

Описание тестового задания для Data Scientist

Описание задания

Для рекламной сети важно уметь точно предсказывать вероятность совершения пользователем конверсии (conversion) на каждый показ (impression) определенной рекламного банера.

По предложенной выборке данных необходимо произвести оценку значимости различных факторов для решения данной задачи

Выборка данных

Файл с данными содержит случайную выборку работы рекламной сети с примерно 100 сайтами за две недели.

- Каждая строчка представляет собой данные метрик conversions и impressions агрегированные по остальным атрибутам.

Field	Description
impressions	<ul style="list-style-type: none">• Number of impressions made by users with features below
conversions	<ul style="list-style-type: none">• Number of conversions made by users after seeing advertisement on impressions
date_time	<ul style="list-style-type: none">• Timestamp of an event rounded by hour
zone_id	<ul style="list-style-type: none">• ID of a zone• Zone - place on a page of a website where ad will be shown
affiliate_id	<ul style="list-style-type: none">• ID of affiliate• Affiliate - website where ad will be shown
banner_id	<ul style="list-style-type: none">• ID of a banner• Banner - particular ad to be shown
advertiser_id	<ul style="list-style-type: none">• ID of an Advertiser
offer_id	<ul style="list-style-type: none">• ID of particular advertisement offer• Several advertisers can share the same offer
geo	<ul style="list-style-type: none">• Country Code• Values: IN, IT, US, etc.
region_code	<ul style="list-style-type: none">• Code of a region of a country
city_code	<ul style="list-style-type: none">• Code of a city
os_type	<ul style="list-style-type: none">• Operation System Type• Values: desktop, mobile, other, -
os	<ul style="list-style-type: none">• Operation System• Values: windows, android, ios, mac, other, linux, windows_phone, blackberry, symbian, -
os_version	<ul style="list-style-type: none">• Operation System Version• Values: win7, win10, android4, android5, ios9, etc.
device_type	<ul style="list-style-type: none">• Type of user device• Values: desktop, phone, tablet, tv, other, -
device	<ul style="list-style-type: none">• User device• Values: desktop, other, iphone, ipad, playstation, xbox, ipod, kindle, -
browser	<ul style="list-style-type: none">• User browser• Values: chrome, safari, firefox, ie, fb, native, edge, opera, uc, other, opera_mini, yandex, puffin, -
netspeed	<ul style="list-style-type: none">• Connection Speed• Values: BROADBAND, XDSL, MOBILE, CABLE, DIALUP, WIRELESS, SATELLITE, T1, -, T3

lang	<ul style="list-style-type: none">• User Language• Values: en, fr, es, pt, ar, ru, tr, it, id, de, etc.
isp	<ul style="list-style-type: none">• ID of internet service provider
mobile_isp	<ul style="list-style-type: none">• ID of internet service provider for mobile traffic
proxy	<ul style="list-style-type: none">• Boolean flag showing whether traffic comes through proxy ip or not
vapp	<ul style="list-style-type: none">• Numeric variable
varpp	<ul style="list-style-type: none">• Numeric variable
varcc	<ul style="list-style-type: none">• Numeric Variable