

SKILLFACTORY

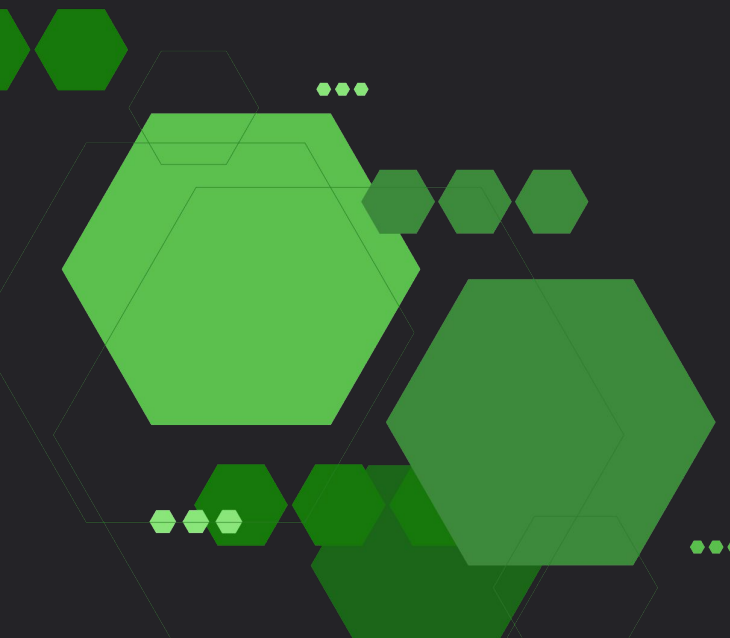


Итоговый вебинар по проекту **Компьютер говорит НЕТ**

Нохрина Анна

ФинТех разработчик CodeRocket

План вебинара

- 
1. Вспомним проект и базовое решение
 2. Обсудим самые распространенные ошибки
 3. Обсудим распространенные сложности
 4. Посмотрим хороший код
 5. Научимся читать чужой код
 6. Соберем идеальный проект
 7. Подведем итоги модуля
 8. Вопросы

Вебинар по проекту “Компьютер говорит НЕТ”

Кому будет полезен этот созвон?

DST-35	s.cheremisov@gmail.com	s_cheremisov	19.03.2021	Активный пользо	1. Cheremisov Stanislav	14.03.2021 17:13:28
DST-35	agzhuravleva@gmail.com	Anastasia_Z	14.03.2021	Активный пользо	1. Anastasia\nhttps://www	14.03.2021 10:22:14
DST-35	yasckevich@gmail.com	Yasckevich	19.03.2021	Активный пользо	1. yasckevich\n2. 0.7360	11.03.2021 19:05:45
DST-35	dotsenko_double@mail.ru	AnnaDo	19.03.2021	Активный пользо	Название команды: А т	11.03.2021 2:45:25
DST-35	alex160798@gmail.com	alex160798	15.03.2021	Активный пользо	Никнейм - Alexey Kobel	10.03.2021 18:42:18
DST-35	l.oborina@ar-manageme	Lidiya_2-2	10.03.2021	Под угрозой	rds-scoring_team-a_ty_t	10.03.2021 11:29:40
DST-35	ek.mezentseva@yandex	EkaterinaMezen	18.03.2021	Активный пользо	1. никнейм на Kaggle - E	09.03.2021 14:13:42
DST-34	yulya_palvan@bk.ru	palvanovayg	18.03.2021	Активный пользо	1. ygpalvanova@gmail.ru	06.03.2021 21:44:00
DST-34	SAMakarov@beeline.ru	smakarov	19.03.2021	Активный пользо	1. https://www.kaggle.co	06.03.2021 19:19:21
DST-34	ania.tok@yandex.ru	Anna_Tokareff	17.03.2021	Активный пользо	Мой никнейм на Kaggle	03.03.2021 22:26:09
DST-36	ooochen_zlaya@mail.ru	Daniil_Gruzdev	17.03.2021	Активный пользо	1. Daniil Gruzdev\n2. 0.7	03.03.2021 20:17:22
DST-33	123julia88@li.ru	123Julia88	19.03.2021	Активный пользо	https://www.kaggle.com/j	03.03.2021 9:56:49
DST-33	zhakupov1987@gmail.com	Aydar	14.03.2021	Активный пользо	1. Aydar Zhakupov\n2. 0	28.02.2021 17:19:34
DST-34	maria.ivoilowa@yandex	Marusha	08.03.2021	Под угрозой	https://github.com/Marus	28.02.2021 16:47:08

И всем, кто почти закончил проект!

Вебинар по проекту “Компьютер говорит НЕТ”

ФИДБЕКИ

в большинстве редми полностью отсутствует или некачественное

https://github.com/anna-kostyakova/SkillFactory_Projects/tree/main/module%204

Работа “куски кода”

Отличие дс от программирования - анализ данных и выводы, которые полезно сохранять для других коллег и для себя самого

Сопровождение комментариями кода в анализе данных. Это графики и выводы под графиками. Что мы увидели? Хорошо это или плохо? Что мы можем этим сделать сейчас или позже.

