

The background is dark gray with a pattern of green hexagons and thin white lines. Some hexagons are solid green, while others are outlines. The lines form a network-like structure connecting the hexagons.

SKILLFACTORY

Вводный вебинар по проекту

Компьютер говорит НЕТ

Нохрина Анна

ФинТех разработчик CodeRocket

Вебинар по проекту “Компьютер говорит НЕТ”

Давайте знакомиться

Нохрина Анна

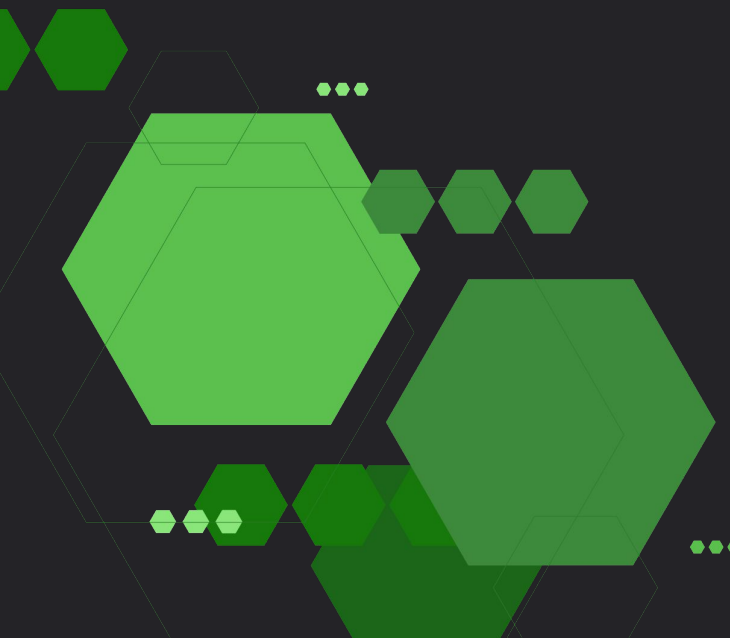
Ментор проекта Компьютер говорит Нет



Вместе мы:

- формируем совместный план по решению задачи
- работаем на протяжении всего проекта
- активно задаем вопросы в [#0_project_4-компьютер_говорит_нет](#)
- если не успеваете сдать проект, пишем в слак [@Анна Нохрина](#)
- подводим итоги проделанной работы

План вебинара

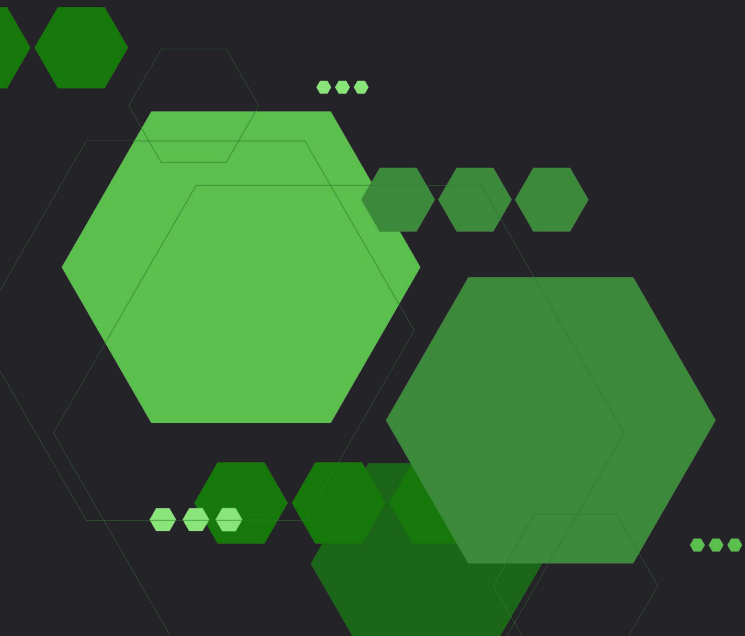
- 
1. Постановка задачи
 2. Обзор навыков для решения задачи
 3. Составление плана работы над проектом
 4. Сложности при выполнении проекта
 5. Рекомендации к выполнению
 6. Вопросы

Постановка задачи

Задача: Предсказать вероятность дефолта клиента банка

Цель: Построить модель на имеющихся данных, которая сумела бы решить нашу задачу

Данные: Имеется датасет о клиентах банка с информацией о КИ и анкетных данных клиентов банка



Что необходимо для решения задачи?

