
Проект 5

НИКИТА КОТЕЛЕВСКИЙ

О чем поговорим сегодня

1. О чем был проект?

2. Общие ошибки

3. Разбор примера

4. Теоретическая справка

5. Следующий модуль.

6. Вопросы

О чем был проект?

Цель проекта — по имеющимся данным, построить модель, предсказывающую дефолт клиента банка.

Основная оценка складывалась за счет:

1. EDA
 2. Корректность подготовки данных для обучения
 3. Обучение модели
-

Общие ошибки

- 1. Нормализация разных сплитов данных (тренировочного, тестового и валидационного)
 - 2. Борьба с несбалансированностью данных (upsampling/downsampling, взвешенная логрессия)
 - 3. Data leakage
 - 4. Осмысленность сетки гиперпараметров для GridSearchCV
-

Теоретическая справка

- Сегодня обсудим следующие теоретические вопросы:
 - - MLE подход
 - - MAP подход (Регуляризация)
 - - Mutual Information (KL дивергенция)
-

Давайте разберем решения!

Немного рекомендаций

Следующий модуль очень важный — основные алгоритмы машинного обучения.

Будем разбирать:

- 1. Древесные модели**
 - 2. Бустинг/стекинг**
 - 3. Кластеризация**
 - 4. Отладка Python**
 - 5. Парсинг**
 - 6. Проект — kaggle соревнование**
-

Вопросы!
