

## **Аннотация**

Исследуется проблема повышения качества моделей прогнозирования временных рядов на примере динамики курса акций. Рассматривается метод включения в модель внешних данных. Вводится предположение, что агрегированные знания опытных инвесторов повышают качество тестируемой модели. Рассматриваются нейросетевые методы машинного обучения и проводятся вычислительные эксперименты на реальных данных.

**Ключевые слова:** временные ряды, нейронные сети, краткосрочное прогнозирование

# Содержание

<b>1</b>	<b>Введение</b>	<b>4</b>
1.1	Обзор предметной области . . . . .	4
1.2	Предложенный метод . . . . .	6
<b>2</b>	<b>Постановка задачи прогнозирования</b>	<b>7</b>
<b>3</b>	<b>Экспериментальные данные</b>	<b>8</b>
3.1	Стационарность . . . . .	8
3.2	Составление выборки . . . . .	10
<b>4</b>	<b>Вычислительный эксперимент</b>	<b>11</b>
4.1	ARIMA . . . . .	12
4.2	Модель автоследования . . . . .	14
4.3	LSTM . . . . .	15
4.4	Горизонт прогнозирования . . . . .	18
4.5	Transformer . . . . .	19
4.6	Код вычислительного эксперимента . . . . .	23
<b>5</b>	<b>Заключение</b>	<b>24</b>

# 1 Введение

**Актуальность темы.** Прогнозирование цен на акции с помощью моделей машинного обучения является сложной задачей из-за высокой степени шума и множества факторов, влияющих на поведение цен. В этом случае помимо данных временного ряда можно использовать агрегированные знания опытных инвесторов.

**Цель работы.** Одним из способов повышения качества алгоритма машинного обучения является обогащение выборки - использование дополнительных данных. Цель данной работы заключается в повышении качества нейросетевых моделей прогнозирования временных рядов на примере динамики курса акций. Для этого предлагается использовать знания опытных инвесторов.

**Новизна.** Предложен подход, основанный на предположении о том, что внешние факторы, влияющие на курс акций, заложены в ответы опытных инвесторов.

**Определение 1.1.** *Модель автоследования — временной ряд, составленный из агрегированных ответов опытных инвесторов о решении продажи или покупки акций.*

## 1.1 Обзор предметной области

Основным подходом краткосрочного прогнозирования временных рядов является использование моделей семейства ARIMA, описанных в работе [1]. Данные модели основаны на использовании авторегрессии и скользящего среднего. В работах [3, 4] рассматриваются тесты стационарности, являющейся важным условием для работы с моделями из данного семейства. Задача прогнозирования цен на акции с использованием модели ARIMA описана в [2].

Нейросетевые методы прогнозирования используют Кодировщик-Декодировщик архитектуру, описанную в [8]. Кодировщик преобразует входную последовательность в векторное представление, ис-

пользуемое Декодировщиком при построении выходной последовательности, однако в качестве Кодировщика и Декодировщика также могут выступать различные архитектуры. В работе [8] в качестве Кодировщика и Декодировщика используется архитектура рекуррентной нейронной сети LSTM [5, 6].

В работе [9] предлагается использовать помимо архитектуры Кодировщика и Декодировщика также механизм внимания - меру сходства состояний модели. Различные виды данного механизма описаны в [10]. А в работе [11] описан механизм самовнимания - метод создания векторных представлений последовательностей.

Современные методы прогнозирования основаны на применении моделей трансформеров, описанных в [12]. Данный подход основан на использовании механизма внимания без применения рекуррентности. Еще одним важным аспектом данных моделей является позиционное кодирование [13] - включение дополнительной информации о порядке передаваемой последовательности.

Последние исследования в области прогнозирования временных рядов используют архитектуры эффективных трансформеров - модификаций базовой архитектуры трансформеров, направленных на повышение их производительности, снижение вычислительных затрат и улучшение обработки последовательностей. В работе [18] описана архитектура Informer, использующая механизм разреженного внимания - уменьшение размерности вектора запроса. Архитектура Autoformer [17] использует слой декомпозиции - разделение временных рядов на сезонные и трендовые компоненты и автокорреляцию в качестве механизма внимания.

Задача прогнозирования временных рядов находит свое применение во многих областях. Так, в [21] в качестве экспериментальных данных используются такие датасеты, как курс валют, метеорологические данные, данные о потреблении электричества и другие данные реального мира.

## 1.2 Предложенный метод

Предлагается в тестируемой модели использовать помимо данных временного ряда также ответы модели автоследования. Ожидается, что качество полученных моделей будет превышать качество моделей без использования ответов модели автоследования.

В качестве экспериментальных данных используются реальные данные подневной динамики курса акций YNDX и ответы опытных инвесторов - авторов стратегий автоследования Тинькофф Инвестиций.

## 2 Постановка задачи прогнозирования

$y_1, y_2, \dots, y_T$  - временной ряд,  $y_i \in \mathbb{R}^n$ .

Требуется получить модель временного ряда:

$$\hat{y}_{t+k}(\mathbf{w}) = f_{t,k}(y_{t-M+1}, \dots, y_t; \mathbf{w})$$

$$k = 1, \dots, K,$$

где

$M$  - размер входного окна,

$K$  - горизонт прогнозирования,

$\mathbf{w}$  - вектор параметров модели.

Функция потерь  $\mathcal{L}$ , используемая при обучении модели:

$$\mathcal{L}(\mathbf{w}, \mathbf{Y}) = \sum_{t=M}^{T-K} \sum_{k=1}^K (f_{t,k}(y_{t-M+1}, \dots, y_t; \mathbf{w}) - y_{t+k})^2,$$

Получаем оптимизационную задачу:

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w} \in \mathbb{W}} \mathcal{L}(\mathbf{w}, \mathbf{Y}).$$

## 3 Экспериментальные данные

Эксперимент проводится для данных курса акций YNDX.

