

Этапы выполнения работы в Colab

1. Описание набора данных, переменных

- (определение целевой переменной, которую нужно предсказать)
- Какие алгоритмы используем в предсказании?
- Регрессия
- Классификация
- Кластеризация

2. Разведочный анализ и подготовка данных (EDA)

- есть ли в данных пропущенные значения (что делаем? заполняем, удаляем? Описываем почему и как)
- смотрим статистическую информацию о числовых переменных (так же анализируем и
- описываем действия или бездействия)
- Для категориальных переменных смотрим значение переменной (анализируем,
- преобразовываем, группируем и т. д.)
- строим всевозможные графические элементы (гистограммы, линейные зависимости,
- ящики с усами и т. д.), которые дадут представление о данных.

Не забываем о целевой переменной- мы должны увидеть какие данные влияют в нашем наборе на target.

Не забываем описывать графические построения.

3. Очистка данных

- На основании проведенного выше анализа занимаемся при необходимости:
 - Пропущенными значениями
 - Удалением индексов
 - Корректировкой выбросов
 - Несбалансированностью классов
- преобразовываем все категориальные переменные в числовые и рисуем тепловую
- карту между всеми переменными

Так же анализируем и при необходимости корректируем данные

4. Моделирование

На основании целевой выбираем алгоритмы обучения

Необходимо выполнить обучение на следующих библиотеках:

1. Scikit-learn случайный лес и градиентый бустинг

2. На выбор одну библиотеку из трех:

— AutoGluon,

— PyCaret,

— h2o-3 (веб интерфейс Flow, localhost 54321, сделать скриншоты в Word).

3. Сравнить результаты, полученные при обучении моделей

После обучения модели желательно визуализировать результаты

Описывайте код, так легче найти ошибки и исправить