



Факультет экономических наук

Сбор, анализ и предиктивная
аналитика открытых
источников данных

Москва
2025

Предсказание цен активов Российского фондового рынка методами машинного обучения и статистического анализа

Выполнил: Решетнев Матвей Андреевич
Группа: БЭК222



Обоснование темы проекта

Актуальность

- Экономическая значимость фондового рынка
- Практическая ценность для участников рынка

Цель: Разработка и обучение модели, способной прогнозировать значение цены закрытия и направление ее движения с высокой точностью, с учетом показателей технического анализа и неявных зависимостей во временных рядах

Задачи:

1. Провести обзор существующих подходов к прогнозированию финансовых показателей и оценить их эффективность
2. Собрать и подготовить данные: по торгам на Мосбирже, по событиям в компаниях (сплит акций, дивиденды), показатели технического анализа
3. Разработать модели, учитывающие временную зависимость данных и включающую ключевые финансовые и технические показатели
4. Изучить и выбрать метрики для сравнения оценки качества предсказаний моделей
5. Протестировать качество построенных моделей на исторических данных и выбрать наилучшую
6. Создать удобный интерфейс для получения предсказаний модели для заинтересованных пользователей



Этапы работы

1. Сбор и обработка данных
2. Разработка базы данных
3. Исследование и разработка моделей статистического анализа и машинного обучения
4. Валидация моделей и выбор наилучшей модели
5. Разработка телеграм-бота для представления результатов



Сбор и обработка данных

Источник: Мосбиржа API

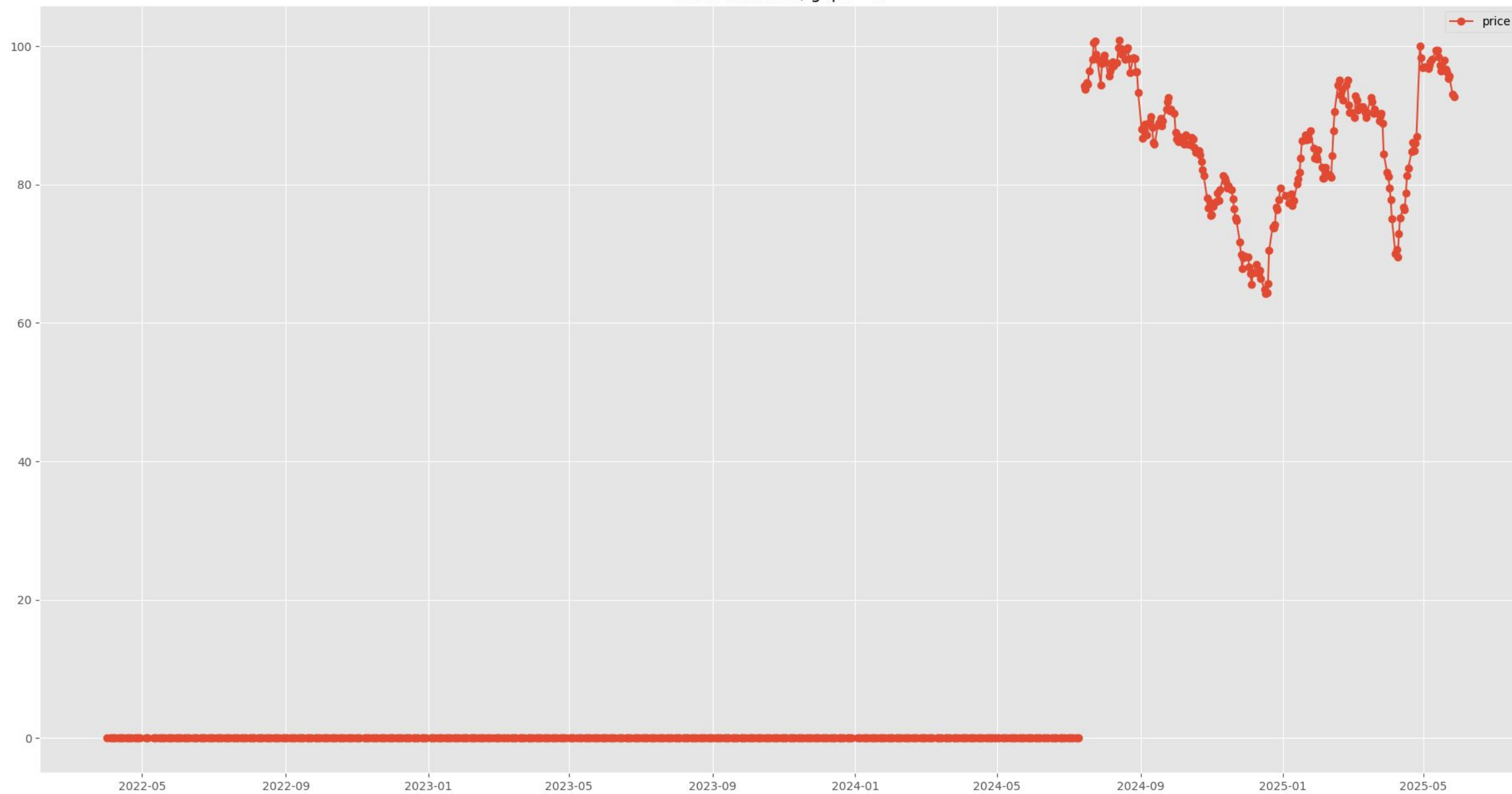
Данные:

