# 爆款商品分析

### 一、logit模型计量分析

从excel表格数据可以看出，“是否为爆款”已经标注出了0、1标签，1为爆款，0为非爆款。某一商品是否为爆款，主要与买家的喜好、商品本身的质量以及商家正确的经营推广有关。因此根据现有数据首先可以提出如下假设：

1. 是否为爆款与店铺名称无关。因为同一店铺既有爆款商品也有非爆款商品，所以店铺名是无关字段。
2. 是否爆款与商品类目无直接关系。因为同一类目的商品（比如T恤）既有成为爆款的也有销量惨淡的。
3. 是否爆款与价格关系不大。尽管实惠往往容易受到买家欢迎，但实惠是建立在质量基础上的，价格不是重要的影响因素。以店铺1的T恤和衬衫为例，在价格相对一致的情况下，有的成为了爆款而有的不是。
4. “近30天的销量”和“总收藏”，这两个因素决定了商品是否能成为爆款。因为这两个字段直接体现出了买家的喜好。

下面以logit回归论证上述假设：

|  |  |  |  |
| --- | --- | --- | --- |
| **Logit model** | **y** | | |
|  | **Coef** | **T-value** | **Sig.** |
| **salesNum30d** | 0.0086 | 5.4746 | \*\*\* |
| **collectNum** | 0.0002 | 2.6282 | \*\*\* |
| **price** | -0.0001 | -1.5666 | — |
| **shopName\_2** | -3.7303 | -3.3245 | \*\*\* |
| **shopName\_3** | -6.0624 | -4.0462 | \*\*\* |
| **shopName\_4** | -5.7516 | -4.0226 | \*\*\* |
| **shopName\_5** | -11.8079 | -4.2788 | \*\*\* |
| **items\_半身裙** | 0.0062 | 0.0076 | — |
| **items\_卫衣/绒衫** | -1.2326 | -1.3904 | — |
| **items\_衬衫** | -3.9664 | -2.5073 | \*\* |
| **items\_连衣裙** | 0.3426 | 0.3111 | — |
| **Model:** | **Logit** | | |
| **No. Observations:** | 280 | | |
| **Pseudo R-squared:** | 0.815 | | |
| **Note: t statistics in parentheses. \* p<.1, \*\* p<.05, \*\*\*p<.01** | | | |

根据logit模型结果结果可知：

1. “近30天的销量”(salesNum30d)和“总收藏”(collectNum)与成为爆款的概率（y）显著正相关。假设4成立。
2. 商品价格（price）与是否成为爆款无关。假设3成立。
3. 相较于店铺1，其他店铺商品显著更不可能成为爆款。说明是否是爆款在店铺之间存在差异。由于不同店铺的经营推广不同，对是否是爆款产生了显著影响。假设一，不成立。
4. 相较于T恤，衬衫更不可能成为爆款，而其他商品相对于T恤而言，能否成为爆款并无显著差异。说明能否成为爆款，与商品类目差异关系不大。假设二成立。

