



Faculty of Informatics, Mathematics and
Computer Science

Applied Mathematics and
Informatics

Nizhny Novgorod
2025

Прогнозирование финансовых временных рядов

Выполнил: Данил Соболев, группа 24МАГИАД

Научный руководитель: Кандидат компьютерных наук С.В. Павлов



Актуальность

- Прогнозирование финансовых временных рядов растет с каждым годом.
- Тренд на использование LLM для решения различных задач.

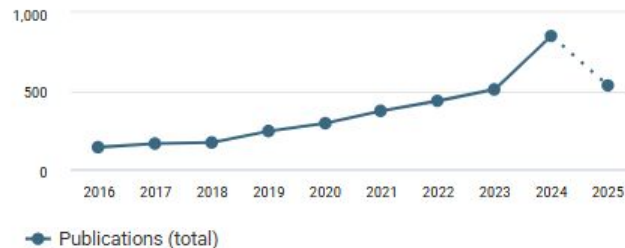
Мотивация

- Мало работ оценивают способности LLM для предсказания временных рядов.
- Отсутствует анализ того, как добавление множества потоков данных влияет на качество предсказаний.

Цели исследования

- Исследовать методы прогнозирования финансовых временных рядов.
- Предложить методологию прогнозирования с использованием нескольких потоков данных с помощью LLM.
- Оценить качество предсказаний различных моделей при использовании разных признаков.

Число публикаций по прогнозированию финансовых временных рядов (2016-2024)



Число публикаций по прогнозированию финансовых временных рядов с помощью LLM (2016-2024)



Рис. 1. Число публикаций (2016–2024). Источник: Dimensions (<https://app.dimensions.ai>)

Временные ряды

Временной ряд — это последовательность значений одного и того же показателя, упорядоченная по времени. Каждому наблюдению соответствует момент времени или порядковый номер.

Классифицируют:

1. Одномерные временные ряды - содержат значения только одного показателя.
2. Многомерные временные ряды - фиксируются одновременно несколько взаимосвязанных показателей.

Метрики качества прогнозов

Средняя абсолютная процентная ошибка:

$$\text{MAPE} = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

Корень среднеквадратичной ошибки:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}$$

Точность направления

$$\text{DA} = \frac{100\%}{n-1} \sum_{t=2}^n \mathbb{I}[(y_t - y_{t-1})(\hat{y}_t - y_{t-1}) > 0]$$

LLM

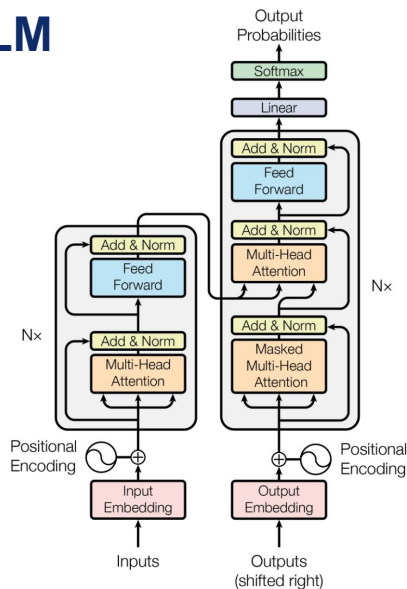


Рис. 2. Архитектура модели трансформера.
Источник: <https://arxiv.org/pdf/1706.03762>

