

Uplift modeling

13 September 2021

Moscow

Необходимо немного исследовать данные и построить Uplift-модель (или желательно несколько) по визитам на предоставленных данных. Рассмотреть метрики для этого типа задач. По ним объяснить какую из моделей лучше использовать и почему, если одна модель почему эту модель нужно использовать.

Данные по рекламным компаниям на клиентах.

1. minethatdata_womens_train.csv - файл с обучающей выборкой

Описание клиента до рассылки (контекст):

1. Index - id пользователя
2. Recency - кол-во месяцев с последней покупки
3. History_segment - категория (по кол-ву \$ потраченных за последний год)
4. History - кол-во \$ потраченных за последний год
5. Mens - индикатор, 1 - клиент купил товары для мужчин за последний год
6. Womens - индикатор, 1 - клиент купил товары для женщин за последний год
7. Zip_code - региональная категория индексов. Urban, Suburban, Rural
8. Newbie - индикатор, новый клиент пришел в магазин в первый раз за последние 12 месяцев
9. Channel - описывает способы приобретения клиентом товаров за последний год
10. Segment - mens e-mail / women's e-mail / no e-mail (то к какой категории отнесен клиент)

Переменные описывающие клиента в течение 2 недель **после** получения e-mail:

1. Visit - индикатор, 1 - клиент посетил сайт в последующие 2 недели после рассылки (то для чего нужно построить uplift модель)
2. Conversion - индикатор, 1 - клиент купил товары на сайте в последующие 2 недели после рассылки
3. Spend - кол-во \$ потраченных на покупки в последующие 2 недели

Термины для гугла:

Uplift modeling, casual inference