- Временные ряды цен открытия, закрытия, средневзвешенные, максимальные, минимальные цены, объем торгов для 44 акций российского рынка
- Даты и объемы выплат дивидендов
- Даты сплитов акций

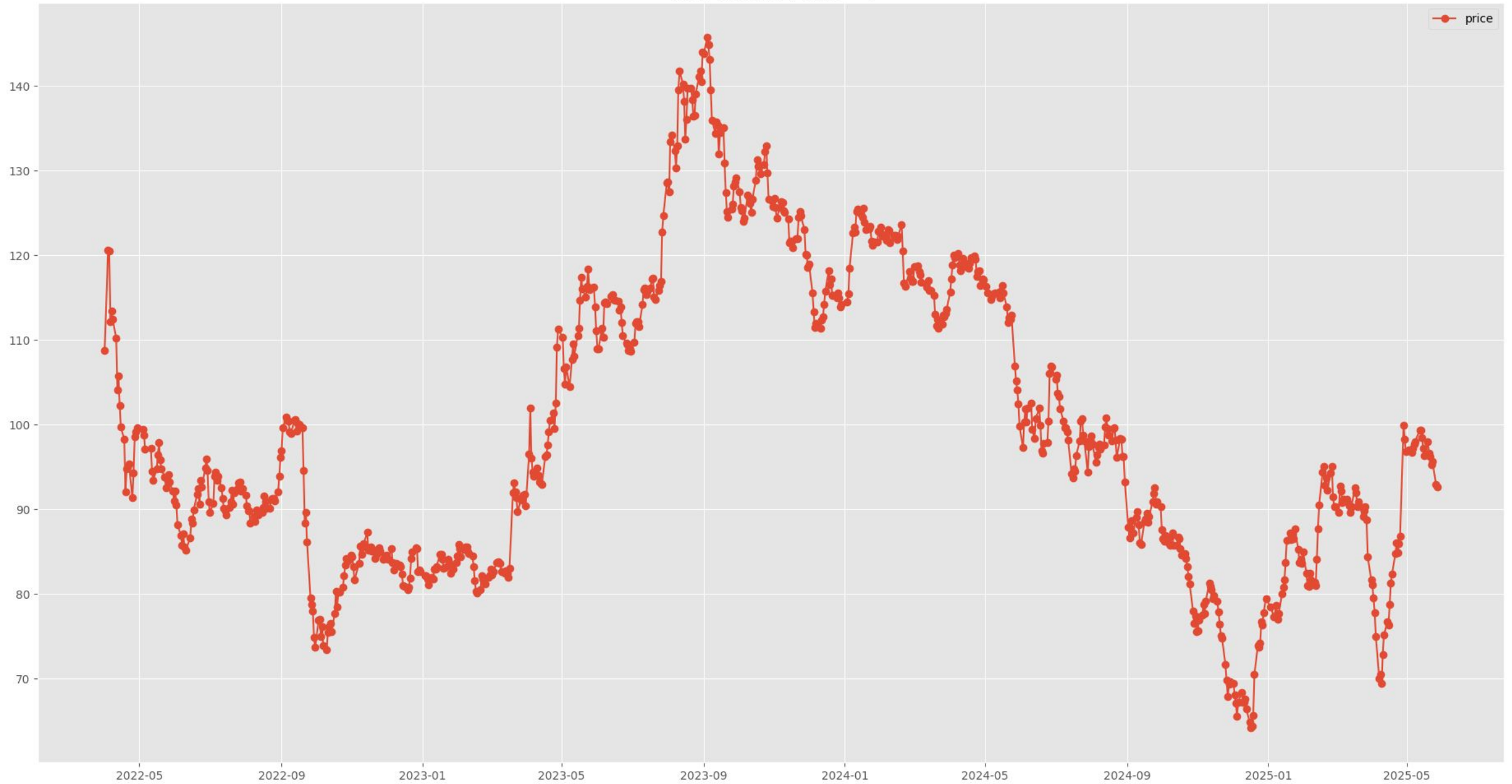
Обработка данных:

- Приведение к одному масштабу после сплита
- Объединение акций, сменивших тикер (TCSB → T, YNDX → YDEX)
- Заполнение пропусков (интерполяция)