Задается временной ряд

$$x_1, x_2, x_3, \dots, x_N, \quad x_i \in \mathbb{R}^5$$

$$x_i = [c_i \ o_i \ h_i \ l_i \ a_i]^T,$$

где

$c_i$  - цена закрытия,

$o_i$  - цена открытия,

$h_i$  - максимальная цена,

$l_i$  - минимальная цена,

$a_i$  - ответ модели автоследования ( $a_i = 0$  в базовом варианте обучения модели)

### 3.1 Стационарность

Исходный ряд приводится к стационарному виду следующими преобразованиями:

- 1) Дифференцирование:  $y'_t = y_t - y_{t-1}$
- 2) Сезонное дифференцирование  $y''_t = y'_t - y'_{t-s}$ ,  $s = 5$

Т.к. торги открыты только в будние дни, то выбранная длина сезона  $s = 5$ .

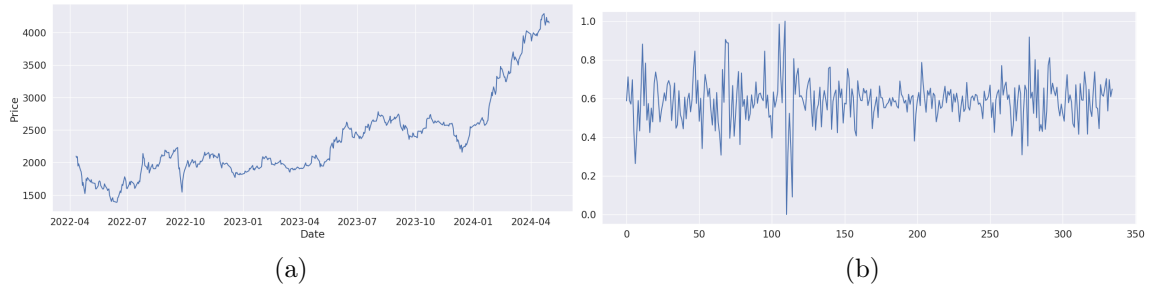


Рис. 1: Временной ряд курса акций YNDX до и после преобразований

Для проверки ряда на стационарность используется критерий KPSS [3, 4]:

Для полученного ряда  $p - value > 0.01$

Для исходного ряда  $p - value < 0.01$

Для визуальной оценки стационарности ряда также построим Q-Q plot:

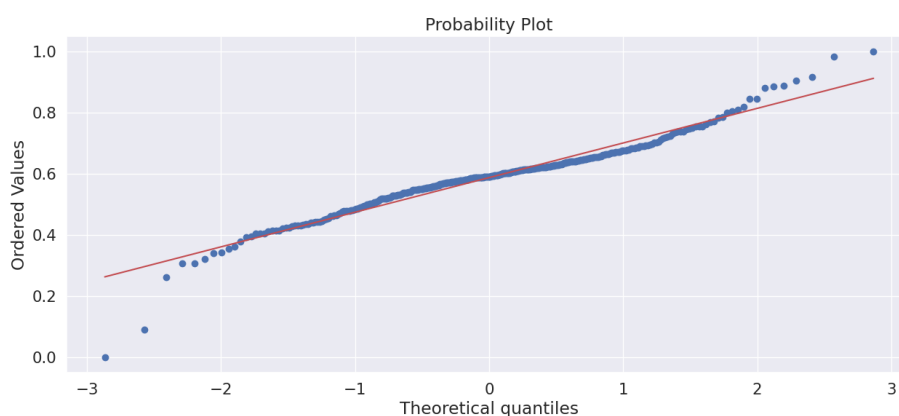


Рис. 2: Q-Q plot полученного ряда

Видно, что точки на графике примерно лежат на прямой линии, следовательно распределение полученного временного ряда близко к нормальному.

Также воспользуемся анализом автокорреляционной функции:

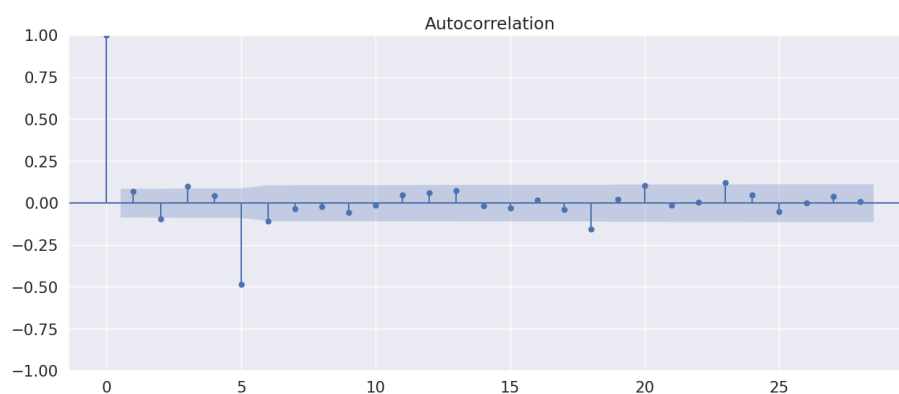


Рис. 3: ACF полученного ряда



Ближкие к нулю значения автокорреляций также указывают на стационарность полученного ряда.

