

# Отчёт о разработке и развертывании сервиса предсказания цен на основе текстовых данных

## Описание проекта

Проект направлен на создание сервиса, который предсказывает изменения цен на финансовые активы, такие как нефть (WTI, Brent) и фондовые индексы (S&P 500, Nasdaq-100). Основное внимание уделено обработке информации: текстовой (новости, аналитика) и её интеграции с рыночными данными для построения точных прогнозов.

## Архитектура сервиса

Сервис реализован с использованием:

- **Streamlit** как пользовательский интерфейс.
- **Виртуального частного сервера**, на котором развернут Docker-контейнер для обеспечения производительности и изоляции.
- API для загрузки рыночных данных и новостей.

---

## Предобработка данных

### 1. Сбор данных

- Рыночные данные собираются через API биржевых и финансовых платформ.
- Новостная информация по ключевым активам извлекается из новостного агрегатора через API.

### 2. Обработка рыночных данных

- Расчёт **технических индикаторов**: средние скользящие (SMA), экспоненциальные скользящие (EMA), волатильность, объемы торгов и т.д.
- Формирование временных рядов с учётом различных временных лагов.

### 3. Обработка текстовых данных

- Сентимент-аналитика с использованием **FinBERT**:
  - Определение тональности новостей (позитивная, негативная, нейтральная).
  - Агрегация сентиментов на уровне дня.
- Построение производных метрик:
  - Скользящие средние сентиментов за неделю и месяц.
  - Разности и ускорения (акселерации) сентиментов.

## 4. Интеграция данных

- Все обработанные данные объединяются в единый датасет.
  - Происходит отбор наиболее значимых признаков для минимизации переобучения.
- 

# Основной функционал

## 1. Анализ данных

- Визуализация временных рядов:
  - Линейные и свечные графики.
  - Гистограммы распределения метрик.
  - Корреляционные матрицы.
- Расчёт описательной статистики (средние значения, медианы, квантили).

## 2. Модели и прогнозирование

- Предобученные модели **Ridge-регрессии** для прогнозирования цен активов.
- Возможность обучения моделей на пользовательских датасетах (при наличии необходимых столбцов).
- Построение прогнозов на заданный временной горизонт.

## 3. Загрузка пользовательских данных

- Пользователи могут загружать собственные датасеты в формате CSV.
  - Интеграция данных в аналитическую платформу с последующей обработкой и прогнозированием.
- 

# Текущие результаты

- Построен и развернут сервис, позволяющий:
  - Проводить анализ временных рядов и текстовой информации.
  - Генерировать прогнозы на основе предварительно обученных моделей.
  - Настраивать параметры анализа для различных активов.
- Визуализированы результаты модели:
  - Сравнение фактических и предсказанных значений.
  - Остаточный анализ (распределение ошибок, Q-Q графики).