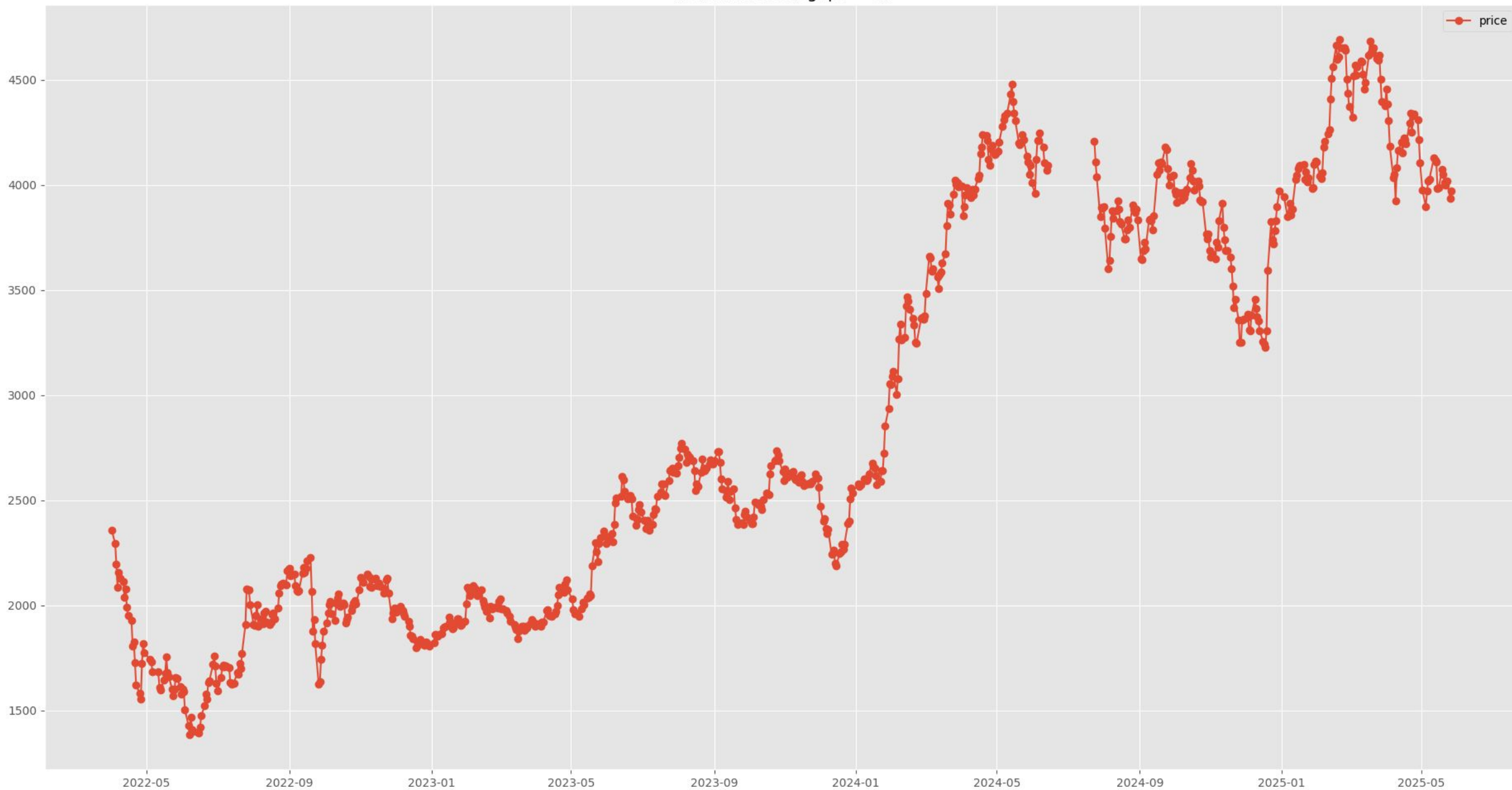
VTBR WAPRICE, gaps = 2



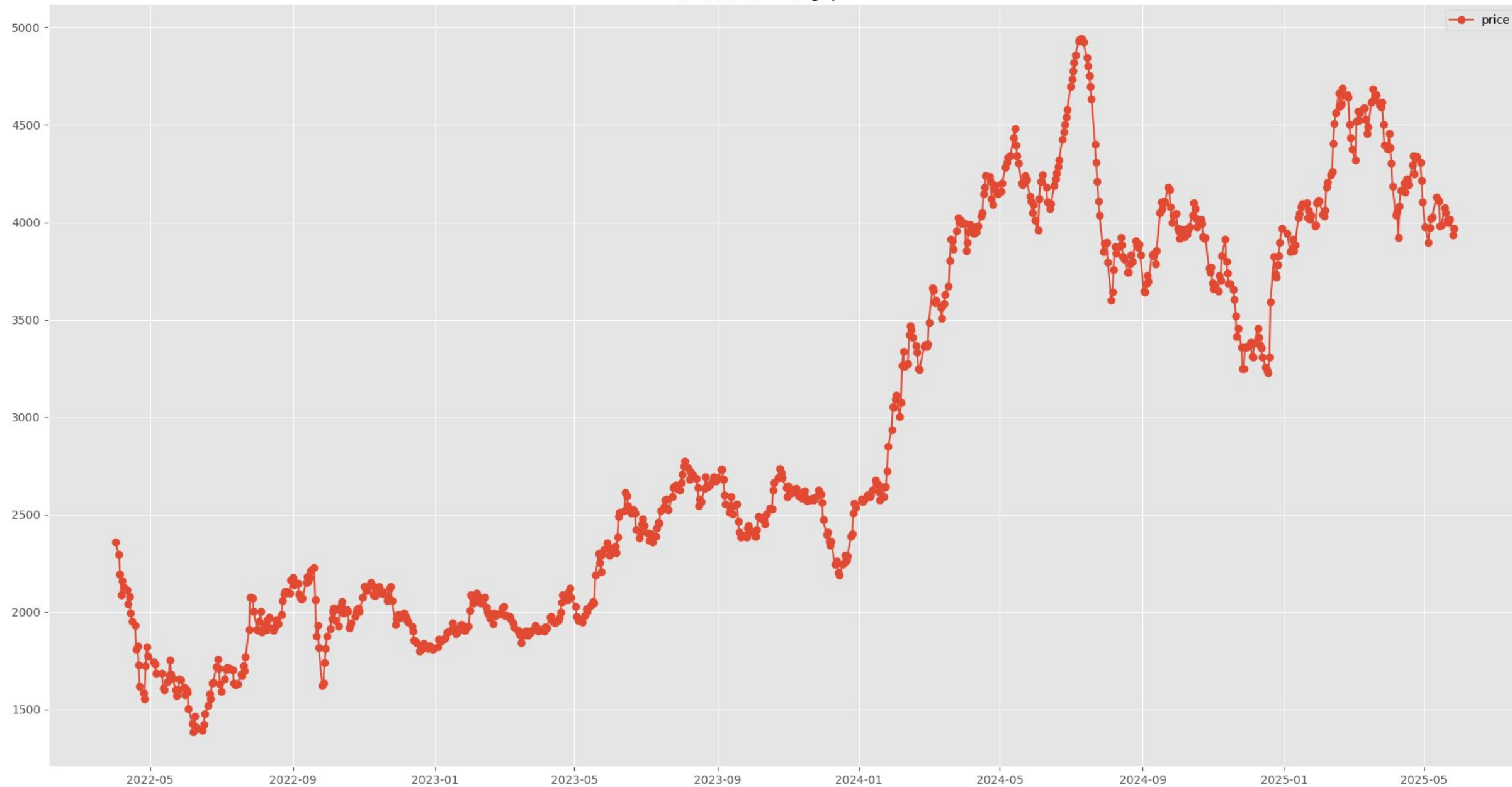
VTBR WAPRICE, gaps = 4



YDEX WAPRICE, gaps = 27



YDEX WAPRICE, gaps = 27





Разработка базы данных

Библиотека: `sqlite3`

Архитектура:

- **prices** — исторические сырые данные по торгам на Мосбирже
- **dividends** — данные о дивидендах компаний
- **splits** — данные о сплитах акций компаний
- **clean_prices** — исторические обработанные данные по торгам на Мосбирже
- **predictions** — обновляемая таблица с предсказаниями моделей

Хранится локально в файле *moex_stock_price_prediction_service.db*



Моделирование

Статистические модели - бейзлайн

- Naive Model
- Auto**ARIMA**
- Auto**CES**
- Auto**ETS**
- AutoRegressive (**AR**)
- RandomWalkWithDrift (**RWD**)