## 3.2 Составление выборки

Методом скользящего окна составляется выборка  $\mathfrak{D} = (\mathbf{X}, \mathbf{Y})$ :

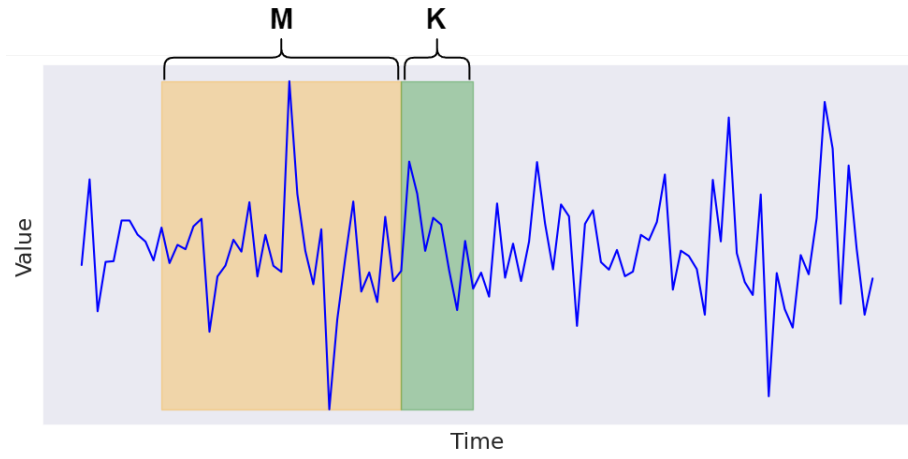


Рис. 4: Метод скользящего окна

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & \dots & x_M \\ x_2 & x_3 & \dots & x_{M+1} \\ x_3 & x_4 & \dots & x_{M+2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N-K-M+1} & x_{N-K-M+2} & \dots & x_{N-K} \end{pmatrix}, \mathbf{Y} = \begin{pmatrix} c_{M+1} & c_{M+2} & \dots & c_{M+K} \\ c_{M+2} & c_{M+3} & \dots & c_{M+K+1} \\ c_{M+3} & c_{M+4} & \dots & c_{M+K+2} \\ \vdots & \vdots & \vdots & \vdots \\ c_{N-K+1} & c_{N-K+2} & \dots & c_N \end{pmatrix},$$

где  $M$  - размер окна,  $K$  - горизонт прогнозирования.

В качестве целевых значений используется цена закрытия.

В соотношении 80/20 выборка делится на обучающую и тестовую части: YNDX-train, YNDX-test.

## 4 Вычислительный эксперимент

Для анализа предложенного метода проводится вычислительный эксперимент для задачи прогнозирования временного ряда на  $K = 5$  шагов.

В качестве тестируемых моделей используются архитектуры Кодировщика - Декодировщика [8] на основе рекуррентной сети LSTM [5, 6] и модели трансформера [12].

Эксперимент проводится для выборок с размером входного окна, равного 5, 15 и 30. Каждая из выборок состоит из обучающей и тестовой части.

Таблица 1: Выборки

Выборка	Пояснение	Размер входного окна	Горизонт прогнозирования
YNDX-Train5	Обучающая часть	5	5
YNDX-Test5	Тестовая часть	5	5
YNDX-Train15	Обучающая часть	15	5
YNDX-Test15	Тестовая часть	15	5
YNDX-Train30	Обучающая часть	30	5
YNDX-Test30	Тестовая часть	30	5

Для решения оптимизационной задачи используется градиентный метод оптимизации Adam [14].

Для анализа качества моделей используются метрики:

- 1) Средняя квадратичная ошибка
- 2) Корреляция Пирсона, усредненная по каждому шагу прогнозирования (далее - Корреляция Пирсона)

## 4.1 ARIMA

В качестве базовой модели прогнозирования используется архитектура ARIMA [1] с порядком дифференцирования  $d = 0$ , поскольку передаваемый ряд приведен к стационарному виду.

$$ARIMA(p, d = 0, q) :$$

$$y_t = \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q},$$

где

$y_t$  - стационарный временной ряд,

$\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q$  - константы ( $\alpha_p \neq 0, \beta_q \neq 0$ ),

$\varepsilon_t$  - гауссов белый шум с нулевым средним и постоянной дисперсией.

Модели-кандидаты обучаются на первых 80 % данных. Итоговый порядок модели определяется с помощью информационного критерия Акаике:

$$AIC = -2 \log \mathcal{L} + 2k,$$

где

$\mathcal{L}$  - значение функции правдоподобия модели,

$k = p + q + 1$  - количество параметров модели

Качество полученной модели  $ARIMA(p = 3, d = 0, q = 3)$  оценивается на оставшихся 20 % данных для задачи прогнозирования на  $K$  шагов.

Таблица 2: Качество моделей

Выборка	Модель	Дополнительные данные	Корреляция Пирсона	MSE
	ARIMA		0,340	0,0183

В таблице 2 представлены результаты сравнения полученных моделей.

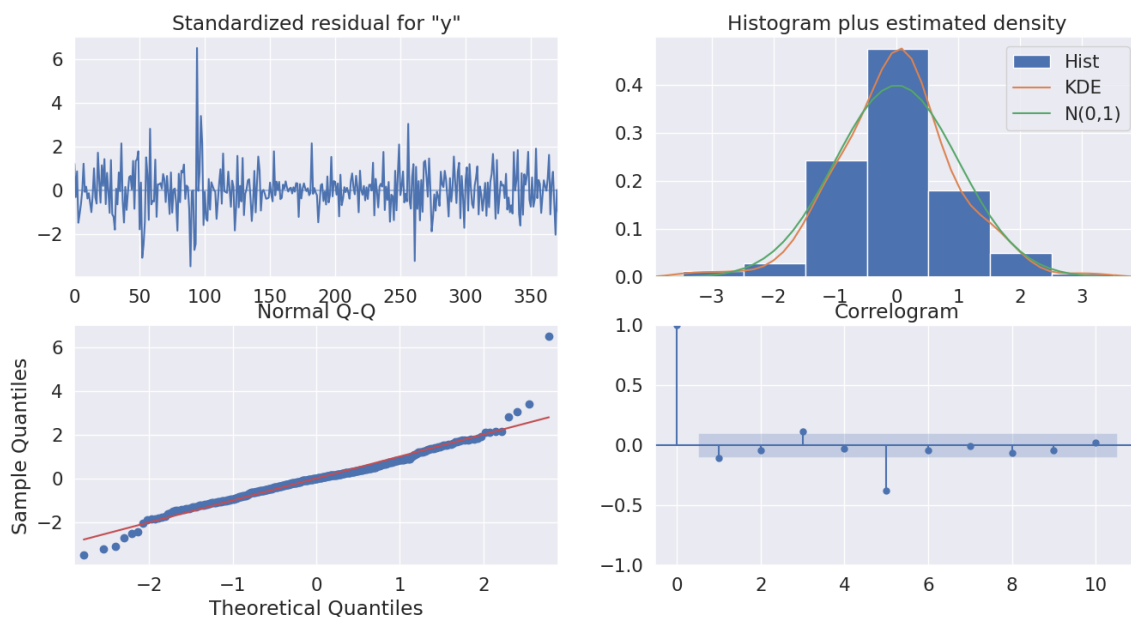


Рис. 5: Остатки ARIMA

Остатки модели - разница между фактическими значениями и значениями, прогнозируемыми моделью.

