

Bank Marketing research

Артём Уткин, Кирщин Иван

November 2023

Содержание

1	Description	1
2	Libraries	1
3	Processing and Data exploration	1
4	Modeling	1
5	Results	1
6	Conclusion	1

1 Description

Данные получены португальским банком при массовом обзвоне клиентов. Цель классификации - предсказать, подпишется ли клиент на срочный депозит (переменная y). Bank Marketing.

2 Libraries

3 Processing and Data exploration

Начнём работу непосредственно с данными. Структурно это будет выглядеть так: последовательно обрабатываем каждый имеющийся признак(колонку), сначала проверяем на незаполненные поля. Если такие имеются, то обрабатываем или не учитываем в выборке. Затем обрабатываем выбросы: либо преобразуем их, либо не учитываем. Если значения имеют вид 'yes', 'no', то заменяем их на бинарные значения $'yes' \rightarrow '1', 'no' \rightarrow '0'$ (признаки: 'default', 'housing', 'loan', целевая переменная y). Если много текстовых значений или значений, которые непонятно как достойно обработать (признак категориальный), будем использовать методы Mean Target Encoding (признаки: 'day of week', 'month') или One-hot Encoding (признаки: 'job', 'contact', 'outcome'). Посмотрим на распределение вида признак-среднее значение целевой переменной, попытаемся вывести новый признак, при котором указанное выше распределение будет напоминать линейную зависимость, применили к признакам 'age', 'campaign'. Присутствует визуализация данных для каждого из признаков.

4 Modeling

5 Results

6 Conclusion