеров, а именно их графических ускорителей. Современные методы анализа временных рядов позволяют обрабатывать большие объёмы данных, выявлять скрытые закономерности и паттерны, предсказывать будущие тенденции и принимать решения на их основе.

Глубокое обучение стало значительным прорывом в области прогнозирования временных рядов благодаря своей способности автоматически извлекать более сложные зависимости из данных. Начиная с простых рекуррентных и сверточных нейронных сетей, исследователи постоянно совершенствуют модели для более точного прогнозирования. С появлением LSTM(Long Short-Term Memory) [2] и GRU(Gated Recurrent Unit) [3] архитектур, глубокое обучение стало способным эффективно учитывать долгосрочные зависимости во временных рядах. Эти типы архитек-

тур позволили значительно улучшить точность прогнозов и зарекомендовали себя на многие годы.

Благодаря схожести постановок задач прогнозирования временных рядов и языкового моделирования, практически все архитектуры из одной области могут быть применены в другой с незначительными изменениями. Машинная обработка естественного языка является активно развивающейся областью, которая получила значительный прирост внимания благодаря статье[4] и предложенной в ней архитектурой Transformer с блоками внимания, именуемыми авторами как Multi-Head Attention Layer и способными особенно качественно захватывать паттерны в данных, о чём свидетельствуют приложенные в статье результаты. Функция Multi-Head Attention блока имеет следующий вид:

$$MultiHead(Q, K, V) = Concat(head_1, ..., head_h)W^O \quad (1)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (2)$$

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (3)$$

Данная архитектура также может быть применена в задаче прогнозирования временных рядов.

### 3 Методика выполнения эксперимента:

Поскольку целью данной работы является исследование эффективности ряда современных нейросетевых архитектур для прогнозирования цен акций российских компаний, а также дальнейшее применение полученных прогнозов для формирования торговой стратегии, эксперимент состоит из двух этапов:

#### 3.1 Этап 1:

Первым и основным этапом данной работы является исследование и сравнение современных моделей глубокого обучения при решении задачи прогнозирования цен акций. В центре внимания данной работы находятся такие архитектуры, как Transformer[4] и TSMixer предложенная в работе [5]. В качестве уже зарекомендовавшего себя подхода используется архитектура LSTM[6],

с эффективностью которой будут сравниваться модели, упомянутые выше. В качестве данных используются почасовые цены акций компаний с 03.01.2020 по 03.01.2024, входящих в индекс Московской Биржи на момент проведения исследования. В качестве признакового описания временных рядов будут использоваться 57 технических индикаторов, рассчитанных на основе потока сделок, заявок и стакана котировок, полученных с помощью официальной библиотеки `MoexAlgo`, написанной на языке программирования Python. Для оценки мер качества моделей в работе используются метрики Mean Squared Error:

$$MSE(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (4)$$

А также Winrate - доля правильно угаданных направлений движения актива:

$$Winrate(y, \hat{y}) = \frac{1}{N} \sum_{i=2}^N [sign(y_i - y_{i-1}) = sign(\hat{y}_i - \hat{y}_{i-1})] \quad (5)$$

В ходе исследования, описанные выше модели пройдут серию из шести испытаний. Во время обучения моделей, роль функции потерь будет выполнять среднеквадратичная функция потерь Mean Squared Error. Испытания приведены ниже:

Таблица 1: Таблица испытаний

	И.1	И.2	И.3	И.4	И.5	И.6
LSTM	•	•	•	•	•	•
Transformer(encoder часть)	•	•	•	•	•	•
Transformer(encoder + decoder)	•		•		•	
TSMixer			•	•	•	•

Где:

И.1: Одномерный ряд(цена закрытия), прогноз на 1 час вперёд.

И.2: Одномерный ряд(цена закрытия), прогноз на 3 часа вперёд.

И.3: Многомерный ряд(признаки - технические индикаторы), прогноз на 1 час вперёд.

И.4: Многомерный ряд(признаки - технические индикаторы), прогноз на 3 часа вперёд.

И.5: Многомерный ряд(признаки - другие активы), прогноз на 1 час вперёд.

И.6: Многомерный ряд(признаки - другие активы), прогноз на 3 часа вперёд.

Как видно в таблице 1, существуют архитектуры, не проходящие некоторые испытания. Данная постановка эксперимента связана исключительно с особенностями этих моделей.

