Курсовой проект

Мегафон.
Вероятность подключения услуги

Табунов Миша
07.06.2020

Задача

У нас появился запрос из отдела продаж и маркетинга. Как вы знаете «МегаФон» предлагает обширный набор различных услуг своим абонентам. При этом разным пользователям интересны разные услуги. Поэтому необходимо построить алгоритм, который для каждой пары пользователь-услуга определит вероятность подключения услуги.

Данные

В качестве исходных данных вам будет доступна информация об отклике абонентов на предложение подключения одной из услуг. Каждому пользователю может быть сделано несколько предложений в разное время, каждое из которых он может или принять, или отклонить.

Отдельным набором данных будет являться нормализованный анонимизированный набор признаков, характеризующий профиль потребления абонента. Эти данные привязаны к определенному времени, поскольку профиль абонента может меняться с течением времени.

Данные train и test разбиты по периодам – на train доступно 4 месяцев, а на test отложен последующий месяц.

Предобработка

Обработка основном даты и была попытка “подшумить данные” : data['diff1_l']

(np.log(data.iloc[1:,3].values))

Признаки отбирал на основе их IV (информационной важности)

Признаки с 10 уникальными значения перевел в категорийные. Остальные ушли в нумерические. Признаки потом трансформировал.

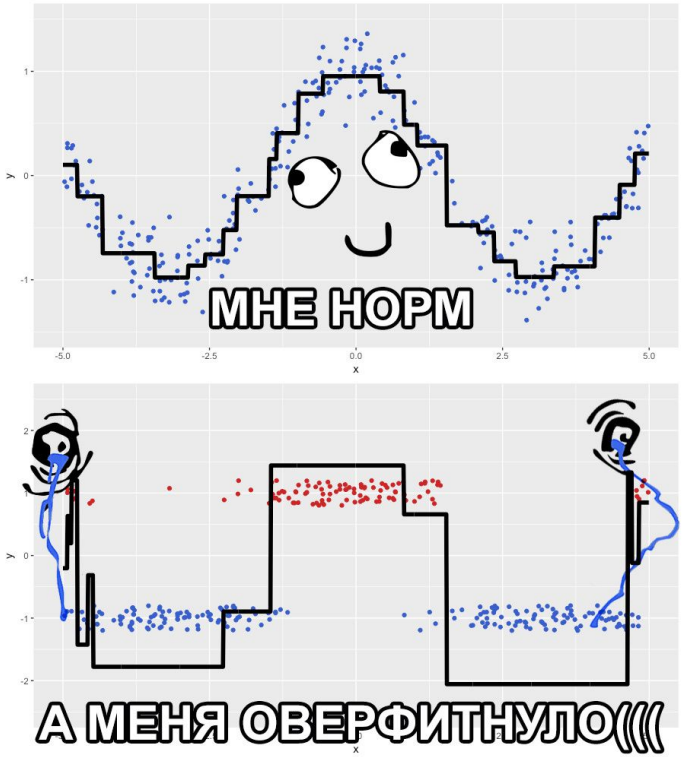
Модели

Обкатывал три модели:

1. LogisticRegression (большие ошибки)

2. XGBoostClassifer (“0” находит отлично, а вот “1” пропускает)

3. BernoulliNB (“1” находит отлично, а вот “1” пропускает)



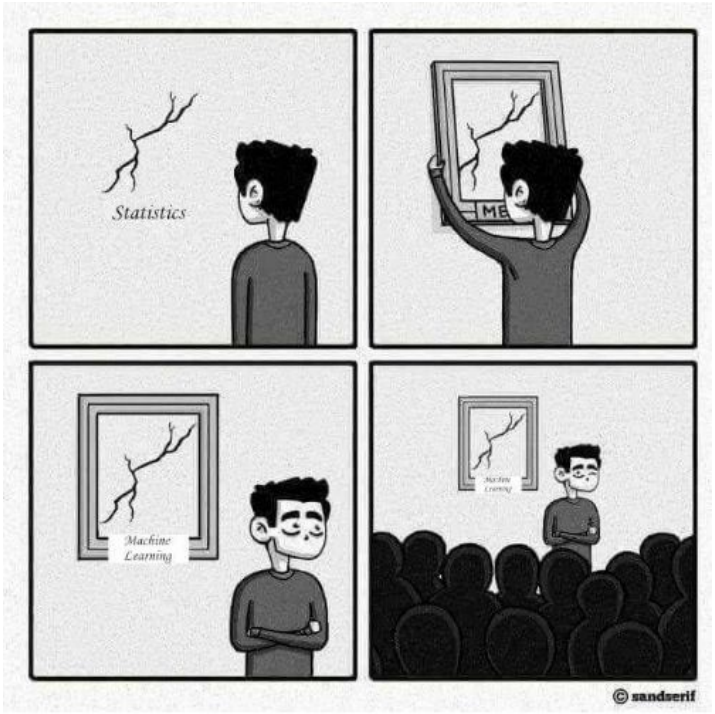
Оценка результатов

Работу модели оценивал по матрице смешений.

Модель подгонял к наибольшей правильности TP/TN

CONFUSION MATRIX

col_0	0.0	1.0
target		
0.0	1026	15
1.0	73	5



Machine Learning

Data Analyst

What my friends think I do

What my Mom thinks I do

What my boss thinks I do

What my customers think I do

What I think I do

What I really do