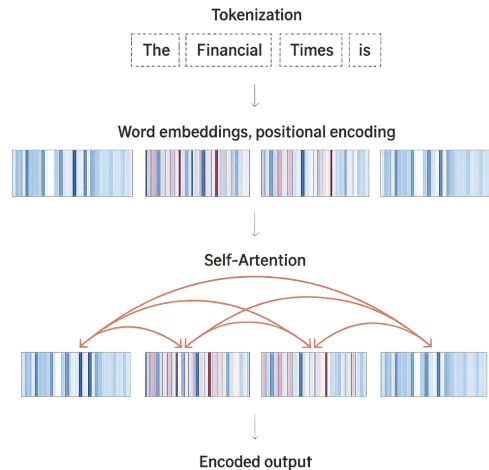


Рис. 3. Кодирование данных. Источник:
<https://ig.ft.com/generative-ai/>

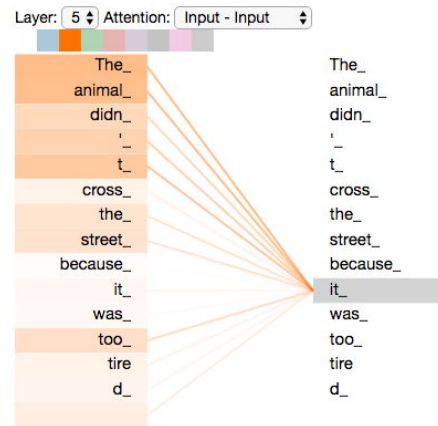


Рис. 4. Определение связей между токенами.
Источник: <http://jalamar.github.io/illustrated-transformer/>

LLMs - это механизмы обнаружения закономерностей, которые угадывают следующий лучший вариант в последовательности.

Одномерные ряды

Для проведения экспериментов использовались ежедневные данные торгов на Московской бирже за период с 2016 по 2025 год по десяти ведущим компаниям индекса МосБиржи. Набор данных охватывает различные отрасли экономики и отражает меняющиеся рыночные условия, имеет длительный временной горизонт и высокую ликвидность, что делает его репрезентативным для задач прогнозирования.

Формат передачи ряда в LLM:

«Ты эксперт по прогнозированию финансовых временных рядов. Последовательность ежедневных цен закрытия: 102.5, 103.0, ..., 102.0, 101.8. Спрогнозируй одношаговый прогноз и верни только число»

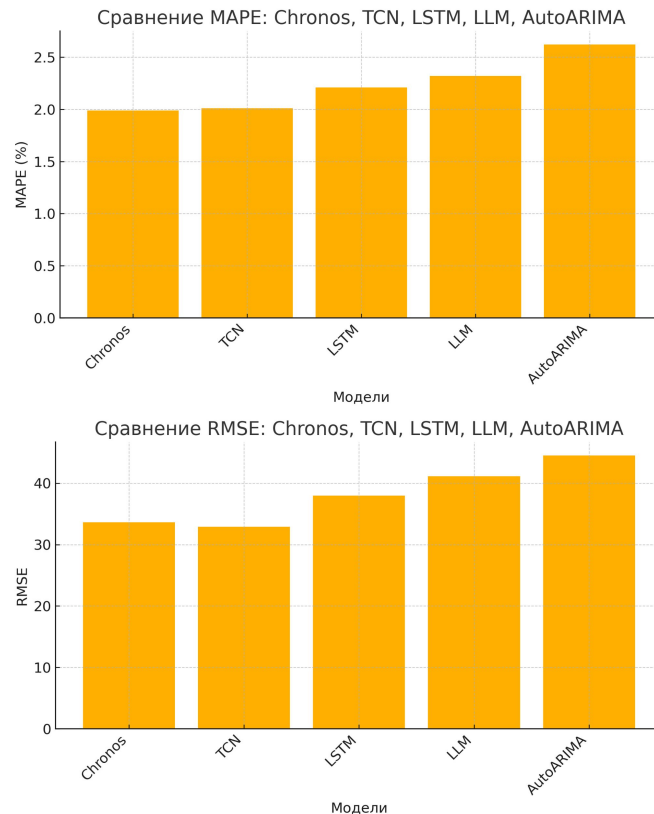


Рис. 5. Метрики прогнозирования одномерных временных рядов

Аномалии и точки перелома

Аномалия - отдельная точка или группа точек, существенно отличающаяся от ожидаемого поведения ряда.

Точка перелома - момент, при котором меняются параметры ряда (среднее, дисперсия, и другие).

В контексте финансовых рядов такие точки позволяют фиксировать смену трендов, начало кризисов и позволяют эффективнее управлять рисками.

Метрики качества детектирования

Точность - доля правильно обнаруженных аномалий среди всех отмеченных моделью

$$\text{Precision} = \frac{|\hat{A} \cap A|}{|\hat{A}|}$$

Полнота - доля правильно обнаруженных аномалий среди всех истинных

$$\text{Recall} = \frac{|\hat{A} \cap A|}{|A|}$$

Гармоническое среднее Precision и Recall, балансирует точность и полноту

$$F_1 = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$



Точечные аномалии

Сравнение алгоритмов детектирования аномалий проводилось на синтетических рядах, сгенерированных по распределению доходностей отрезка данных «Аэрофлота» без явных выбросов.

