

Customer Retention Prediction Challenge

The Case of a High-Street Barber Chain

1. Background

At Meizhi, we advise our client companies on how to drive business growth with a data approach. The dataset used in this case is a simulation of the kind of data that a tech-savvy high-street barber chain has accumulated over the course of expanding its business.

2. Database Connections

1. Type: MySQL
2. Host: 47.100.4.142
3. Port: 3306
4. Username: model
5. Password: ModelPwd123
6. Database: model

3. Dataset

1. orders

Field	Type	Comment
index	bigint(20)	主键
order_no	bigint(20)	订单号
customer_id	int(11)	用户id
store_id	int(11)	门店id
admin_id	int(11)	发型师id
all_price	float	应付金额
coupon_price	float	优惠金额
pay_price	float	实付金额
status	int(11)	订单状态 (0是产生了没有支付, 1是支付了没有服务, 2是服务等待评价, 负数的都是退款, 3是评价完成, 4是历史遗留的一个功能评价后分享给好友)

Thursday, 1 November 2018

type	int(11)	订单类型（服务订单： 1,2,3,5,7,10,20,26,28,4,8,9,13,14,15,19,25,27,29,30,32；产品 订单： 11,12,16,17,21,22,23,24,31）
pay_time	datetime	付款时间
serve_start_time	datetime	服务开始时间
serve_end_time	datetime	服务结束时间
appointment_time	datetime	预约时间

2. customers

Field	Type	Comment
customer_id	int(11)	用户id
customer_create_time	datetime	用户创建时间
sex	int(11)	用户性别（0未知 1男 2女）
age_group	int(11)	用户年龄区间（1小孩 2青年 3 中年 4老人）
hair_cut_num	int(11)	用户累计剪发次数

3. stores

Field	Type	Comment
store_id	int(11)	门店id
seat	int(11)	门店座位数
province_id	int(11)	省份id（2北京 25上海 31杭州 32重庆）
start_time	time	营业开始时间
end_time	time	营业结束时间
open_time	datetime	门店开业时间
is_temp	char(20)	1代表TRUE，0代表FALSE

4. stylists

Field	Type	Comment
stylist_id	int(11)	发型师id

sex	int(11)	发型师性别
class	int(11)	发型师等级
birthday	date	发型师生日
entry_time	datetime	发型师入职时间

5. coupon_use_details

Field	Type	Comment
coupon_id	bigint(20)	卡券id
user_id	int(10)	用户id
card_type	enum('GENERAL_COUPON','GIFT','DISCOUNT','CASH','FIXED','GROUPON')	GENERAL_COUPON: 优惠券, GIFT: 兑换券, DISCOUNT: 折扣券, CASH: 代金券, GROUPON: 团购券,
coupon_price	double	卡券金额
get_time	datetime	用户领取优惠券时间
use_time	datetime	用户使用优惠券时间
use_status	int(10)	使用状态 0.未领取 1.已领取 2.已使用 3.已作废
order_no	text	使用的订单号
send_type	int(10)	代金券发放类型: 1.网页领取 2.推送 3.老用户分享4.新用户下单反馈老用户
use_type	tinyint(1)	优惠券类型 1剪发服务优惠券 2商品的优惠券

6. comments

Field	Type	Comment
user_id	int(20)	用户id
comment_char_no	int(11)	评价内容字数
comment_star	int(11)	评价星级 (1-5星)
order_no	varchar(20)	订单号
judgement_label	text	评价标签

4. Questions

Thursday, 1 November 2018

1. Please use at least two different machine learning models to predict customer retention. Compare and contrast the models while staying very careful of the following aspects:
 1. How should retention be defined in order to make business sense?
 2. What metrics/aspects should be used in assessing the model?
 3. How much does each step in each model such as the model sensitivity to the various parameters contribute to reducing/explaining the unexplained?
 4. From a business sense, is there a model that is truly superior, i.e., not only may the model have greater empirical explaining power per the dataset given but also the working detail of the model fits the hairstyling business logic?
2. Please submit your solution in the form of an essay alongside your codes. There is no time limit - you decide what time frame you need to present your best capabilities.