Инструкция пользователя: Интерфейс предсказаний для молекулярных данных

1. Общая информация

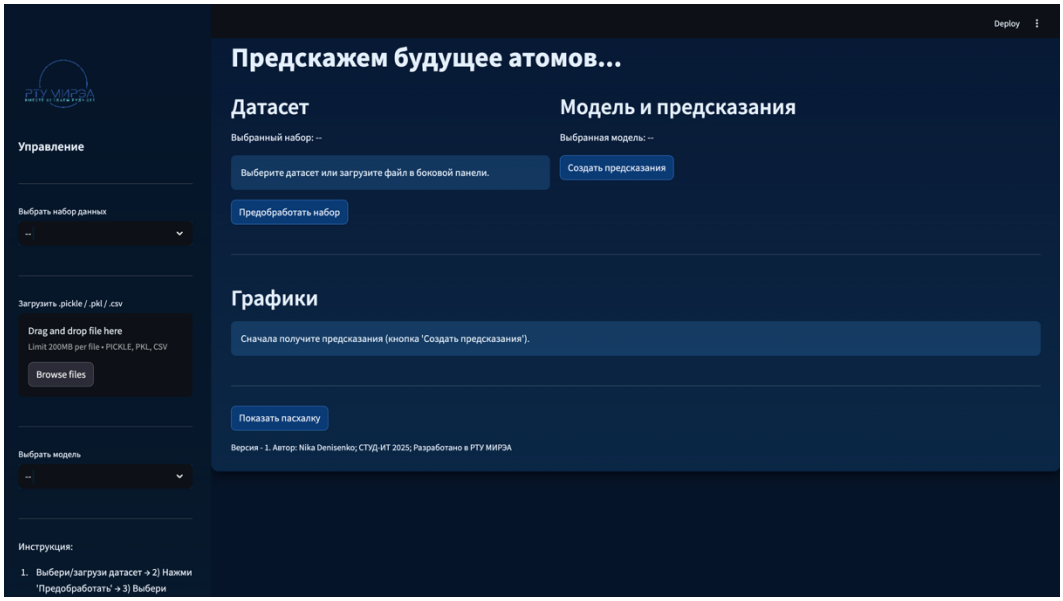
Приложение предназначено для предобработки данных о молекулах и создания предсказаний с использованием машинного обучения. Пользователь может загружать собственные датасеты, выбирать модели, предобрабатывать данные и визуализировать результаты.

Версия приложения: **1.0**
Автор: **Вероника Денисенко, СТУД-ИТ 2025**
Разработка: **РТУ МИРЭА**

---

2. Запуск и интерфейс

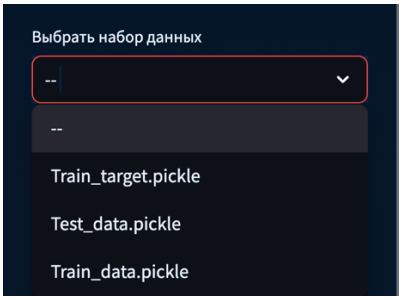
1. Запустите приложение Streamlit. streamlit run app_streamlit.py
2. Интерфейс разделён на две основные колонки:
    - **Левая колонка:** Работа с датасетом
    - **Правая колонка:** Выбор модели и создание предсказаний



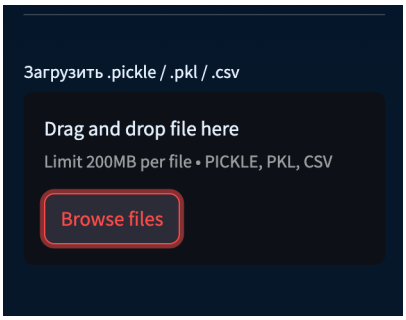
3. Работа с датасетом

3.1 Выбор или загрузка данных

- В боковой панели выберите один из доступных локальных файлов (`.csv`, `.pickle`, `.pkl`) в списке **"Выбрать набор данных"**.



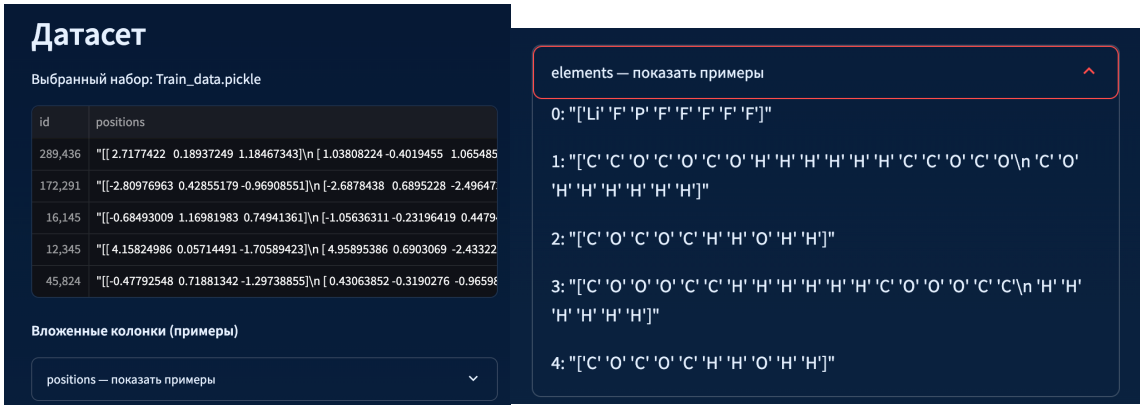
- Для загрузки собственного файла используйте **"Загрузить .pickle / .pkl / .csv"**.