Формат передачи ряда в LLM:

«Ты — эксперт по детектированию аномалий. Последовательность ежедневных цен закрытия за последние W дней: «102.5, 103.0, ..., 103.8». Определи индексы аномалий (начиная с 1) и верни только номера дней»

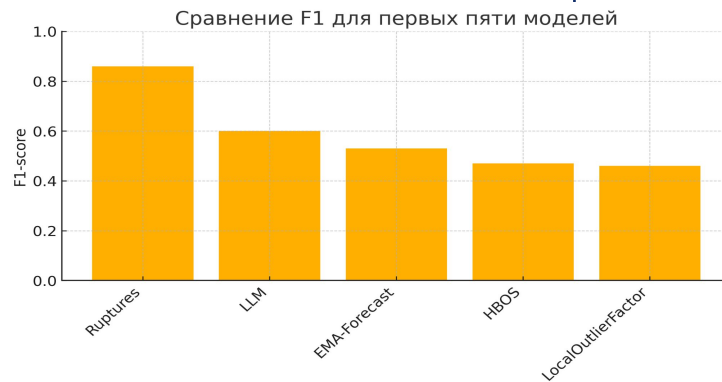


Рис. 6. Метрики детектирования точечных аномалий

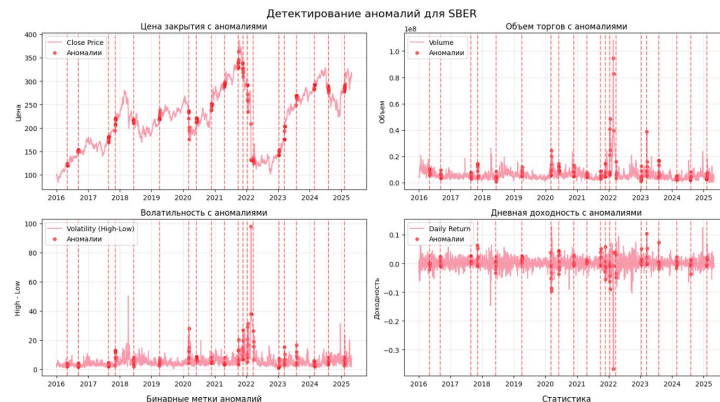


Рис. 7. Результаты детектирования аномалий

Оценка новостного фона

Использованы ежедневные новостные заголовки из открытых источников, привязанные к компаниям.

Оценки были получены с помощью LLM:

“Ты - финансовый эксперт. Оцени влияние новости {заголовок} на цену акции: ПОЛОЖИТЕЛЬНОЕ / НЕЙТРАЛЬНОЕ / НЕГАТИВНОЕ”

Метки были переведены в числа (1/0/-1) и агрегированы по дня для каждой компании.