По итогам первого этапа работы будут получены результаты качества всех моделей с последующим заполнением данной таблицы и нахождением наиболее точной среди них.

### 3.2 Этап 2:

Опираясь на полученные результаты, второй этап представляет собой построение торговой стратегии на основе лучшей модели и сравнение её эффективности с бенчмарком в виде портфеля, состоящего исключительно из бумаги Индекса МосБиржи(ИМОЕХ). Иными словами, в рамках данного этапа мы отвечаем на вопрос: "Хватает ли точности прогнозов для того, что бы эффективно использовать их при составлении портфеля и получать большую доходность, чем при портфеле, состоящем исключительно из одного лишь индекса?".

Для нахождения оптимального портфеля в момент времени  $t$ , будем использовать следующую постановку задачи оптимизации, предложенную в работе [7]:

$$\begin{aligned} \max_{u_t} \quad & R_t^p(u) = \hat{r}_t^T h_t + \hat{r}_t^T u_t - c^T u_t - \gamma \cdot ((h_t + u_t)\Sigma(h_t + u_t)) \\ \text{subject to} \quad & u_t \in \mathbb{R}^{n+1}, h_t + u_t \geq 0 \\ & 1^T u_t + c^T u_t = 0 \end{aligned} \tag{6}$$

где :

$u_t$  - вектор торговых операций в рублях(целевая переменная)

$\hat{r}_t$  - вектор ожидаемых доходностей

$h_t$  - портфель на момент начала периода  $t$

$c^T u_t$  - транзакционные издержки

$\gamma$  - коэффициент принятия риска  
 $(h_t + u_t)\Sigma(h_t + u_t)$  - функция оценки риска портфеля  
 $1^T u_t + c^T u_t = 0$  - условие самофинансирования

В ходе работы алгоритма, каждый час производится перебалансировка портфеля путём решения данной задачи оптимизации. Более подробная схема алгоритма имеет вид:

---

**Algorithm 1** Торговый алгоритм

---

**Require:**

$h_0 = (0, 0, \dots, \text{initial balance}), c, \gamma, \Sigma, T$  ;

**Ensure:**

$h_T$  - итоговый портфель после торговой сессии

- 1: **for**  $t = 0 : T$  **do**
  - 2:   Решаем задачу оптимизации  $u_t^* = \arg \max_u R_t^p(u)$
  - 3:   Округляем  $u_t^*$  по ценам на момент  $t$
  - 4:   Совершаем торговые операции  $h_t^* = h_t + u_t^*$
  - 5:   Дожидаемся конца периода и фиксируем реализованную доходность  
 $h_{t+1} = (1 + r_t) \circ h_t^*$
  - 6:   Получаем прогнозы доходностей на конец следующего периода  $r_{t+1}^{\hat{}}$
  - 7: **end for**
- 

## 4 Результаты эксперимента:

### 4.1 Демонстрация результатов этапа 1:

Таблица 2: Таблица со значениями MSE метрики на тестовом множестве(январь 2023 - декабрь 2023)

	И.1	И.2	И.3	И.4	И.5	И.6
LSTM	4.86	4.61	3.21	3.92	2.53	5.11
Transformer(encoder часть)	3.81	3.22	4.51	3.02	3.87	5.57
Transformer(encoder + decoder)	4.13		4.52		3.83	
TSMixer			2.52	2.58	2.11	3.96

Таблица 3: Таблица с усреднёнными значениями Winrate метрики на тестовом множестве(январь 2023 - декабрь 2023)

	И.1	И.2	И.3	И.4	И.5	И.6
LSTM	0.48	0.50	0.49	0.50	0.49	0.48
Transformer(encoder часть)	0.48	0.48	0.48	0.50	0.49	0.50
Transformer(encoder + decoder)	0.47		0.49		0.51	
TSMixer			0.51	0.50	0.49	0.49

Для более понятного сравнения моделей по метрике Winrate, их также следует сравнивать с наивными прогнозами о том, что актив всегда растёт, значение следующего шага всегда выше предыдущего. Это даст понимание того, действительно ли хорошее значение Winrate связано с техническими преимуществами модели или же это не более чем простое угадывание. Ниже приведена таблица с результатом наивных прогнозов:

Таблица 4: Winrate для наивных прогнозов(январь 2023 - декабрь 2023)