- WindowAverage (**MA**)

Модели машинного обучения

- CatBoostRegressor
- RandomForestRegressor



Модели Deep Learning

Chronos - pretrained модель Amazon

- Chronos_zero_shot - модель без файнтюнинга
- Chronos_finetuned - small-модель с дообучением на наших данных

RNN - рекуррентная нейронная сеть

- RNN_no_exog - обучена на одномерных рядах
- RNN_many_exog - обучена на многомерных рядах с календарными признаками и индикаторами тех. анализа

SOFTS - Series-cOre Fused Time Series

- Обучена на полных многомерных рядах

TimesNet

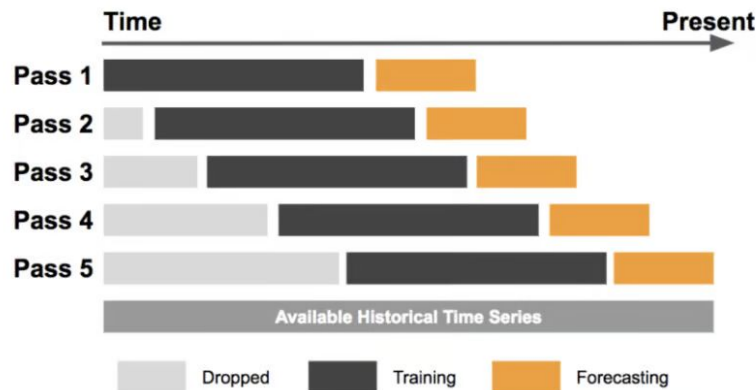
- Обучена на одномерных рядах

StemGNN - Spectral Temporal Graph Neural Network

- Обучена на одномерных рядах

Валидация моделей

Sliding Window



Expanding Window



Метрики:

