

Проектная работа

Ознакомьтесь с проектом, который нужно выполнить в конце курса.

Описание проекта

Оператор мобильной связи «Мегалайн» выяснил: многие клиенты пользуются архивными тарифами. Они хотят построить систему, способную проанализировать поведение клиентов и предложить пользователям новый тариф: «Смарт» или «Ультра».

В вашем распоряжении данные о поведении клиентов, которые уже перешли на эти тарифы (из проекта курса «Статистический анализ данных»). Нужно построить модель для задачи классификации, которая выберет подходящий тариф. Предобработка данных не понадобится — вы её уже сделали.

Постройте модель с максимально большим значением *assurasy*. Чтобы сдать проект успешно, нужно довести долю правильных ответов по крайней мере до 0.75. Проверьте *assurasy* на тестовой выборке самостоятельно.

Инструкция по выполнению проекта

1. Откройте файл с данными и изучите его. Путь к файлу: `datasets/users_behavior.csv`. [Скачать датасет](#)
2. Разделите исходные данные на обучающую, валидационную и тестовую выборки.
3. Исследуйте качество разных моделей, меняя гиперпараметры. Кратко напишите выводы исследования.
4. Проверьте качество модели на тестовой выборке.
5. Дополнительное задание: проверьте модели на вменяемость. Ничего страшного, если не получится: эти данные сложнее тех, с которыми вы работали раньше. В следующем курсе подробнее об этом расскажем.

Описание данных

Каждый объект в наборе данных — это информация о поведении одного пользователя за месяц. Известно:

- *calls* — количество звонков,
- *minutes* — суммарная длительность звонков в минутах,
- *messages* — количество sms-сообщений,
- *mb_used* — израсходованный интернет-трафик в Мб,
- *is_ultra* — каким тарифом пользовался в течение месяца («Ультра» — 1, «Смарт» — 0).

Как будут проверять мой проект?

Мы подготовили критерии оценки проекта, которыми руководствуются ревьюеры. Прежде чем приступить к решению кейса, внимательно их изучите.

На что обращают внимание ревьюеры, проверяя проект:

- Как вы изучаете данные после загрузки?

- Корректно ли разделяете данные на выборки?
- Как выбираете размеры выборок?
- Правильно ли вы оцениваете качество моделей в исследовании?
- Какие модели и гиперпараметры вы используете?
- Какие выводы об исследовании делаете?
- Правильно ли тестируете модели?
- Насколько высокое значение *accuracy* получаете?
- Соблюдаете структуру проекта и поддерживаете аккуратность кода?