	Winrate
Прогноз постоянного роста	0.53

## 4.2 Обсуждение результатов этапа 1:

Изучая таблицу 2, становится понятно, что дополнительные признаки не принесли ощутимого результата. Более того, в некоторых случаях дополнительные данные только сбивали модель, мешая ей находить паттерны. И если теоретически предположить, что дополнительная информация не имеет смысловой нагрузки, то модель должна была сформировать свои веса таким образом, что эта информация бы игнорировалась, однако на практике достичь этого не удалось. Иными словами: правило "Чем больше информации, тем выше точность" не сработало. Как можно наблюдать из таблицы 3, ни одна из моделей не смогла качественно уловить все скрытые тенденции данных и сделать хороший прогноз. Особенно это становится понятно, когда мы сравниваем полученный результат с таблицей наивных прогнозов, которые дают Winrate равный 0.48 и 0.53 соот-



ветственно. Поскольку период тестовых данных (январь 2023 - декабрь 2023) пришелся на рост рынка, то прогноз постоянного роста даёт точность выше, чем 0.50. Одним из объяснений таких низких результатов качества может служить среднеквадратичная функция потерь, которая по своей сути штрафует модель не за то, угадала ли она направление движения, а за то, насколько близко прогноз и фактическое значение лежат друг с другом. Более подробное обсуждение этой проблемы приводится далее.

#### 4.3 Демонстрация результатов этапа 2:

Как можно видеть из постановки задачи оптимизации, используемый алгоритм является упрощённым и не позволяет занимать короткие позиции, не учитывает дивидендные сезоны и имеет простую функцию оценки риска. Все эти ограничения при желании могут быть сняты в угоду получения более сложного и эффективного алгоритма. Цель данного алгоритма - симитировать принципы принятия решений рядового игрока, который не прибегает к более сложным техникам, таким как занятие коротких позиций или выставление двойных ордеров. Единственным источником принятия решений для него служат прогнозы моделей, а единственное принимаемое решение - это покупать актив на час или нет.

Для проверки алгоритма на эффективность была проведена серия экспериментов, в которой в качестве гиперпараметров выступали: транзакционные издержки, параметр принятия риска, порог уверенности в прогнозе на конкретный актив.

Подробнее о пороге уверенности в прогнозе: как следует из описания этапа 1, для каждого актива была обучена своя модель, а это значит, что у нас есть статистика о том, какие активы прогнозируются лучше, а какие хуже. Следовательно, перед началом работы алгоритма мы можем провести фильтрацию и игнорировать прогнозы тех активов, winrate которых ниже заданного порога.

Результаты серии экспериментов приведены ниже. В качестве модели была взята архитектура TSMixer, которая принимает на вход временной ряд с признаковым описанием из 57 технических индикаторов и делает прогноз на час вперёд. Оценкой качества алгоритма является итоговая стоимость портфеля после окон-

чания его работы. Начальная сумма равна 100.000 рублей.

Таблица 5: Результаты экспериментов для транз. издержек = 0.0004 и стартового капитала = 100.000 рублей

	2	1	1/2	1/30	1/90
Threshold = 0.515	99.617 руб.	99.230 руб.	98.558 руб.	78.694 руб.	60.914 руб.
Threshold = 0.50	99.587 руб.	99.183 руб.	98.402 руб.	76.913 руб.	56.440 руб.
Threshold = 0.48	99.568 руб.	99.129 руб.	98.265 руб.	74.315 руб.	53.187 руб.

Таблица 6: Результаты экспериментов для транз. издержек = 0.001 и стартового капитала = 100.000 рублей

	2	1	1/2	1/30	1/90
Threshold = 0.515	99.078 руб.	98.143 руб.	96.402 руб.	46.327 руб.	22.927 руб.
Threshold = 0.50	99.049 руб.	98.101 руб.	96.250 руб.	45.030 руб.	21.353 руб.
Threshold = 0.48	99.032 руб.	98.057 руб.	96.126 руб.	42.996 руб.	20.705 руб.

