

Презентация итоговой работы

“Сервис предсказания совершения
целевого действия пользователем на сайте
“СберАвтоподписка” ”

Выполнил: Валиев Р.И.
специализация Machine
Learning Engineer

Цели и задачи

Цели

- Научиться предсказывать совершение целевого действия (ROC-AUC ~ 0.65)
- Внедрить модель машинного обучения в сервис предсказания целевого действия

Задачи:

- Изучение датасета
- Подготовка данных
- Создание фич
- Обучение моделей
- Внедрение модели в сервис(localhost web app)

Изучение датасета

GA Sessions

- Одна строка = один визит на сайт
- Содержит атрибуты на основании которых обучается модель/делаются предсказания

GA Hits

- Одна строка = одно событие в рамках одного визита на сайт
- Event_action – атрибут содержащий целевые данные о конверсии

Далее, в целях обучения модели с “учителем” по атрибуту **session_id** объединяем исследуемые таблицы

Подготовка данных

Cr = 0

- 97% датасета
- Строки с пропусками данных удаляем

Cr = 1

- 3% датасета
- В связи с ценностью строки с пропусками заполняем модой по атрибуту(по cr1)

Создание фич

Генерация ФИЧ МЕСЯЦ / ЧАС посещения

- Из данных о дате и времени сессии создаем фичи “ЧАС” и “Месяц” предполагая что время суток и года может влиять на CR

Генерация ФИЧИ разрешение экрана

- В целях создания количественной из категориальной переменной, для дальнейшего ONE, перемножили разрешение экрана

Генерация ФИЧИ lat long

- С помощью модуля геору из названия города и страны получаем координаты lat long, для дальнейшего ONE

Обучение моделей

Логистическая регрессия

- LogisticRegression ROC-AUC 0,74%

Модель случайного леса

- RandomForestClassifier ROC-AUC 0,68%

Многослойный персептрон

- MLPClassifier ROC-AUC 0,73%

В соответствии с наилучшим показателем ROC-AUC 0,74 для дальнейшего внедрения выбрана модель логистической регрессии

Внедрение модели

С помощью фреймворка FAST API реализована интеграция модели предсказания в локальный сервис. Ниже продемонстрирован пример работы API:

POST http://127.0.0.1:8000/predict

Params Authorization Headers (9) Body Pre-request Script Tests Settings

none form-data x-www-form-urlencoded raw binary GraphQL JSON

```
1 {
2   "session_id": "9055768928391199041.1623652678.1623652678",
3   "client_id": "2108460508.1623652673",
4   "visit_date": "2021-06-14",
5   "visit_time": "09:00:00",
6   "visit_number": 1,
7   "utm_source": "ZpYIoDJMcFzVoPFsHGJL",
8   "utm_medium": "banner",
9   "utm_campaign": "LEoPHuyFvzoNfnzGgfc",
10  "utm_adcontent": "vCImpaGBnIQhyYnKXqp",
11  "utm_keyword": "puhZPIYqKXeFPaUviSjo",
12  "device_category": "mobile",
13  "device_os": "Android",
14  "device_brand": "Huawei",
15  "device_model": "NaN",
16  "device_screen_resolution": "360x720",
```

Body Cookies Headers (4) Test Results

Pretty Raw Preview Visualize JSON

```
1 {
2   "session_id": "9055768928391199041.1623652678.1623652678",
3   "client_id": "2108460508.1623652673",
4   "target": 1.0
```

POST http://127.0.0.1:8000/predict

Params Authorization Headers (9) Body Pre-request Script Tests Settings

none form-data x-www-form-urlencoded raw binary GraphQL JSON

```
1 {
2   "session_id": "9055401700113249881.1639446112.1639446112",
3   "client_id": "2108375006.163945",
4   "visit_date": "2021-12-14",
5   "visit_time": "04:41:52",
6   "visit_number": 1,
7   "utm_source": "ZpYIoDJMcFzVoPFsHGJL",
8   "utm_medium": "banner",
9   "utm_campaign": "TmThBvoCcwkcZZUWACYq",
10  "utm_adcontent": "NaN",
11  "utm_keyword": "NaN",
12  "device_category": "mobile",
13  "device_os": "Android",
14  "device_brand": "Samsung",
15  "device_model": "NaN",
```

Body Cookies Headers (4) Test Results

Pretty Raw Preview Visualize JSON

```
1 {
2   "session_id": "9055401700113249881.1639446112.1639446112",
3   "client_id": "2108375006.163945",
4   "target": 0.0
```