$$\text{MASE} = \frac{\frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|}{\frac{1}{n-m} \sum_{t=m+1}^n |y_t - y_{t-m}|} \quad \text{MAPE} = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad \text{SMAPE} = \frac{100\%}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{(|y_t| + |\hat{y}_t|)/2}$$

$$\text{DA} = \frac{1}{n-1} \sum_{t=2}^n \mathbf{1}(\text{sign}(y_t - y_{t-1}) = \text{sign}(\hat{y}_t - y_{t-1}))$$

$$\text{HarmonicMetric} = \frac{1}{\frac{1}{\text{MASE}_{\text{mean}}} + \frac{1}{1 - \text{DA}_{\text{mean}}}}$$



Выбор наилучшей модели

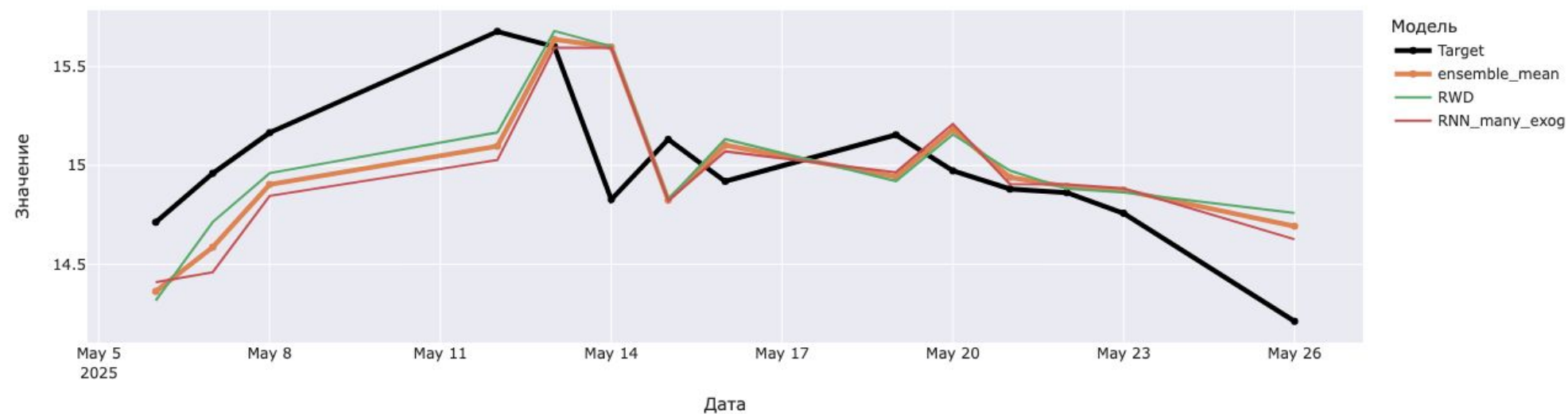
	MASE_median	MASE_mean	MAPE_median	MAPE_mean	SMAPE_median	SMAPE_mean	Directional_Accuracy_median	Directional_Accuracy_mean	Type	harmonic_metric
CatBoostRegressor	1.021628	17.127108	0.015493	0.353800	0.007735	0.066256	1.000000	0.741883	ML/DL	0.508569
Chronos_zero_shot	0.869229	1.022899	0.013832	0.016740	0.006924	0.008344	0.642857	0.548701	ML/DL	0.626284
RWD	0.707623	0.882467	0.011901	0.014074	0.005945	0.007019	0.535714	0.509740	Baseline	0.630334
RNN_many_exog	0.790539	0.940294	0.012786	0.014699	0.006400	0.007354	0.571429	0.524351	ML/DL	0.631735
CES	0.758950	0.908864	0.012322	0.014511	0.006152	0.007233	0.464286	0.485390	Baseline	0.657140
RNN_no_exog	0.776577	0.936866	0.013084	0.014841	0.006536	0.007404	0.357143	0.491883	ML/DL	0.658883
AutoARIMA	0.707527	0.887251	0.012067	0.014194	0.006022	0.007080	0.321429	0.443182	Baseline	0.684230
WindowAverage(7)	1.235269	1.468713	0.021859	0.024129	0.010831	0.011889	0.500000	0.548701	Baseline	0.690442
Naive	0.712339	0.883129	0.011933	0.014119	0.005957	0.007039	0.392857	0.426948	Baseline	0.695077
AutoRegressive	0.763405	0.917551	0.012482	0.014782	0.006233	0.007369	0.357143	0.423701	Baseline	0.707947
StemGNN	1.092148	1.476071	0.018691	0.022710	0.009290	0.011201	0.642857	0.534091	ML/DL	0.708261
AutoETS	0.714716	0.885894	0.011920	0.014173	0.005953	0.007068	0.285714	0.405844	Baseline	0.711272
WindowAverage(6)	1.190863	1.424808	0.021308	0.023368	0.010587	0.011545	0.500000	0.524351	Baseline	0.713206
WindowAverage(14)	1.454360	1.713490	0.026854	0.028568	0.013244	0.013994	0.535714	0.538961	Baseline	0.726581
Chronos_finetuned	0.971523	1.152041	0.015734	0.018233	0.007899	0.009155	0.428571	0.469156	ML/DL	0.726792
WindowAverage(5)	1.152022	1.364199	0.019888	0.022275	0.009906	0.011035	0.500000	0.501623	Baseline	0.730048
WindowAverage(2)	0.865799	1.049940	0.014057	0.016748	0.007021	0.008339	0.392857	0.428571	Baseline	0.740073
RNN_few_exog	0.789637	0.970301	0.013206	0.015296	0.006589	0.007624	0.285714	0.389610	ML/DL	0.749371
WindowAverage(3)	0.997822	1.183505	0.016812	0.019035	0.008389	0.009462	0.357143	0.451299	Baseline	0.749784
WindowAverage(4)	1.090919	1.284377	0.018399	0.020791	0.009180	0.010321	0.428571	0.459416	Baseline	0.760908
SOFTS	11.647752	13.301091	0.200546	0.200199	0.112395	0.112183	0.607143	0.573052	ML/DL	0.827340
TimesNet	1.714129	1.978888	0.031019	0.033339	0.015213	0.016268	0.321429	0.441558	ML/DL	0.871068



