

Описание к проекту

РАЗРАБОТКА АЛГОРИТМА МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ
ПОВЕДЕНИЯ КИБЕРФИЗИЧЕСКОЙ СИСТЕМЫ НА ОСНОВЕ ИСТОРИЧЕСКИХ ДАННЫХ

Выполнил:

Гавриленко Ольга Руслановна

Санкт – Петербург
2024

Оглавление

1.	Общие сведения о программе	3
1.1.	Frontend – разработка	3
1.2.	Backend – разработка	3
1.3.	Аналитическое исследование	4
2.	Техническая схема архитектуры решения и компонентов	5
3.	Выводы	6

1. Общие сведения о программе

Целью программы «Разработка алгоритма машинного обучения для прогнозирования поведения киберфизической системы на основе исторических данных» является прогнозирование давления на приеме насоса и оценивании точности результатов с помощью коэффициента детерминации. Данный проект создан для того, чтобы при заказе скважинных штанговых насосов, заказчик мог принять решение и быть в нем уверенным, основываясь на точности предсказания давления на приеме насоса.

Проект составляет 3 взаимосвязанные части:

- frontend – разработка – клиентская часть;
- backend – разработка – серверная часть;
- аналитическое исследование.

1.1. Frontend – разработка

HTML – форма представляет клиентскую часть. В ней заказчик вводит параметры для предсказания давления. Данная форма представляет из себя 7 строк, в которые вводятся параметры скважины, а также кнопка, по которой отправляется запрос на сервер. Форма сделана с помощью – HTML, а дизайн прописан с помощью – CSS.

1.2. Backend – разработка

Разработка внутренней части программы описана на языке Python с помощью Framework «Flask». Flask — это легковесный веб-фреймворк для языка Python, который предоставляет минимальный набор инструментов для создания веб-приложений. У Flask много преимуществ, которые выделяют его среди других фреймворков:

- простой синтаксис — Python;

—удобные шаблоны — можно быстро создавать прототипы веб-приложений;

—большое количество инструментов для гибкой настройки сайтов под любые нужды.

В проекте для разработки backend – части был выбран данный Framework, потому что модели машинного обучения реализованы на Python, а значит для удобства переноса моделей, полученных при моделировании будет удобно воспользоваться той же библиотекой.

Для отправки введенных данных из клиентской формы на сервер используется HTTP – протокол с запросом «POST».

Как упоминалось ранее, в серверной части также присутствуют модели прогнозирования. О них подробнее написано в пункте 1.3. данного документа и в документе «Математическое обоснование».

1.3. Аналитическое исследование

Аналитика в проекте представлена как моделирование прогнозов давления на выходе с помощью 3-х методов реализуемой в библиотеке sklearn:

—Линейная регрессия (самый простой регрессионный метод)

— Случайный лес (более современный регрессионный метод);

—Градиентный бустинг (также современный регрессионный метод, отличающийся по модели от случайного леса);

Моделирование реализуется с помощью машинного обучения (англ. machine learning, ML) — это класс методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а обучение за счет применения решений множества сходных задач. Для построения таких методов используются средства математической статистики, численных методов, математического анализа, методов оптимизации, теории

вероятностей, теории графов, различные техники работы с данными в цифровой форме.

2. Техническая схема архитектуры решения и компонентов

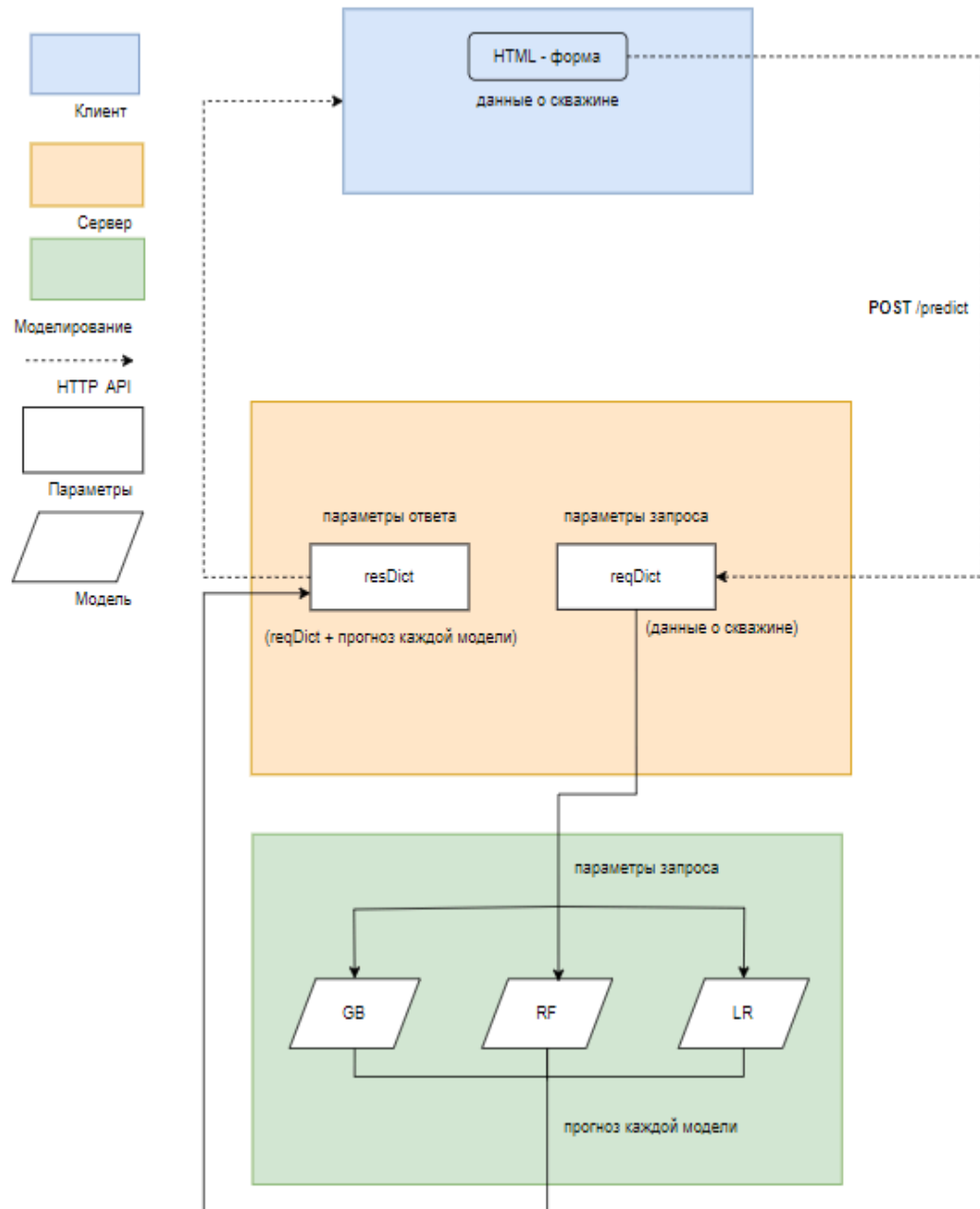


Рисунок 1 – Диаграмма описания технической части проекта

3. Выводы

В данной работе проведено исследование, которое показывает, что в открытых данных присутствуют выбросы, влияющие на результаты прогнозирования. Модели дают наилучшие с точки зрения точности результаты при значении давления от 50 и меньше. В открытом доступе нет дополнительных данных для построения более точного прогноза.

Из данной работы, можно сделать вывод, что наиболее точным методом прогнозирования является градиентный бустинг по оценке коэффициента детерминации.