Таблица 7: Стоимость единицы акции IMOEX на начало и конец эксперимента

03.01.2023	29.12.2023	Доходность
2.157	3.092	43.35%

Согласно результатам из таблиц 5 и 6, ни одна комбинация гиперпараметров не смогла показать качественный результат, который давал хотя бы положительную доходность. Несмотря на это, полученные значения итоговой стоимости портфеля полностью совпадают с общими представлениями. Чем более высокий порог мы задаём, тем меньшим количеством активов мы торгуем, хотя прогнозы по ним более качественные и как следствие - меньшие убытки. Кроме того, более агрессивная торговля в данной ситуации приводит к большим потерям, что также совпадает с общими представлениями о принципах торговли.

#### 4.4 Обсуждение результатов:

По окончании серии экспериментов и завершению анализа полученных результатов, хочется сделать следующий вывод:

Отчётливо видно, что несмотря на высокую эффективность современных моделей глубокого обучения в задачах предсказания временных рядов, на финансовых данных они по-прежнему не могут показать свою точность. Более того, используемое признаковое описание рядов хоть и позволяет добиться прироста в точности, однако не помогает моделям улучшить прогнозы настолько, что бы ими можно было эффективно пользоваться. Также одним из важных факторов является функция потерь, по которой обучаются модели. В большинстве работ по прогнозированию временных рядов, в том числе и финансовых, используется среднеквадратичная функция потерь Mean Squared Error, описанная в формуле 4. По мнению автора, её использование может быть оправдано на разных классах задач, где важно добиться приближения к истинному значению, однако при прогнозировании цен активов появляется другое, намного более важное требование: важно не сколько само приближение прогноза к истинному значению, столько угадывание направления его движения, ведь именно это является ключевой гарантией успешной транзакции, однако MSE функция не способна оказывать прямое воздействие на эту способность модели в процессе обучения.

В качестве идей по дальнейшей модернизации испытаний можно отметить следующие:

- **Переход к задаче классификации с дальнейшим изменением алгоритма торговли:** в рамках данной модификации предлагается совершить переход от задачи регрессии к задаче классификации и оптимизировать модели используя функцию бинарной кросс-энтропии. Данный подход игнорирует необходимость приближения прогноза к истинным значениям, однако учитывает необходимость корректного угадывания движения цены. Переход к данной постановке также требует изменений в структуре торгового алгоритма.
- **Переход к прогнозированию волатильности на несколько периодов вперёд:** в рамках данной модификации предлагается моделировать волатильность и при высоком ожи-

даемом значении выставят два ордера на покупку и продажу в один момент. Иными словами: если ожидается сильное колебание цены вниз и вверх, то покупаем внизу и продаём в тот момент, когда цена нормализовалась. Данный подход также требует пересмотра принципов работы торгового алгоритма.

- **Выставление порогов:** основной идеей данного метода является покупка только тех активов, ожидаемая доходность которых выше некоторого порога. Данный подход имеет шанс увеличить эффективность торговли, так компенсирует некоторую погрешность прогнозов.
- **Использование другой функции потерь:** как говорилось ранее, MSE приближает прогноз к истинному значению, однако лишь косвенно влияет на угадывание знака доходности. Всё это подталкивает к необходимости использования другой функции. Функция Winrate, описанная выше, не является дифференцируемой, а значит не позволяет проводить обучение градиентными методами.
- **Дополнительное признаковое описание рядов:** как упоминалось ранее, одних лишь технических индикаторов оказалось недостаточно для эффективной работы моделей, а значит следует рассмотреть перспективы использования дополнительной информации в качестве признакового описания.

## Список литературы

- [1] Anna Zinenko. Forecasting financial time series using singular spectrum analysis. *Business Informatics*, 2023.
- [2] Ralf C. Staudemeyer and Eric Rothstein Morris. Understanding lstm – a tutorial into long short-term memory recurrent neural networks, 2019.
- [3] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling, 2014.

- [4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023.
- [5] Si-An Chen, Chun-Liang Li, Nate Yoder, Sercan O. Arik, and Tomas Pfister. Tsmixer: An all-mlp architecture for time series forecasting, 2023.
- [6] Benjamin Lindemann, Timo Müller, Hannes Vietz, Nasser Jazdi, and Michael Weyrich. A survey on long short-term memory networks for time series prediction. *Procedia CIRP*, 99:650–655, 2021. 14th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 15-17 July 2020.
- [7] Stephen Boyd, Enzo Busseti, Steven Diamond, Ronald N. Kahn, Kwangmoo Koh, Peter Nystrup, and Jan Speth. Multi-period trading via convex optimization, 2017.