



## Задача «Модель прогнозирования выбора образовательных курсов пользователями»

### Введение

В данный момент продвижение онлайн-курсов Нетологии осуществляется без применения моделей машинного обучения. Следующий этап развития продвижения курсов - попытка попасть в индивидуальные пожелания клиентов, улучшив клиентский опыт и увеличив выручку. Для этого был выбран подход NBO (Next Best Offer): перед рассылкой коммуникаций необходимо понимать, какой из доступных курсов максимизирует вероятность отклика покупателя. В этом и должна помочь модель, подбирается наиболее актуальный для клиента курс.

### Условие задачи

Цель соревнования — разработать модель машинного обучения, которая на основе истории действий пользователя на сайте будет предсказывать его повторную покупку образовательного курса.

### Описание входных значений

- train.csv — файл для обучения, содержащий 2000 строчек;
- test.csv — файл для предсказания;
- sample\_solution.csv — пример файла для отправки.

А также в данных присутствует 59 уникальных столбцов.

id	уникальное значение каждой строки
month_id	последний день месяца среза
student_id	id студента
carts_created_at	дата создания заявки
program_id	id продукта
<b>target</b>	<b>строка для предсказания</b>



Метрики прохождения программы в месяц среза:

spent_time_total	время за учебой в условных единицах
spent_time_to_complete_hw	время, потраченное на выполнение дз
avg_hw_mark	средняя оценка за дз

completed_hw	кол-во выполненных дз
failed_hw	кол-во проваленных дз
reworked_hw	кол-во переделанных дз
interacted_hw	кол-во дз, которыми занимался студент в этом месяце

test_with_good_mark	кол-во тестов с удовлетворительной оценкой
test_with_great_mark	кол-во тестов с отличной оценкой

webinars	кол-во просмотренных вебинаров
avg_quiz_result	средняя оценка за квизы
notes	кол-во оставленных заметок
hw_leader	был лидером в групповых заданиях
lessons	кол-во уроков, которые открывал пользователь
activity	общая активность студента в личном кабинете

Покупки до приобретения текущей программы:

bought_d1	кол-во прошлых покупок в направлении 1
bought_d2	кол-во прошлых покупок в направлении 2
bought_d3	кол-во прошлых покупок в направлении 3
bought_d4	кол-во прошлых покупок в направлении 4
bought_d5	кол-во прошлых покупок в направлении 5

Метрики относящиеся непосредственно к покупке:

payment_type	тип оплаты
promo	участвовал ли в любой промокампании
price	цена программы
communication_type	тип создания заявки (ОП или сам через сайт)
auto_payment	автопокупка (1 или 0)



ABC	сегмент клиента по готовности купить
city	город пользователя
country	страна пользователя
gender	пол
age_indicator	возраст в условных единицах
speed_recall	скорость перезвона клиенту (1 - быстро, 7 - долго)

os	операционная система пользователя
browser	название браузера
platform	платформа

Информация по коммуникации с менеджерами отдела продаж:

m_avg_duration	средняя длительность звонка
m_avg_talk_duration	средняя длительность разговора
m_missed_calls	кол-во пропущенных звонков
m_total_calls	всего звонков
m_was_conversations	кол-во звонков, когда состоялся диалог
m_total_duration	длительность всего

Информация по коммуникации с телемаркетологами:

p_avg_duration	средняя длительность звонка
p_avg_talk_duration	средняя длительность разговора
p_missed_calls	кол-во пропущенных звонков
p_total_calls	всего звонков
p_was_conversations	звонков, когда состоялся диалог
p_total_duration	длительность всего

support\_feedback\_avg  
средняя удовлетворенность службой поддержки за год

Средняя оценка уроков студентов в каждом направлении за прошедший год:

feedback_avg_d1	средний балл по курсам направления 1
feedback_avg_d2	средний балл по курсам направления 2



feedback\_avg\_d3

feedback\_avg\_d4

feedback\_avg\_d5

средний балл по курсам направления 3

средний балл по курсам направления 4

средний балл по курсам направления 5

## Метрика

В качестве метрики выступает взвешенная сумма Recall и Precision по следующей формуле:

$$Result = 0.2 * Recall + 0.8 * Precision$$