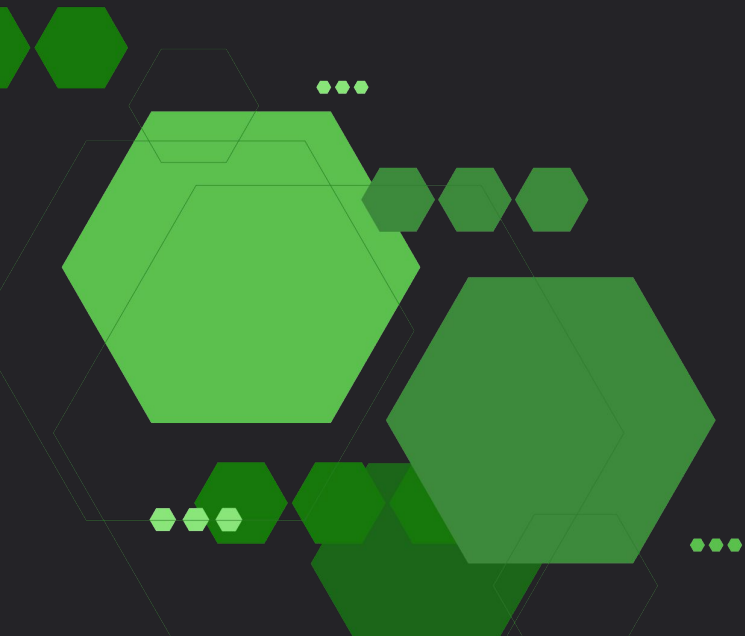
Процесс создания новых переменных или выбора тоже сопровождается мыслями.

Подведение итогов в каждой логической части проекта, где мы что-то получаем (первый взгляд на данные, обработанный датасет, наивную модель..)

А так же сопровождение выводами процесс обучения и итоговый результат

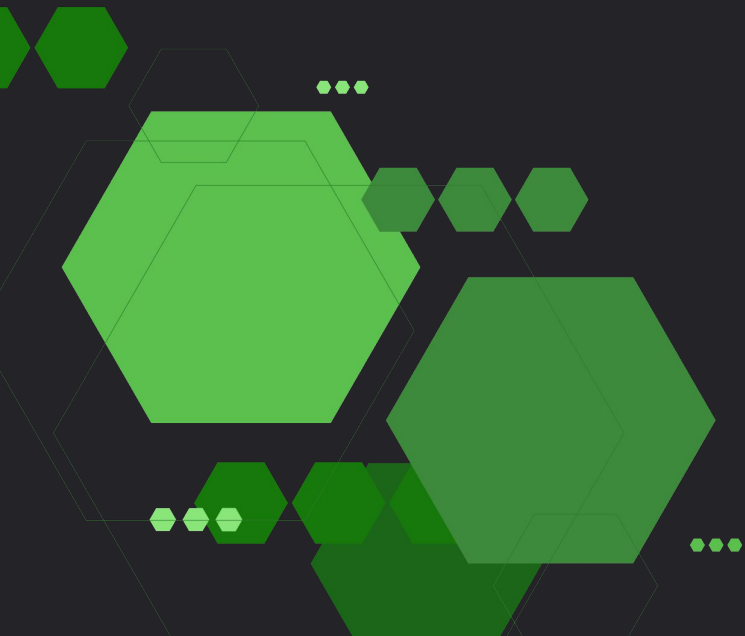
Анализ данных проводится 1 раз

После обучения модели полезно возвращаться к самому началу и проводить эксперименты с разными вариациями замены пропусков, отбора признаков, анализом и сравнивать результаты



После построения модели проект резко заканчивается

Проводить постанализ очень полезно.
Построение матрицы ошибок и выводы о
результате обучения модели помогают нам
лучше понять ее работу.
Так мы видим к чему еще можно стремиться

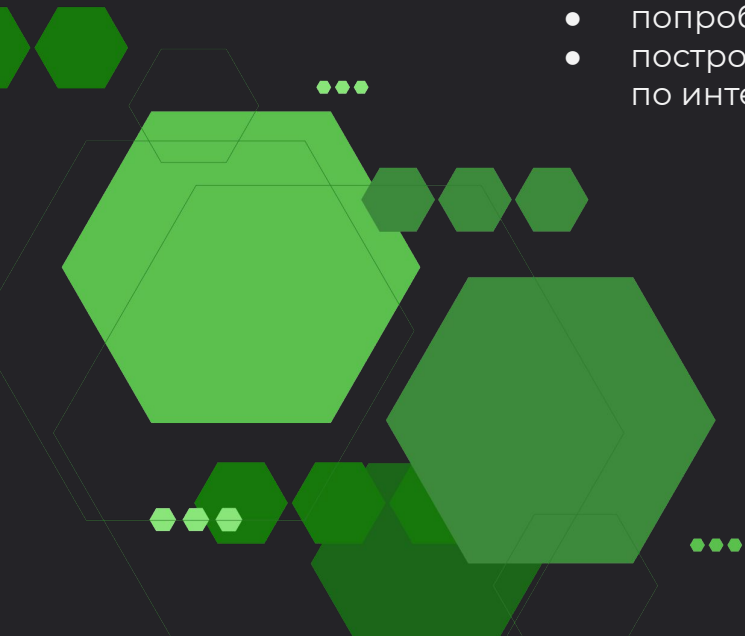


“Трудные моменты”