Рис. 8. Результаты оценки новостного фона

Использованные признаки

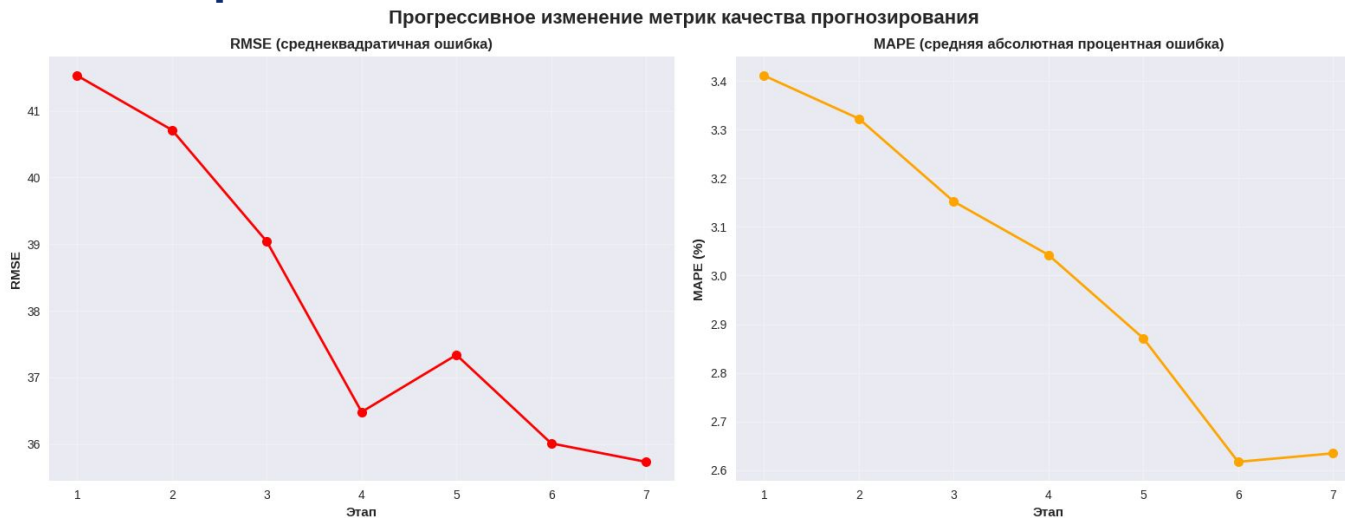
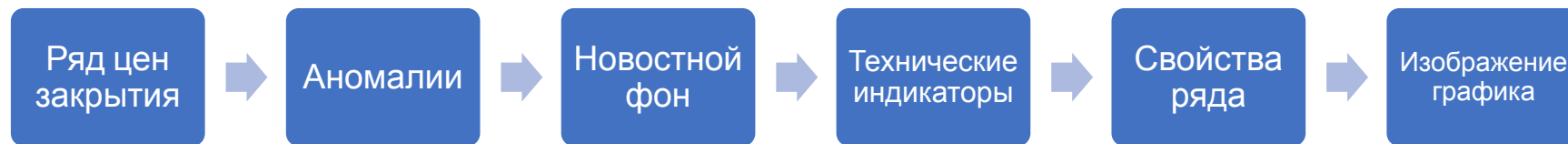


Рис. 9. Изменение метрик RandomForest при повышении количества признаков



Комбинированный подход

«Ты эксперт по прогнозированию финансовых временных рядов. Имеются следующие характеристики временных рядов:

- Цена закрытия (руб.): 102.5, ..., 101.8,
- Объемы торгов (тыс. акций): 10, ..., 25.
- Доходности (в %): +0.49, ..., +6.04.

...

- Аномалии (0 - нет, 1 - да): 0, ..., 1.
- Новостной фон (-100 - крайне отрицательный / 100 крайне положительный): 20, ..., 90.
- [Изображение графика цен и индикаторов]

Спрогнозируй следующее значение ряда цен и верни одно число»

Использование нескольких потоков данных позволяет лучше захватывать зависимости и тенденции временного ряда.