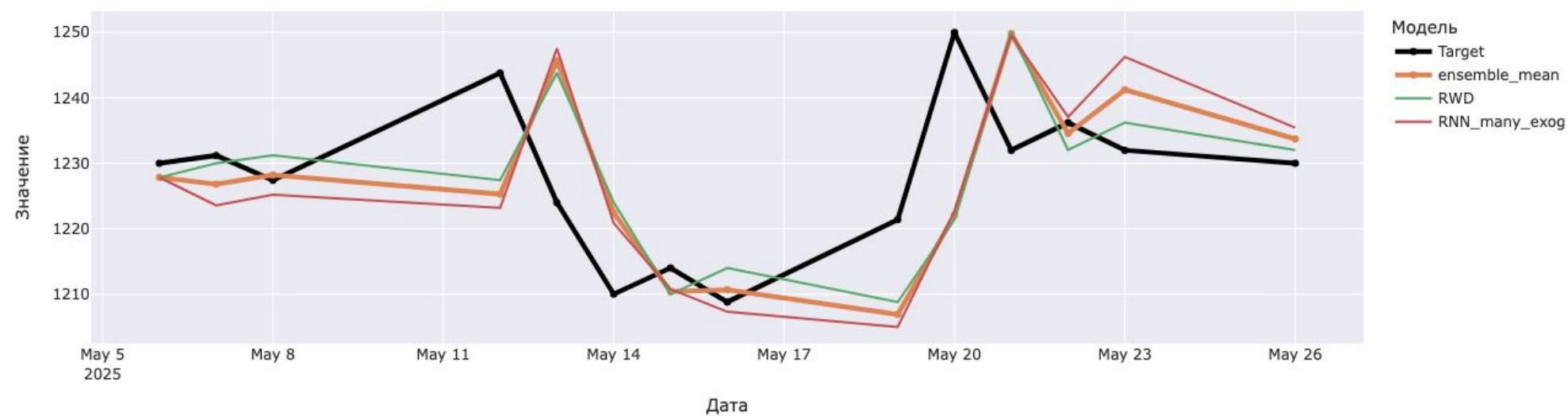
Выбор наилучшей модели

	MASE_median	MASE_mean	MAPE_median	MAPE_mean	SMAPE_median	SMAPE_mean	Directional_Accuracy_median	Directional_Accuracy_mean	Type	harmonic_metric
CatBoostRegressor	1.021628	17.127108	0.015493	0.353800	0.007735	0.066256	1.000000	0.741883	ML/DL	0.508569
ensemble_mean	0.765434	0.906407	0.012192	0.014203	0.006098	0.007094	0.607143	0.524351	ML/DL	0.623899
Chronos_zero_shot	0.869229	1.022899	0.013832	0.016740	0.006924	0.008344	0.642857	0.548701	ML/DL	0.626284
RWD	0.707623	0.882467	0.011901	0.014074	0.005945	0.007019	0.535714	0.509740	Baseline	0.630334
RNN_many_exog	0.790539	0.940294	0.012786	0.014699	0.006400	0.007354	0.571429	0.524351	ML/DL	0.631735
CES	0.758950	0.908864	0.012322	0.014511	0.006152	0.007233	0.464286	0.485390	Baseline	0.657140
RNN_no_exog	0.776577	0.936866	0.013084	0.014841	0.006536	0.007404	0.357143	0.491883	ML/DL	0.658883
AutoARIMA	0.707527	0.887251	0.012067	0.014194	0.006022	0.007080	0.321429	0.443182	Baseline	0.684230
WindowAverage(7)	1.235269	1.468713	0.021859	0.024129	0.010831	0.011889	0.500000	0.548701	Baseline	0.690442
Naïve	0.712339	0.883129	0.011933	0.014119	0.005957	0.007039	0.392857	0.426948	Baseline	0.695077
AutoRegressive	0.763405	0.917551	0.012482	0.014782	0.006233	0.007369	0.357143	0.423701	Baseline	0.707947