综上，可得结论：

1. 爆款与商品销量、收藏量直接相关，与商品价格、类目无关。

2. 不同店铺的经营活动对能否成为爆款也产生了显著的影响。

### 二、树模型决策点分析

为了判断店铺“近30天的销量”和“总收藏”达到多少时才会成为爆款，从已有数据中对爆款产品进行简单平均，可以大致看出，店铺表现为如下两个特点：

1. 高销售低收藏（店铺1）。
2. 低销售高收藏（店铺2-5）。

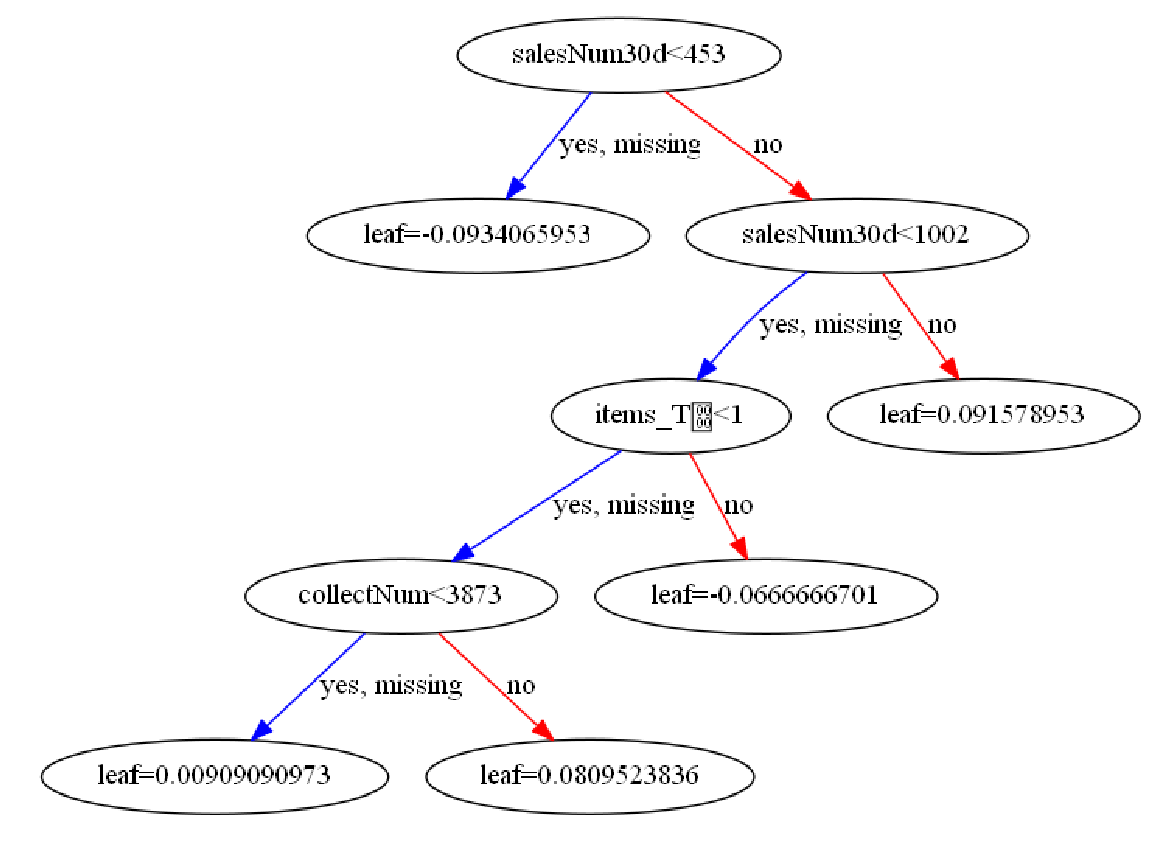
|  |  |  |
| --- | --- | --- |
| **shopName** | **salesNum30d** | **collectNum** |
| **1** | 4950.645 | 1008.194 |
| **2** | 1592.517 | 17193.41 |
| **3** | 1965.811 | 43119.22 |
| **4** | 1549.5 | 34807.91 |
| **5** | 3347.231 | 33994.12 |

通过简单分组求均值的方式，过于粗糙。此处采用随机森林模型、以及xgboost两种算法来训练模型，在保证预测准确率的情形，根据树模型的决策分裂点可以提取出爆款产品的分界点。

首先对数据进行如下处理：

|  |  |  |
| --- | --- | --- |
| 原数据列 | 新数据列 | 处理 |
| 店铺名称 | shopName | Onehot |
| 类目 | items | Onehot |
| 近30天销量 | salesNum30d | 不做处理 |
| 总收藏 | collectNum | 不做处理 |
| 价格 | price | 不做处理 |
| 是否为爆款 | y | 不做处理 |

经过训练，randomForest以及xgboost的accuracy均达到0.9857。此时以xgboost的CART树为例，可得出如下规则：

即是否爆款仅与“近30天销售量”以及“总收藏”、“是否体恤”有关；“近30天销售量”至少应大于453件；“近30天销售量”小于1002件时，非T恤的收藏量应达到3873等等。

【分析处理较粗糙，有很多经不起推敲的地方。。。】

杨双杰 17796401944

西南财经大学 财务管理(2019)