На рис. 5 показана диагностика полученной модели:

- 1) График остатков позволяет увидеть, равномерно ли распределены остатки вокруг нуля
- 2) Гистограмма остатков помогает визуально оценить, нормально ли распределены остатки
- 3) Q-Q plot сравнивает распределение остатков с нормальным распределением
- 4) График автокорреляционной функции остатков показывает автокорреляцию остатков на разных лагах

Исходя из полученных графиков следует вывод, что остатки модели не содержат дополнительной информации, важной для улучшения качества модели.

## 4.2 Модель автоследования

В качестве предложенного метода повышения качества моделей прогнозирования временных рядов предлагается использовать агрегированные знания опытных инвесторов.

Автоследование — способ инвестирования, при котором все желающие могут подключиться к стратегии более опытного инвестора (он же автор стратегии) и автоматически повторять все его сделки на своем счете.

Для создания модели автоследования выделяются инвесторы - авторы стратегий автоследования Тинькофф инвестиций.

$$\text{Ответ инвестора} = \frac{\text{Сумма сделки}}{\text{Объем портфеля}}$$

При этом не рассматриваются сделки, сумма которых превышает объем портфеля. Путем усреднения ответов инвесторов о продаже или покупке акций составляется временной ряд  $a_1, \dots, a_N, a_i \in [-1, 1]$ .

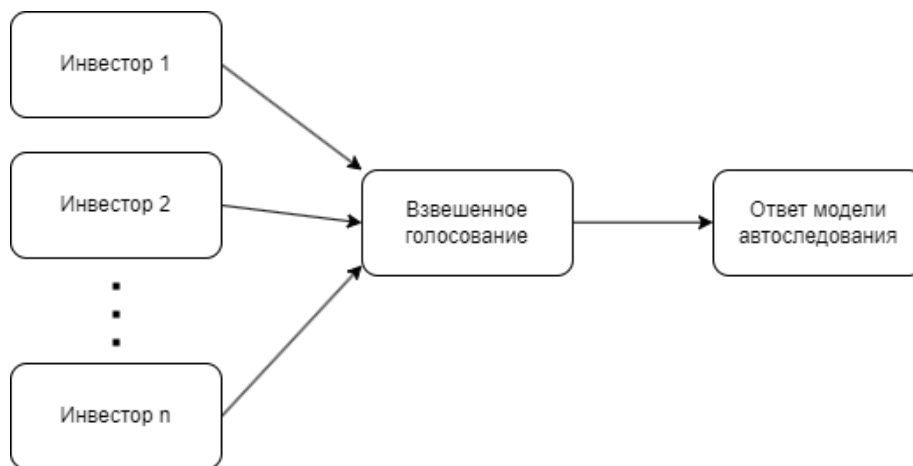


Рис. 6: Модель автоследования

Полученный временной ряд передается в тестируемую модель в качестве дополнительных данных.

### 4.3 LSTM

В качестве тестируемой модели используется архитектура Кодировщика - Декодировщика на основе рекуррентной сети LSTM.

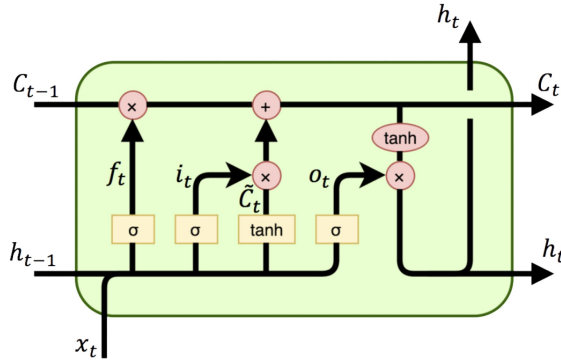


Рис. 7: Ячейка LSTM [7]

$$\begin{aligned} f_t &= \sigma(W_f[h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i) \\ \tilde{C}_t &= \tanh(W_C[h_{t-1}, x_t] + b_C) \\ C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \\ o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) \\ h_t &= o_t \odot \tanh(C_t) \end{aligned}$$

Кодировщик, состоящий из последовательных ячеек LSTM, обрабатывает входную последовательность временного ряда и передает Декодировщику внутреннее представление с последнего шага, которое содержит информацию о всей входной последовательности. Декодировщик, также состоящий из последовательных ячеек LSTM, получает на вход последнее внутреннее представление Кодировщика и использует свои собственные предсказания для генерации последующих значений выходной последовательности.