StemGNN	1.092148	1.476071	0.018691	0.022710	0.009290	0.011201	0.642857	0.534091	ML/DL	0.708261
AutoETS	0.714716	0.885894	0.011920	0.014173	0.005953	0.007068	0.285714	0.405844	Baseline	0.711272
WindowAverage(6)	1.190863	1.424808	0.021308	0.023368	0.010587	0.011545	0.500000	0.524351	Baseline	0.713206
WindowAverage(14)	1.454360	1.713490	0.026854	0.028568	0.013244	0.013994	0.535714	0.538961	Baseline	0.726581
Chronos_finetuned	0.971523	1.152041	0.015734	0.018233	0.007899	0.009155	0.428571	0.469156	ML/DL	0.726792
WindowAverage(5)	1.152022	1.364199	0.019888	0.022275	0.009906	0.011035	0.500000	0.501623	Baseline	0.730048
WindowAverage(2)	0.865799	1.049940	0.014057	0.016748	0.007021	0.008339	0.392857	0.428571	Baseline	0.740073
RNN_few_exog	0.789637	0.970301	0.013206	0.015296	0.006589	0.007624	0.285714	0.389610	ML/DL	0.749371
WindowAverage(3)	0.997822	1.183505	0.016812	0.019035	0.008389	0.009462	0.357143	0.451299	Baseline	0.749784
WindowAverage(4)	1.090919	1.284377	0.018399	0.020791	0.009180	0.010321	0.428571	0.459416	Baseline	0.760908
SOFTS	11.647752	13.301091	0.200546	0.200199	0.112395	0.112183	0.607143	0.573052	ML/DL	0.827340
TimesNet	1.714129	1.978888	0.031019	0.033339	0.015213	0.016268	0.321429	0.441558	ML/DL	0.871068

CLOSE — прогнозы лучших моделей для AFKS



CLOSE — прогнозы лучших моделей для TRNFP



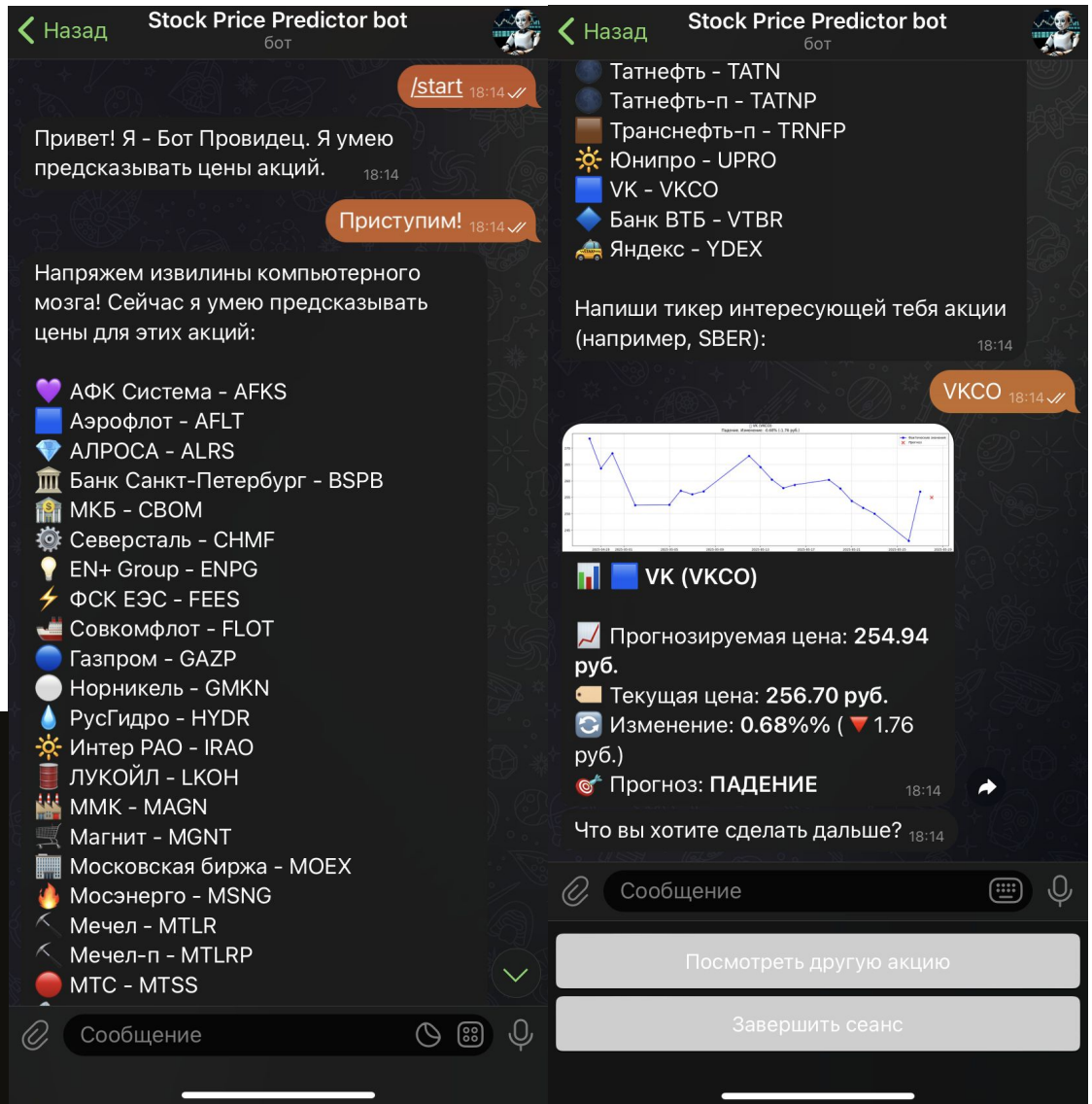
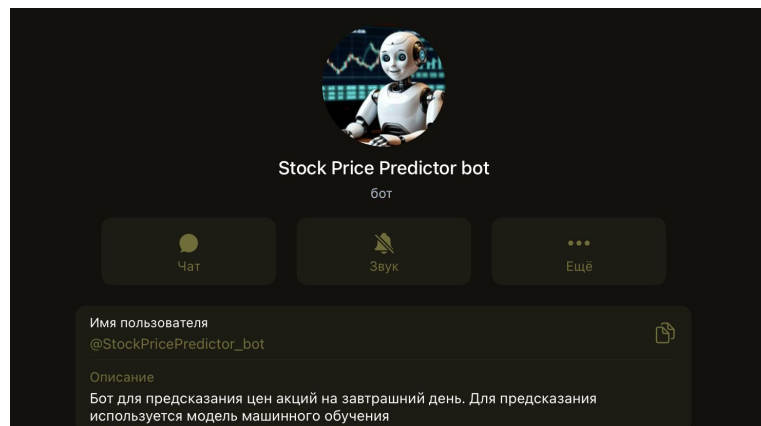


