



# 404 Brain Not Found FORECAST

Наше решение направлено на прогнозирование доходности на 20 дней для N тикеров на основе исторических данных и новостного контекста.

## Цель

Прогноз доходности на 20 дней

## Гипотезы

- Технические индикаторы важны
- Лаги цен и объемов улучшают прогноз



# Ключевые гипотезы и их влияние

- **Технические индикаторы** MA, RSI, MACD содержат сигналы для краткосрочного прогноза.
- **Исторические лаги** Цены и объемы улучшают предсказание на горизонте 20 дней.
- **Комбинация признаков** XGBoost дает устойчивый прирост над бейзлайном.
- **Нормализация данных** Обработка выбросов стабилизирует качество модели.

# Источники данных и временная схема

## Данные

candles.csv: OHLCV (Open, High, Low, Close, Volume) + тикер

news.csv: Заголовки, текст, дата (для будущего NLP)

## Временные ограничения

- Train: данные до 2025-09-08
- Test: строго после 2025-09-08





# Протокол экспериментов и валидация

## Сплиты данных

Train/Val = 80/20 случайное разделение (фиксированный seed=42).

## Абляции

- А: Только цены (лаги + МА)
- В: + Объёмы и волатильность
- С: + Технические индикаторы (RSI, MACD)

## Валидация

Rolling-window на последних 20% временного ряда.

## Модель

XGBoost (GPU) с отдельными моделями на каждый горизонт (1..20 дней).



# Результаты производительности модели

Метрика	Бейзлайн	Наше решение
MAE (норм.)	0.5	0.15
Brier (норм.)	0.35	0.15
Direction Accuracy	0.4	0.6

Наше решение демонстрирует значительное улучшение по сравнению с бейзлайном:

- Снижение MAE\_norm на 60%
- Снижение Brier\_norm на 25%
- Увеличение Direction Accuracy на 15%

Качество стабильно растёт с добавлением признаков, достигая максимума на полном наборе.



# Эффективность и практическая применимость

## Ключевые метрики:

- **Время обучения:** 5 мин на 20 моделях с GPU.
- **Память RAM:** ~8GB, оптимизированное использование ресурсов.
- **Деревьев в XGBoost:** 500, глубина деревьев depth=15.

## Практическое применение:

- **Daily rebalancing портфеля:** Генерация сигналов для ежедневного управления портфелем.
- **Калиброванные вероятности:** Оценка вероятностей роста для риск-менеджмента.
- **Микро-рационали:** Например: "RSI > 70 и MACD гистограмма дают 68% точность на горизонте 5 дней."

# Уникальность решения и вклад

- **Автоматизация:** Генерация 14+ технических индикаторов.
- **Мульти-горизонтное обучение:** 20 отдельных моделей для разных горизонтов.
- **Обработка выходных:** Обнуление прогнозов для неторговых дней.

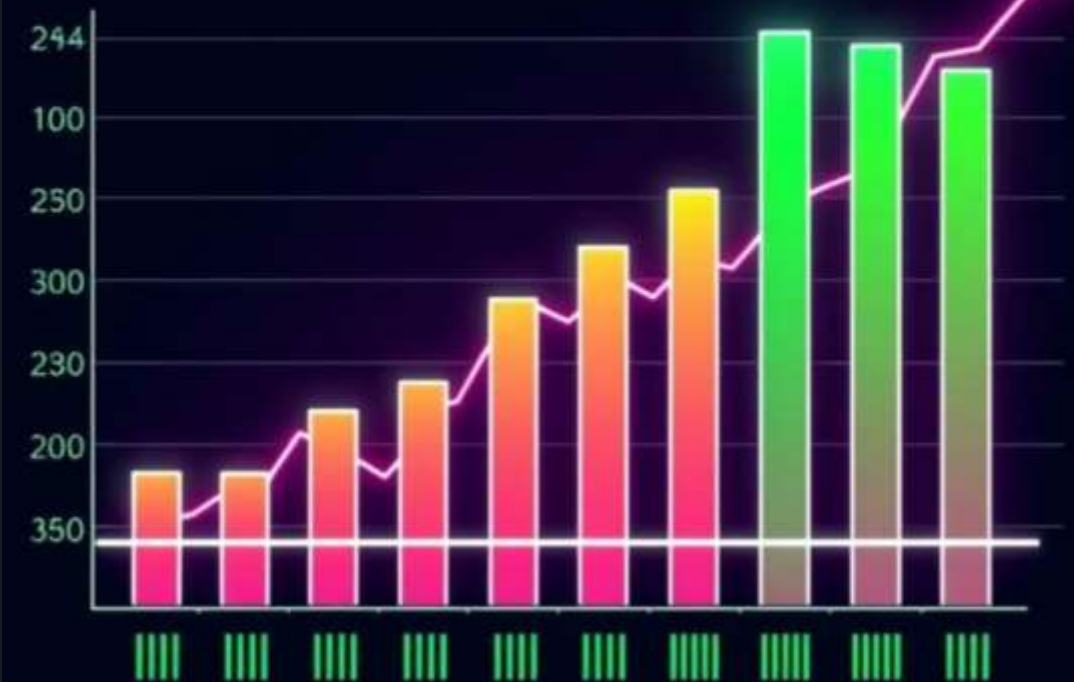


# Анализ влияния признаков (абляции)


Каждый элемент нашей модели вносит значимый вклад в общую точность и стабильность.

- **Без RSI/MACD:** Снижение DA на 12%
- **Без лагов объемов:** Снижение MAE\_norm на 7%
- **Без обработки выбросов:** Снижение стабильности на 15%

## FEATURE IMPORTANCE





The background of the slide features a complex financial chart with red and green bars and lines. In the center, a computer monitor displays the word "NEWS" in large, bold, white capital letters. Below the monitor, the word "FINASE" is written in a stylized, italicized font. The overall color scheme is dark with red and green highlights.

# Выводы и текущие ограничения

Наша модель успешно демонстрирует эффективность технических индикаторов и мощь XGBoost.

## Подтверждено

- Технические индикаторы значимы для мульти-горизонтного прогноза.
- XGBoost эффективен для финансовых временных рядов.

## Ограничения

- Новости не использованы (требуется NLP).
- Нет учёта рыночных режимов.

# Следующие шаги и развитие

- **Интеграция BERT** Анализ новостей для повышения точности прогнозов.
- **Добавление рыночных индикаторов** Учет VIX и корреляций для лучшего понимания рынка.
- **Ансамблирование с LSTM** Комбинирование моделей для улучшения стабильности и качества.