- не растёт метрика, несостоятельность метрики
- непонятно как оценить модель
- баланс между ошибкой первого и второго рода
- долгий подбор гиперпараметров hyperopt, optuna

Как можно улучшить проект?

- построить матрицу ошибок и совокупность метрик
- провести много экспериментов
- попробовать бороться с несбалансированностью классов?
- попробовать зануление весов при помощи регуляризации
- построить несколько моделей машинного обучения и оценить их по интерпретируемости, метрике, сложности и скорости



Идеальный проект

1. Первичный анализ данных при помощи pandas-profiling
2. Построение наивной модели
3. построение нескольких метрик
4. Борьба с неран. классами
5. Анализ-сравнение нескольких алгоритмов*
6. Оформление качественного редми на гихабе

Подведем итоги

1. Логрегрессия подходит для задач бинарной классификации
2. Логрегрессия возвращает вероятности
3. Корреляция, гаусово распределение, шум

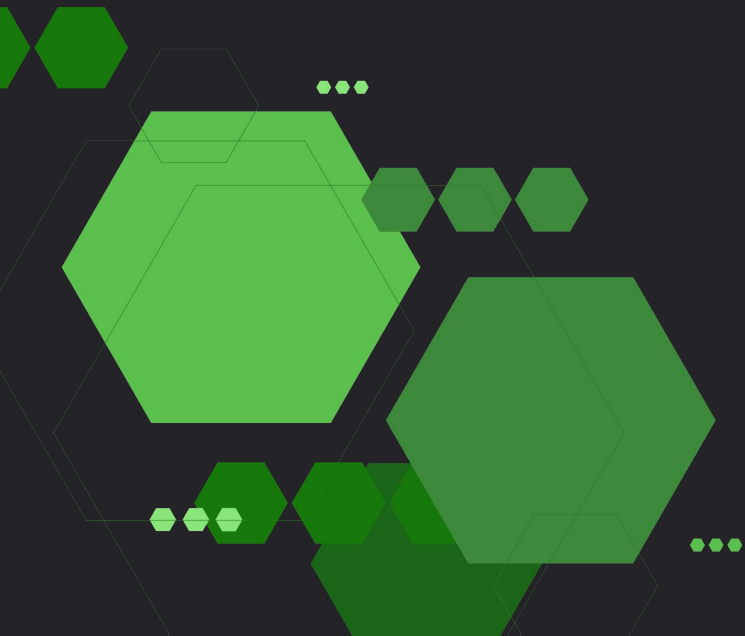
Сэмплинг:

<https://machinelearningmastery.com/random-oversampling-and-undersampling-for-imbalanced-classification/>

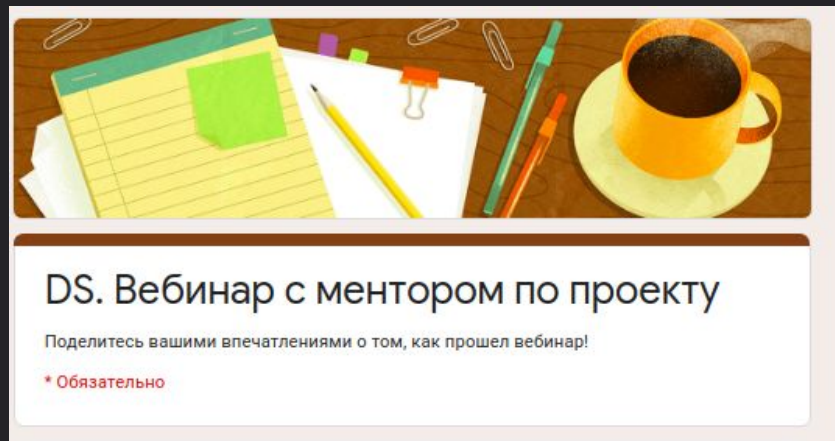
<https://towardsdatascience.com/oversampling-and-undersampling-5e2bbaf56dcf>

<https://www.kaggle.com/residentmario/undersampling-and-oversampling-imbalanced-data>

Вопросы



Поделитесь вашими впечатлениями о том, как прошел вебинар!



[Нажмите, чтобы оставить ОС ментору](#)