- После загрузки файл автоматически добавляется в сессию.

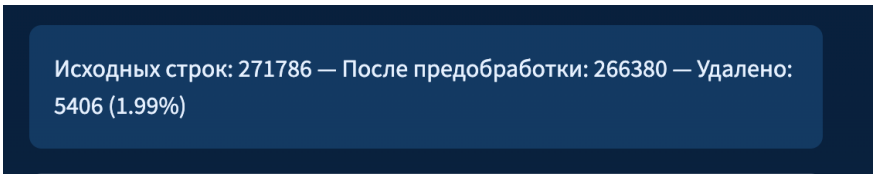
3.2 Просмотр данных

- После выбора датасета в левой колонке отображается предварительный просмотр первых строк.
- Если в колонках содержатся сложные структуры (списки, массивы), для них показываются укороченные примеры.
- Разверните вкладки с вложенными колонками для просмотра детализированных данных.



3.3 Предобработка данных

1. Нажмите кнопку **"Предобработать набор"**.
2. Приложение выполнит:
    - Очистку данных (удаление строк без координат, с выбросами энергии или заряда).
    - Генерацию новых признаков (количество атомов, связи, расстояния, градиенты и дипольные моменты).
    - Удаление исходных колонок с массивами.
3. После завершения предобработки появится информация:
    - Количество исходных строк
    - Количество удалённых строк
    - Список признаков после предобработки
4. Если строки были удалены, их можно просмотреть в отдельной вкладке **"Показать примеры удалённых строк"**.

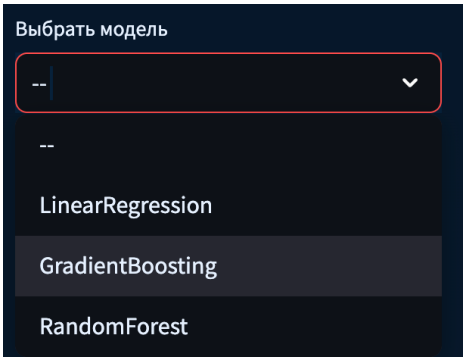


---

4. Работа с моделями

4.1 Выбор модели

- В боковой панели выберите модель из списка:
    - Загруженные модели с диска (`.pkl`)
    - Встроенные модели: `LinearRegression`, `RandomForest`, `GradientBoosting`



4.2 Создание предсказаний

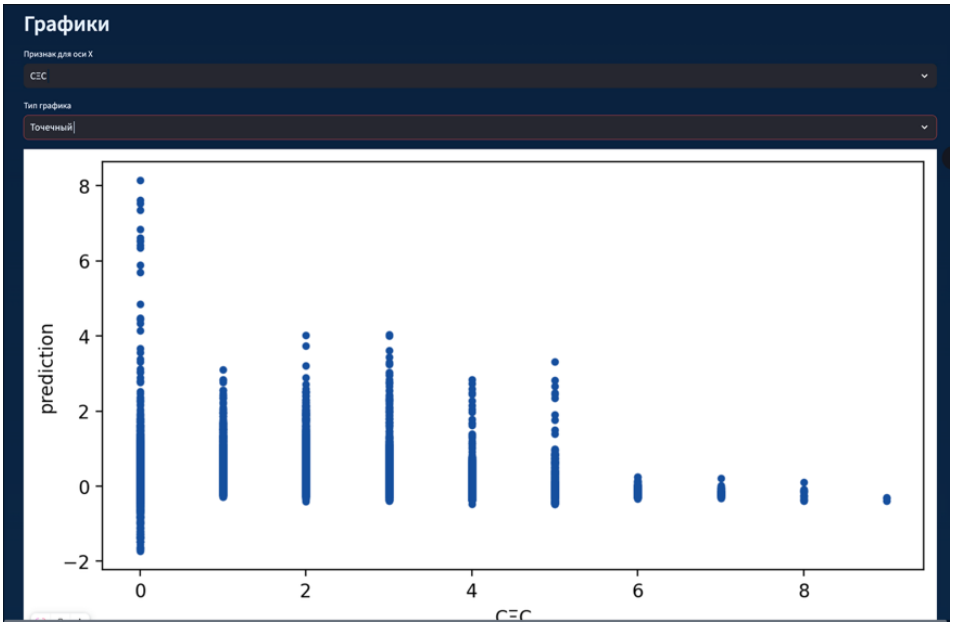
1. Убедитесь, что выбран предобработанный датасет и модель.
2. Нажмите **"Создать предсказания"**.
3. Результаты сохраняются в сессии и отображаются в правой колонке:
    - Таблица предсказаний
    - Возможность скачать CSV-файл с результатами

– Просмотр конкретного предсказания по выбранному индексу с указанием примерной неопределённости

---

5. Визуализация предсказаний

1. В разделе **"Графики"** выберите:
   - Признак для оси X (`index` или любой числовой признак после предобработки)
   - Тип графика: **Линейный**, **Точечный**, **Гистограмма**
2. Приложение автоматически фильтрует некорректные значения (`NaN`, `Inf`).
3. Графики отображаются непосредственно в интерфейсе.



---

6. Дополнительно

- **Информация об авторе:** кнопка **"Об авторе"** в боковой панели
- **Пасхалка:** кнопка **"Показать пасхалку"** активирует визуальные эффекты

---

7. Рекомендации

- Для больших датасетов рекомендуется использовать предобработку и визуализацию частями (head/фильтры), чтобы ускорить работу.
- Перед созданием предсказаний убедитесь, что все данные корректно предобработаны.
- Загружайте модели только из доверенных источников.