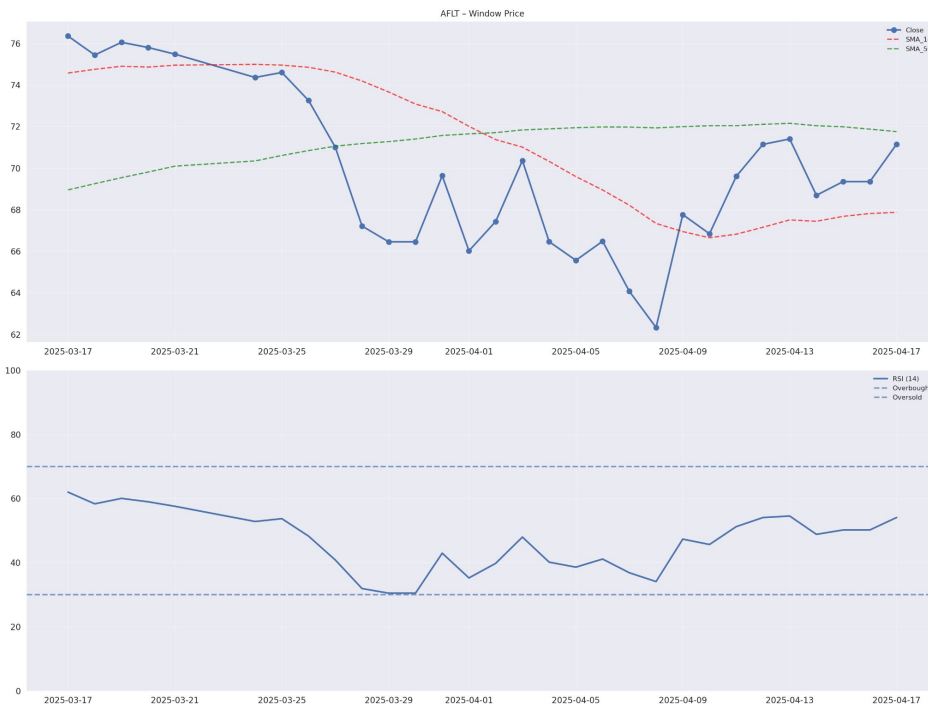


Рис. 10. Визуализация ряда передаваемого LLM

Результаты

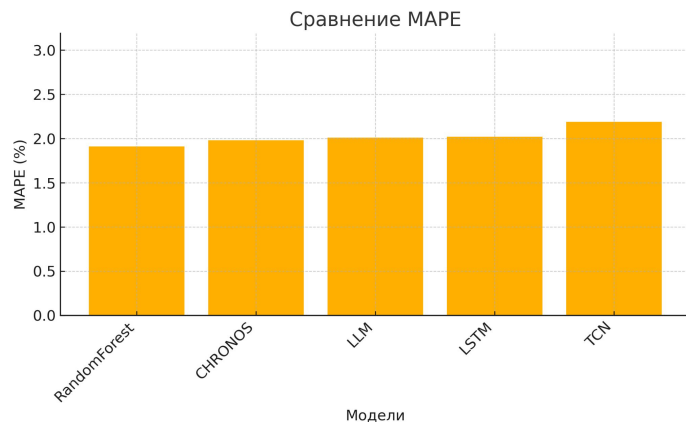


Рис. 11. Метрики моделей при прогнозировании многомерных рядов

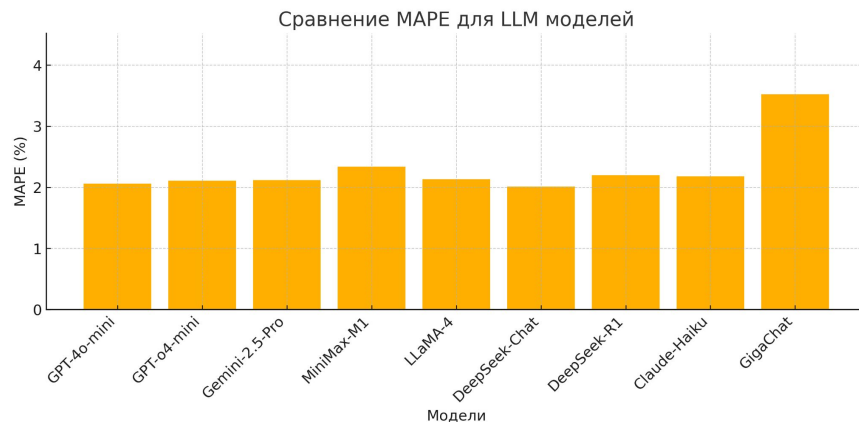


Рис. 12. Метрики LLM моделей при прогнозировании многомерных временных рядов

LLM-модели без дообучения с дополнительными признаками демонстрируют точность, сопоставимую со специализированными и более сложными решениями. При этом между актуальными LLM-моделями нет значимых различий в качестве предсказаний.



Заключение

В исследовании были выполнены следующие задачи:

Получение и использование дополнительных признаков: аномалии, новостной фон, технические индикаторы и свойства рядов.

Разработка комбинированного подхода для LLM.

Прогнозирование одномерных и многомерных временных рядов и оценка моделей.

Результаты:

Комбинированный подход повышает предсказательную точность LLM на $\approx 5\%$.

LLM демонстрируют способность прогнозировать временные ряды даже без дообучения, что упрощает их прикладное применение.

Дальнейшее расширение контекста (соц. сети, отчёты и др.) может обеспечить дополнительное обнаружение скрытых закономерностей и улучшить качество прогноза.

Прогнозирование финансовых временных рядов остается сложной задачей из-за множества неизвестных факторов, снижающих надежность моделей.