Какие навыки у нас уже есть?

- уверенно владеем основами python, pandas, numpy...
- умеем проводить качественный разведывательный анализ
- есть опыт участия на kaggle
- построили свою первую модель и можем построить еще!

Чему еще необходимо научиться?

- изучить новый алгоритм логистической регрессии и научиться с ним работать
- изучить разные метрики оценки модели и найти самую подходящую
- попробовать подобрать гиперпараметры к модели для ее улучшения

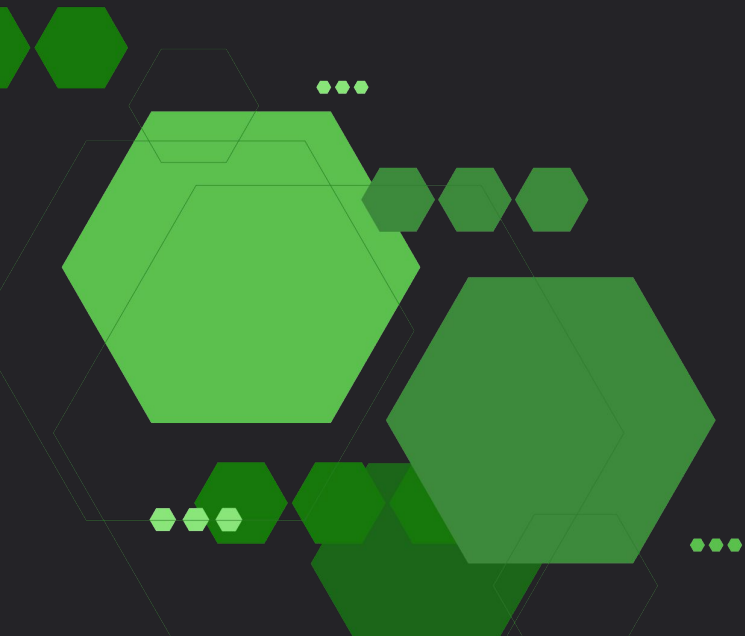
Вебинар по проекту “Компьютер говорит НЕТ”

Зачем мне решать этот проект?

- финтех — одна из передовых сфер, где повсюду применяется ml, dl
- четкое понимание того, где деньги и как их посчитать
- проект про жизнь
- понять, что нейронные сети не везде хороши

Как будем решать проект?

1. первичный осмотр
2. визуализация данных, поиск выбросов, пропусков,
3. базовая модель



Как будем решать проект?

1. Посмотреть на данные - визуализация
2. Чистим данные, ЕДА, ищем выбросы, (визуализация, iqr), описание датасета. Пустые значения убрать, + продвинутые методы замены пропусков -> датасет очищенный
3. наивная модель,
4. анализ по типу признаков
5. генерация признаков
6. строим модель
7. выбор метрик, строим много метрик
8. нормализация, стандартизация
9. важность и отбор признаков

Полезное

- ЕДА - это просто, когда используешь Pandas Profiling!
https://github.com/testdriver87/sf_rds_projects/tree/master/rds_06
- Как пользоваться Pandas Profiling?
<https://habr.com/ru/company/ruvds/blog/451478/>
- Пишем красивый и понятный код вместе с PEP-8
<https://pythonworld.ru/osnowy/pep-8-rukovodstvo-po-napisaniyu-koda-na-python.html>
- Как понять какой признак важен sklearn feature_selection?
https://scikit-learn.org/stable/modules/classes.html#module-sklearn.feature_selection
- Оформляем репо правильно
https://github.com/yurybikuzin/skillfactory_rds4_recommendation_challenge

Сложности при выполнении проекта

решение		проблема	
Сделать основную часть работы, смотреть ноутбуки коллег, работать в команде	✓	✗	Не хватает времени
Пересмотреть видео с работой на каггл, спросить в канале	✓	✗	Сложности при сабмите и создании ноутбука на каггле
Посмотреть другие метрики, вернуться к ЕДА	✓	✗	Не растет метрика чтобы я ни делал
Закреп в слаке, спросить у ментора, оставить задачу	✓	✗	Сложности при подборе гиперпараметров
Сменить команду, делать одному, написать тьютору	✓	✗	Вялая команда