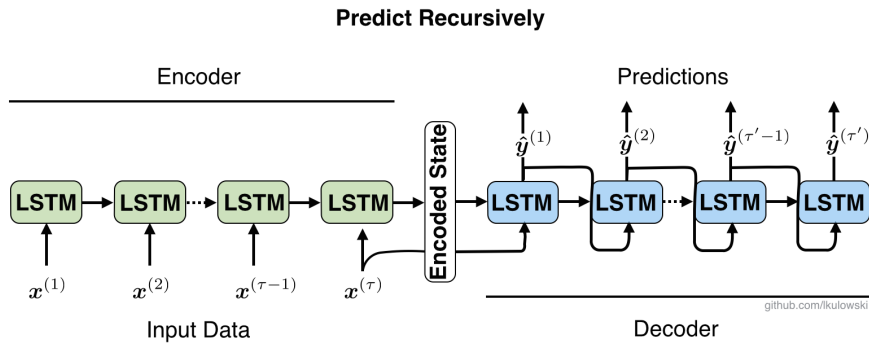


Рис. 8: Кодировщик - Декодировщик архитектура на основе LSTM

Проводится сравнение тестируемой модели с моделями, где в качестве дополнительных данных используются:

- 1) Ответы модели автоследования
- 2) Нормальный шум  $\mathcal{N}(0, 1)$

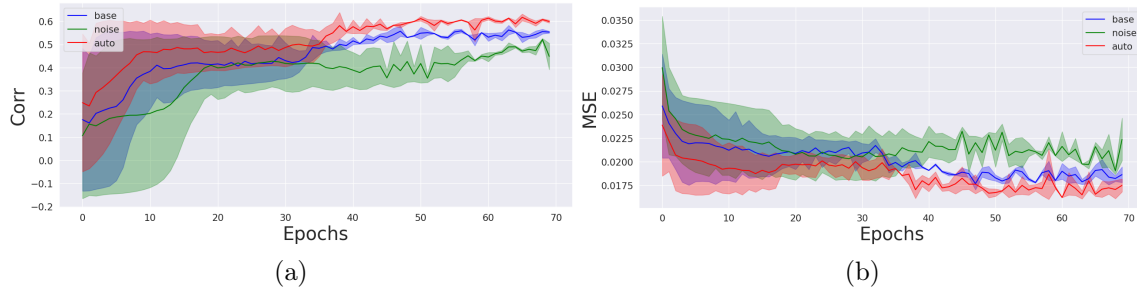


Рис. 9: Качество аппроксимации на тестовой выборке. Все результаты усреднены по 3 запускам. а) Корреляция Пирсона; б) Средняя квадратичная ошибка

На рис. 9 показаны графики зависимости корреляции Пирсона и средней квадратичной ошибки на отложенной тестовой выборке между истинными значениями ряда и ответами модели.

На графиках видно, что модель, использующая ответы модели автоследования, показывает лучшее качество прогнозирования, при этом наблюдается снижение средней квадратичной ошибки.

Таблица 3: Качество моделей

Выборка	Модель	Дополнительные данные	Корреляция Пирсона	MSE
YNDX-Train5	Seq2Seq LSTM	—	$0,476 \pm 0,027$	$0,0175 \pm 0,0003$
		Автоследование	$0,510 \pm 0,036$	$0,0171 \pm 0,0001$
YNDX-Train15	Seq2Seq LSTM	—	$0,553 \pm 0,004$	$0,0186 \pm 0,0008$
		Автоследование	$0,599 \pm 0,006$	$0,0175 \pm 0,0007$
YNDX-Train30	Seq2Seq LSTM	—	$0,478 \pm 0,006$	$0,0195 \pm 0,0005$
		Автоследование	$0,538 \pm 0,007$	$0,0186 \pm 0,0017$

