

FORECAST.AI

Topic: Forecast

Team: back2back

FINAM X HSE AI TRADE HACK

Преимущества и особенности

Быстрота

Предсказание цены занимает < 20 min

Глубина

Модель учитывает сентимент новостей

Полнота

21 метрика для обучения модели

Точность

84%

Использованные модели

Finbert.

Анализ сентимента новостей
Модель специально обучалась
на финансовых новостях

Ансамбль моделей

Была реализована ансамблевая архитектура,
основанная на комбинации нескольких алгоритмов
регрессии.

Основу составляет ансамбль из пяти разнотипных
моделей:

- Random Forest Regressor
- Gradient Boosting Regressor
- Extra Trees Regressor
- Ridge Regression
- Lasso Regression

Ансамбль моделей

Модель включает четыре компонента:

- Метод случайного леса
- Градиентный бустинг над гистограммами
- Ридж-регрессию с L2-регуляризацией
- Метод экстра-деревьев

Ключевая особенность

Взвешенное агрегирование предсказаний, где каждому алгоритму присвоен коэффициент значимости на основе его прогностической способности

Преимущества выбранного подхода заключаются в следующем:

- 1) Комбинация разнородных алгоритмов позволяет охватить различные типы зависимостей в данных — как линейные, так и нелинейные.
- 2) Ансамблирование снижает дисперсию ошибок и уменьшает риск переобучения за счет балансировки индивидуальных погрешностей моделей.
- 3) Взвешенное голосование обеспечивает оптимальное использование сильных сторон каждого алгоритма при минимизации их слабых мест.

Метрики

ОСНОВНЫЕ МЕТРИКИ

- 1. MAE (Mean Absolute Error)
- 2. MAE_norm (Нормированный MAE)
- 3. Brier Score
- 4. Brier_norm (Нормированный Brier)
- 5. DA (Directional Accuracy)
- 6. Final_Score

МЕТРИКИ ДЛЯ OPEN

- 7. MAE_open
- 8. RMSE_open
- 9. MAPE_open

МЕТРИКИ ДЛЯ HIGH

- 10. MAE_high
- 11. RMSE_high
- 12. MAPE_high

МЕТРИКИ ДЛЯ LOW

- 13. MAE_low
- 14. RMSE_low
- 15. MAPE_low

МЕТРИКИ ДЛЯ CLOSE

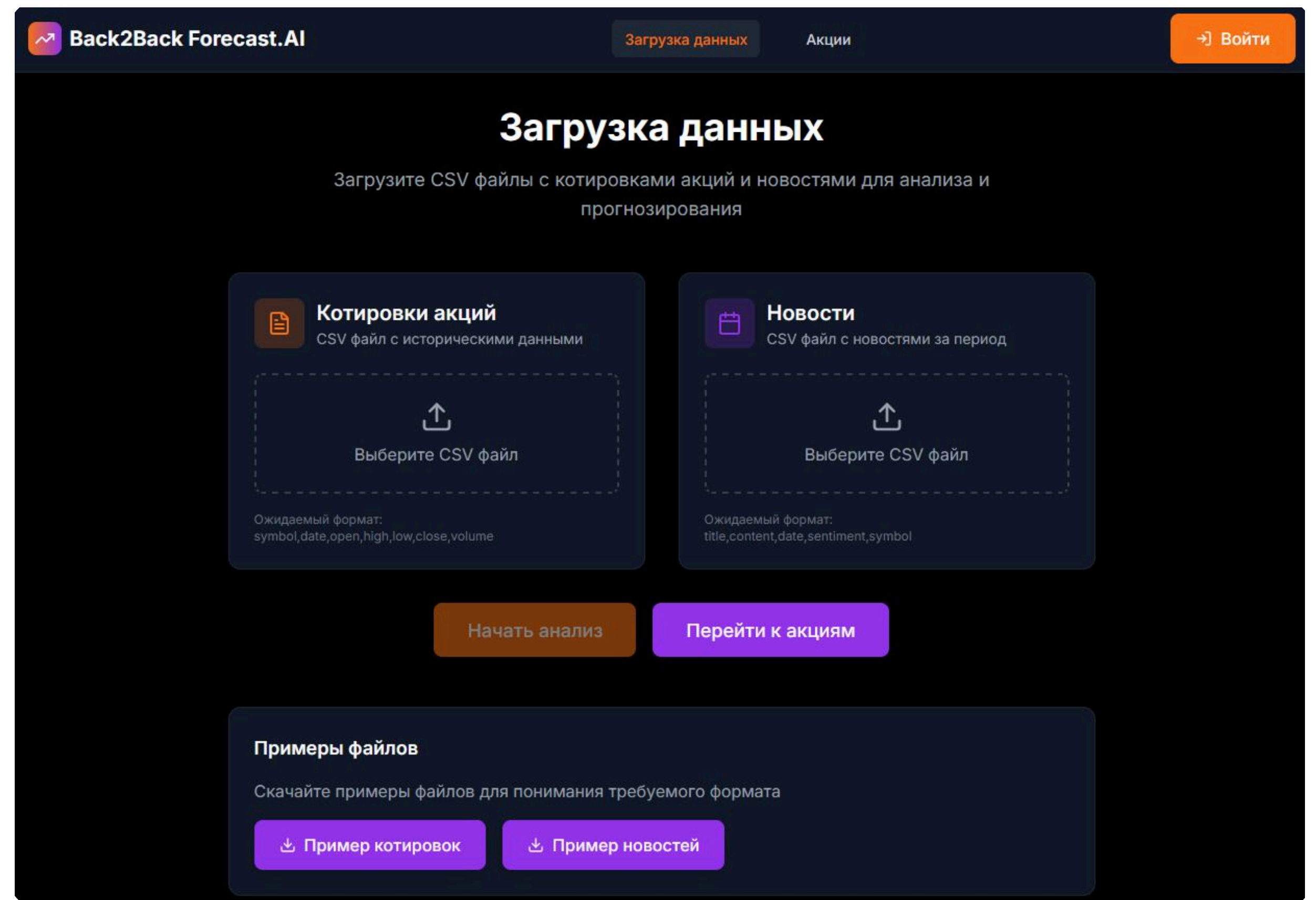
- 16. MAE_close
- 17. RMSE_close
- 18. MAPE_close