Критерии оценивания

- Качество кода (соблюдение стандартов оформления пер-8, комментирование кода, наличие выводов). Оформление проекта на Github— 3 балла.
- Качество разведывательного анализа данных (визуализация, очистка данных, работа с выбросами) — 3 балла.
- Работа с признаками (выбор признаков, генерация новых признаков) — 3 балла.
- Качество решения: Результат метрики ROC AUC, f1, accuracy, matrix confusion — 3 балла.
- Качество решения: обоснованность выводов и общая логика решения — 3 балла.
- *Работа в команде — 3 балла.
- *Настройка гиперпараметров — 3 балла.
- Работа сдана к дедлайну — 3 балла.

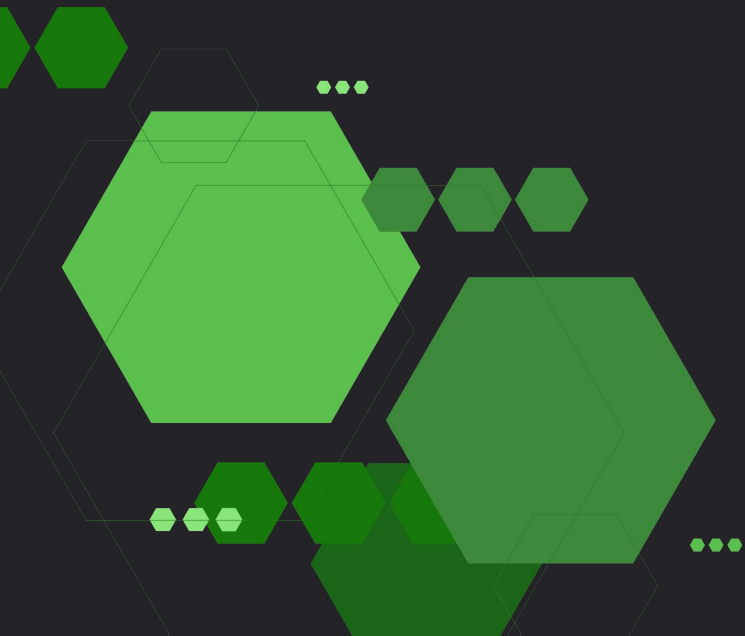
За каждый критерий можно получить от 0 до 3 баллов.

Вебинар по проекту “Компьютер говорит НЕТ”

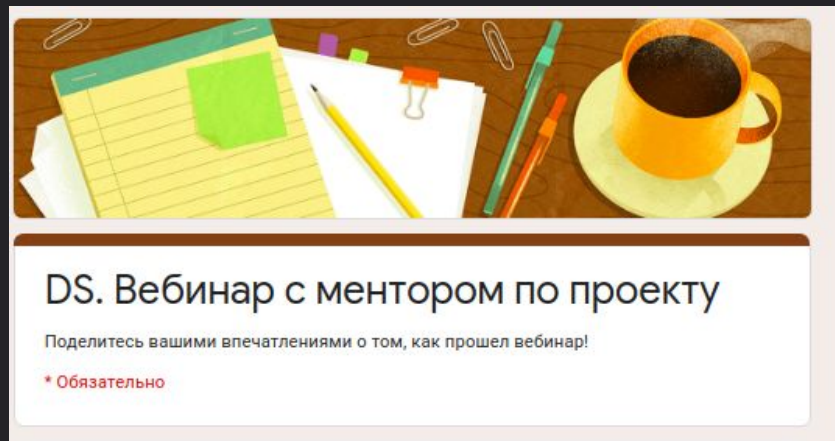
Для любопытных

- <https://catboost.ai/>
- <https://xgboost.readthedocs.io/en/latest/>
- <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>
- https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegressionCV.html

Вопросы



Поделитесь вашими впечатлениями о том, как прошел вебинар!



[Нажмите, чтобы оставить ОС ментору](#)