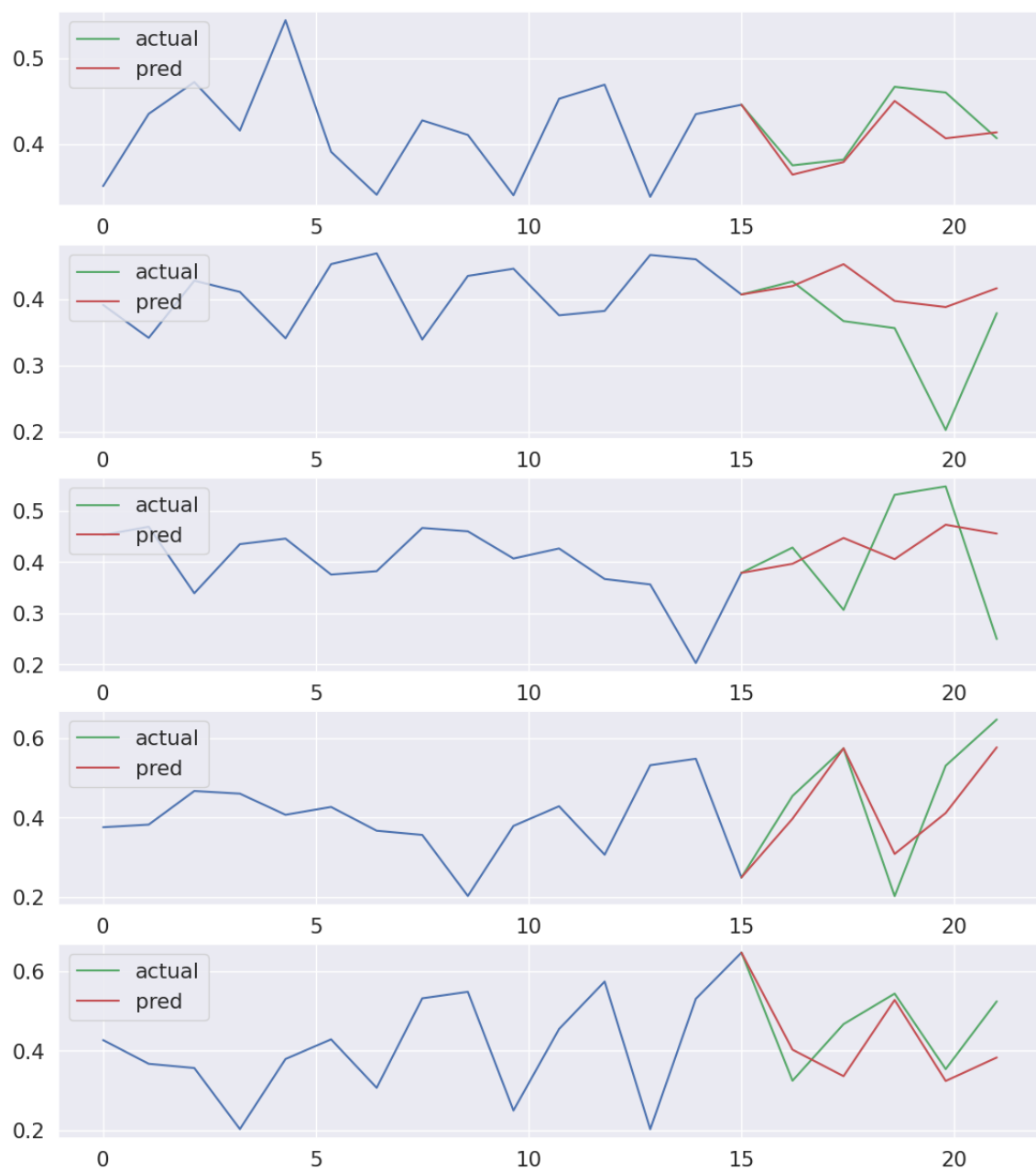


Рис. 10: Прогнозы LSTM

На рис. 10 показаны примеры прогнозов модели Seq2Seq LSTM на отложенной тестовой выборке.



## 4.4 Горизонт прогнозирования

Выбор горизонта прогнозирования является одним из основных аспектов в задаче прогнозирования временных рядов. Поскольку используется рекурсивный метод построения прогнозов, то с каждым шагом накапливается ошибка прогноза. Поэтому правильный выбор горизонта позволяет сбалансировать качество прогноза и его пользу для принятия решений.

Проводится сравнение модели Seq2Seq LSTM в зависимости от выбранного горизонта прогнозирования.

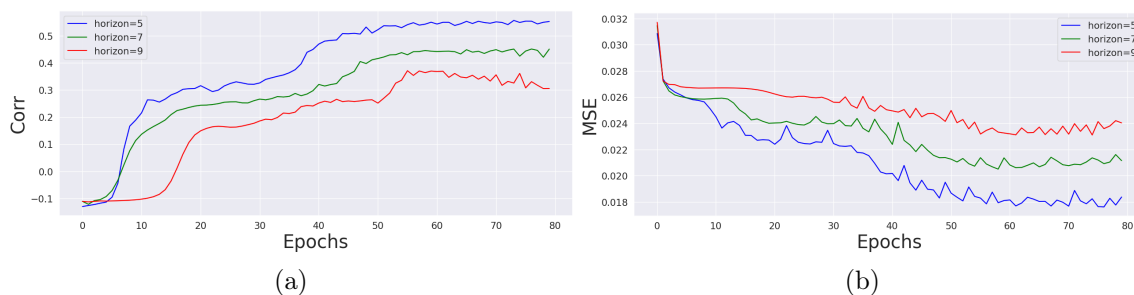


Рис. 11: Качество аппроксимации на тестовой выборке. а) Корреляция Пирсона; б) Средняя квадратичная ошибка

На рис. 11 показаны графики зависимости корреляции Пирсона и средней квадратичной ошибки на отложенной тестовой выборке между истинными значениями ряда и ответами модели.

На графиках видно, что с увеличением горизонта прогнозирования качество модели ухудшается. Также наблюдается более поздний выход корреляции Пирсона на плато.

## 4.5 Transformer

В качестве тестируемой модели используется архитектура Кодировщика - Декодировщика на основе модели трансформера [12].

1. Добавляются позиционные векторы  $p_i$ :  

$$h_i = x_i + p_i, \quad H = (h_1, \dots, h_n)$$
2. Многомерное самовнимание:  

$$h_i^j = \text{Attn}(W_q^j h_i, W_k^j H, W_v^j H)$$
3. Конкатенация:  

$$h_i' = \text{MH}_j(h_i^j) \equiv [h_i^1, \dots, h_i^J]$$
4. Сквозная связь + нормировка уровня:  

$$h_i'' = \text{LN}(h_i' + h_i; \mu_1, \sigma_1)$$
5. Полносвязная 2х-слойная сеть FFN:  

$$h_i''' = W_2 \text{ReLU}(W_1 h_i'' + b_1) + b_2$$
6. Сквозная связь + нормировка уровня:  

$$z_i = \text{LN}(h_i''' + h_i''; \mu_2, \sigma_2)$$

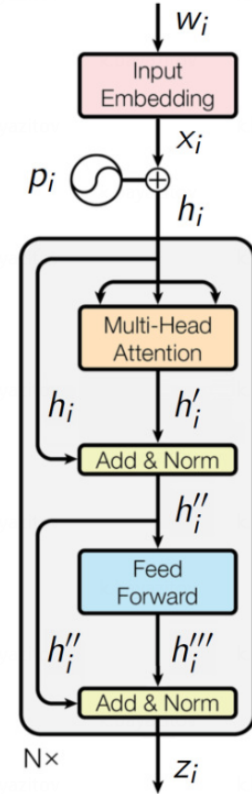


Рис. 12: Трансформер - Кодировщик

Трансформер - Кодировщик обрабатывает входную последовательность и передает свои выходы Трансформеру - Декодировщику.

Для всех  $t = 1, 2, \dots$  :

1. Маскирование данных:  

$$h_t = y_{t-1} + p_t; \quad H_t = (h_1, \dots, h_t)$$
2. Многомерное самовнимание:  

$$h'_t = \text{LN} \circ \text{MH}_j \circ \text{Attn}(W_q^j h_t, W_k^j H_t, W_v^j H_t)$$
3. Многомерное внимание на кодировку Z:  

$$h''_t = \text{LN} \circ \text{MH}_j \circ \text{Attn}(\tilde{W}_q^j h'_t, \tilde{W}_k^j Z, \tilde{W}_v^j Z)$$
4. Двухслойная полносвязная сеть:  

$$y_t = \text{LN} \circ \text{FFN}(h''_t)$$
5. Линейный предсказывающий слой:  

$$p(\tilde{w}|t) = \text{SoftMax}_{\tilde{w}}(W_y y_t + b_y)$$

Генерация  $\tilde{w}_t$

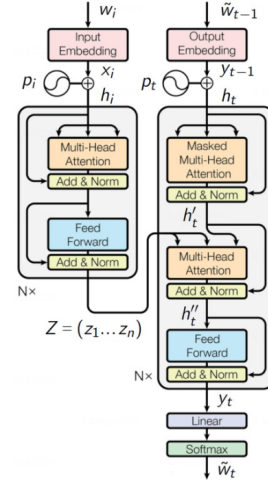


Рис. 13: Трансформер - Декодировщик

Трансформер - Декодировщик использует информацию от Трансформера - Кодировщика и генерирует выходную последовательность.