МЕТРИКИ ДЛЯ VOLUME

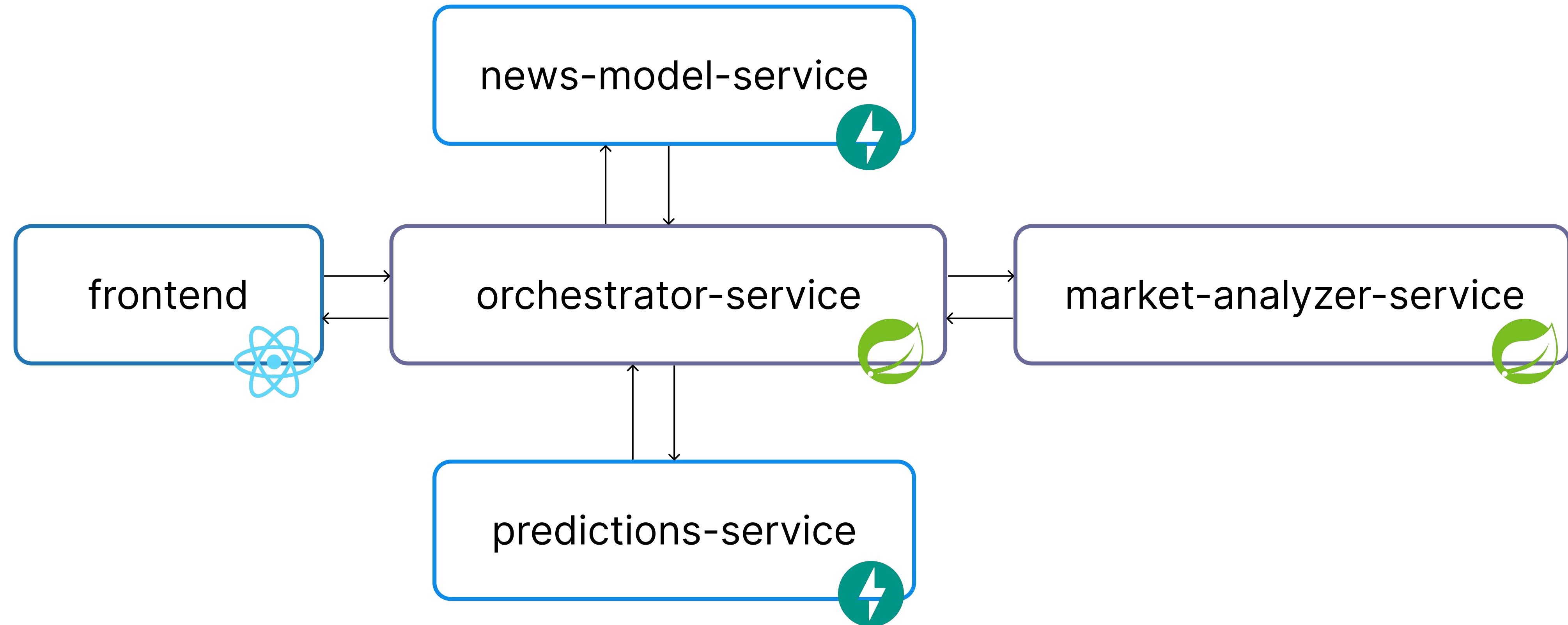
- 19. MAE_volume
- 20. RMSE_volume
- 21. MAPE_volume

Взаимодействие с сервисом

- Загрузить CSV-файлы котировок и новостей
- Нажать на кнопку <Начать анализ>



Архитектура



Design



хороший фронтенд - не работающий фронтенд

back2back teem



Даниил Лулаков

front + ML
DevOps + back



Виктор Гезенцвей

back + DevOps



Валерий Беляев

Analyst + ML



Полина Сорокина

Hard ML