Разработка бота

Разработан на основе Telegram API

Функции:

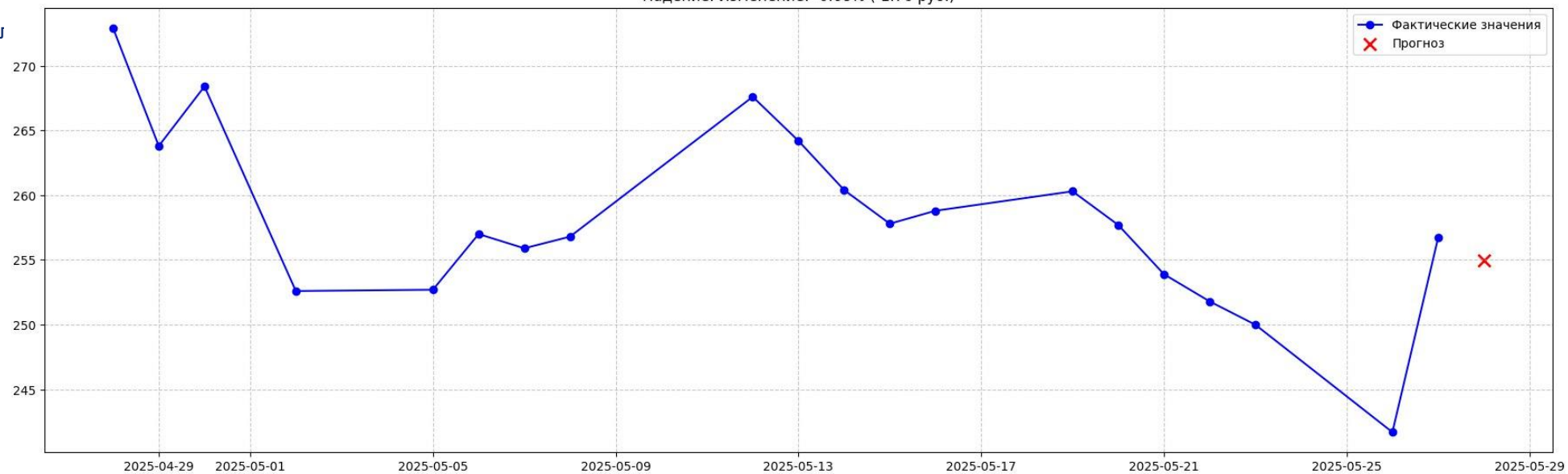
- Запуск по кнопке `/start`
- Приветствие, список доступных акций
- Обращение к таблице предсказаний
- Прогноз на день вперед для выбранной акции
- Генерация графика





Факул

VK (VKCO)
Падение. Изменение: -0.68% (-1.76 руб.)



ЛУКОЙЛ (LKOH)
Падение. Изменение: -0.32% (-20.64 руб.)