Проводится сравнение тестируемой модели с моделью, где в качестве дополнительных данных используются ответы модели автоследования.

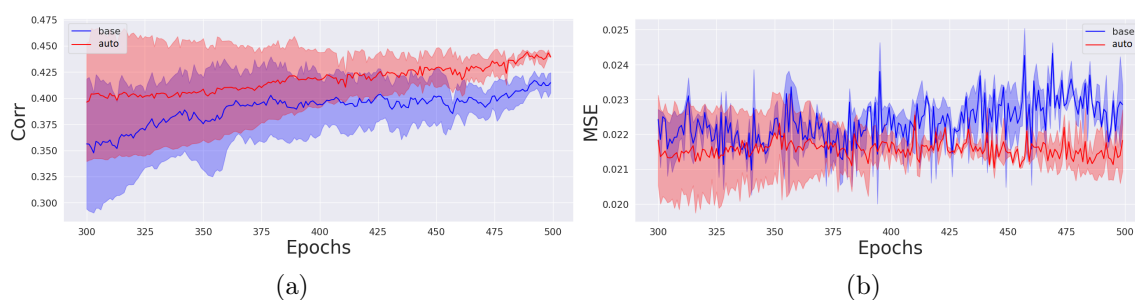


Рис. 14: Качество аппроксимации на тестовой выборке. Все результаты усреднены по 3 запускам. а) Корреляция Пирсона; б) Средняя квадратичная ошибка

На рис. 14 показаны графики зависимости корреляции Пирсона и средней квадратичной ошибки на отложенной тестовой выборке между истинными значениями ряда и ответами модели.

На графиках видно, что модель, использующая ответы модели автоследования, показывает лучшее качество прогнозирования, при этом наблюдается снижение средней квадратичной ошибки.

Таблица 4: Качество моделей

Выборка	Модель	Дополнительные данные	Корреляция Пирсона	MSE
YNDX-Train5	Seq2Seq Transformer	—	$0,379 \pm 0,039$	$0,0225 \pm 0,0005$
		Автоследование	$0,401 \pm 0,035$	$0,0219 \pm 0,0004$
YNDX-Train15	Seq2Seq Transformer	—	$0,415 \pm 0,010$	$0,0229 \pm 0,0006$
		Автоследование	$0,440 \pm 0,001$	$0,0218 \pm 0,0009$
YNDX-Train30	Seq2Seq Transformer	—	$0,406 \pm 0,011$	$0,0211 \pm 0,0010$
		Автоследование	$0,425 \pm 0,009$	$0,0210 \pm 0,0009$

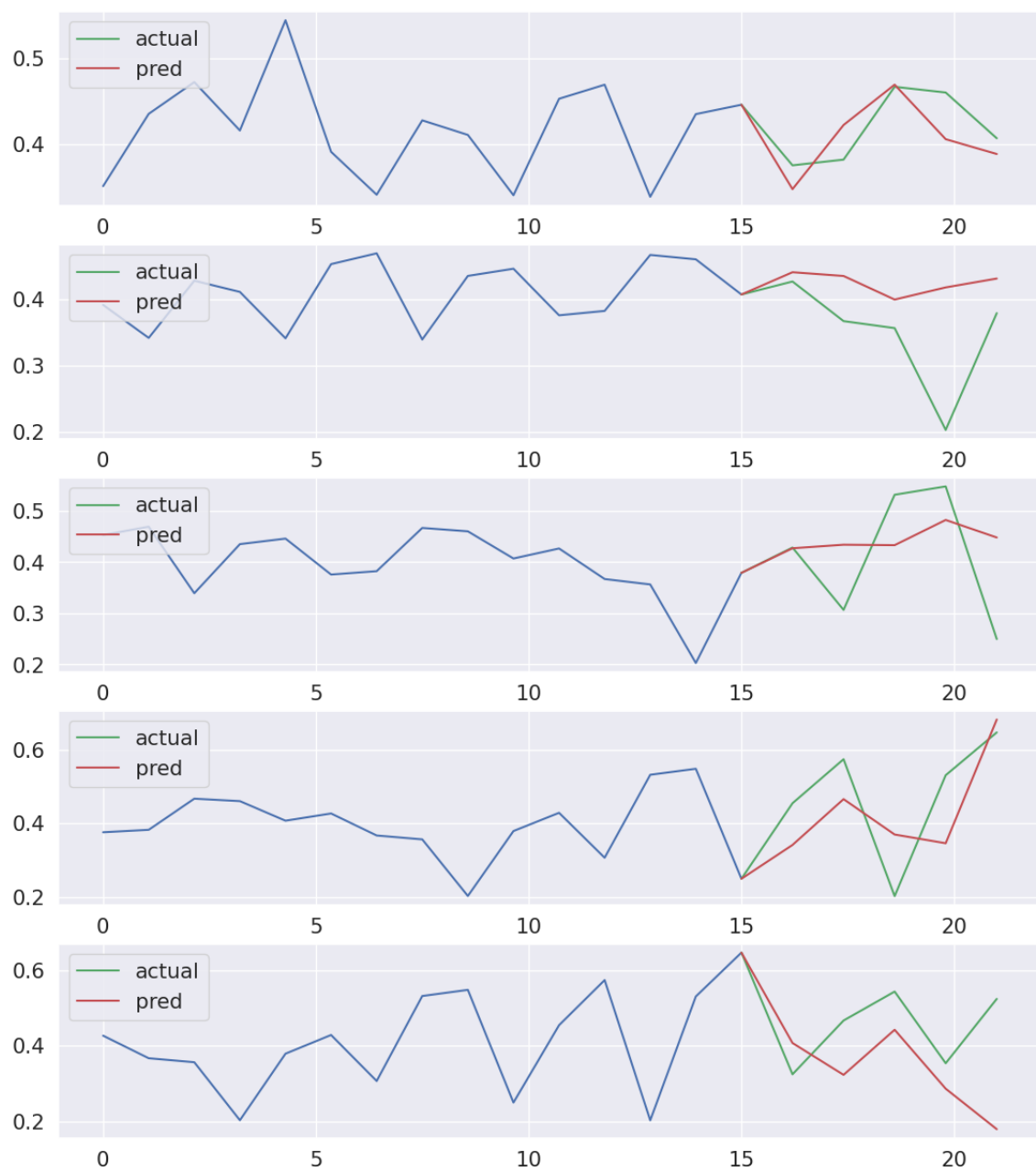


Рис. 15: Прогнозы Transformer

На рис. 15 показаны примеры прогнозов модели Seq2Seq Transformer на отложенной тестовой выборке.

## 4.6 Код вычислительного эксперимента

Весь код вычислительного эксперимента представлен в [23]. Также доступны письменный отчет и результаты экспериментов.

## 5 Заключение

Таблица 5: Результаты экспериментов

Выборка	Модель	Дополнительные данные	Корреляция Пирсона	MSE
	ARIMA		0,340	0,0183
YNDX-Train5	Seq2Seq LSTM	—	$0,476 \pm 0,027$	$0,0175 \pm 0,0003$
		Автоследование	$0,510 \pm 0,036$	$0,0171 \pm 0,0001$
	Seq2Seq Transformer	—	$0,379 \pm 0,039$	$0,0225 \pm 0,0005$
		Автоследование	$0,401 \pm 0,035$	$0,0219 \pm 0,0004$
YNDX-Train15	Seq2Seq LSTM	—	$0,553 \pm 0,004$	$0,0186 \pm 0,0008$
		Автоследование	$0,599 \pm 0,006$	$0,0175 \pm 0,0007$
	Seq2Seq Transformer	—	$0,415 \pm 0,010$	$0,0229 \pm 0,0006$
		Автоследование	$0,440 \pm 0,001$	$0,0218 \pm 0,0009$
YNDX-Train30	Seq2Seq LSTM	—	$0,478 \pm 0,006$	$0,0195 \pm 0,0005$
		Автоследование	$0,538 \pm 0,007$	$0,0186 \pm 0,0017$
	Seq2Seq Transformer	—	$0,406 \pm 0,011$	$0,0211 \pm 0,0010$
		Автоследование	$0,425 \pm 0,009$	$0,0210 \pm 0,0009$

В работе рассмотрена проблема повышения качества моделей прогнозирования временных рядов на примере динамики курса акций. Рассмотрены методы прогнозирования на основе рекуррентных сетей и моделей трансформеров. Был предложен подход включения в модель дополнительных данных — агрегированных знаний опытных инвесторов.

В ходе экспериментов, проведенных на реальных данных, было показано, что предложенный метод работает и повышает качество тестируемых моделей. Результаты экспериментов представлены в таблице 5.

Из таблицы видно, что качество модели зависит от размера входного окна: модели, обученные на выборке с входным окном размера 15, имеют наилучшее качество. Также во всех экспериментах качество тестируемой модели повышается при использовании дополнительных данных модели автоследования.

## Список литературы

- [1] *Лукашин Ю.П.* Адаптивные методы краткосрочного прогнозирования временных рядов. Финансы и статистика, 2003
- [2] *Ariyo A.A., Adewumi A.O.* Stock price prediction using the ARIMA model, 2014
- [3] *Patterson K.* Unit Root Tests In Time Series Volume 1, 2011
- [4] *Herranz E.* Unit root tests, 2017
- [5] *Hochreiter S., Schmidhuber J.* Neural Computation 9(8), 1997
- [6] *Greff K., Schmidhuber J.* LSTM: A Search Space Odyssey, 2017
- [7] *Christopher Olah* Understanding LSTM Networks, 2015  
<https://colah.github.io/posts/2015-08-Understanding-LSTMs>
- [8] *Cho K.* On the Properties of Neural Machine Translation: Encoder–Decoder Approaches, 2014
- [9] *Bahdanau D.* Neural Machine Translation By Jointly Learning To Align And Translate, 2016
- [10] *Dichao Hu* An Introductory Survey on Attention Mechanisms in NLP Problems, 2018
- [11] *Lin Z., Bengio Y.* A Structured Self-Attentive Sentence Embedding, 2017
- [12] *Vaswani A. et al.* Attention Is All You Need, 2017
- [13] *Shaw P., Uszkoreit J., Vaswani A.* Self-Attention with Relative Position Representations, 2018
- [14] *Kingma D. P., Ba J.* Adam: A Method for Stochastic Optimization, 2017



- [15] *Wu N., Green B.* Deep Transformer Models for Time Series Forecasting, 2020
- [16] *Li S., Jin X.* Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting, 2019
- [17] *Wu H., Xu J.* Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting, 2021
- [18] *Zhou H., Zhang S.* Informer: Beyond efficient transformer for long sequence time-series forecasting, 2021
- [19] *Liu S., Yu H.* Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting, 2021
- [20] *Zhou T., Ma Z.* FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting, 2022
- [21] *Liu Y., Hu T., Zhang H.* iTransformer: Inverted Transformers Are Effective for Time Series Forecasting, 2023
- [22] *Воронцов К. В.* Машинное обучение, курс лекций
- [23] *Отчет* <https://github.com/kbayazitov/StockPricingForecasting>