

基于决策树挖掘的淘宝女鞋的消费分析

赵雅菲 河北大学管理学院

摘要：本文根据最近一个月淘宝女鞋的销售数据，基于决策树模型，得出影响销售的因素。
关键词：决策树；淘宝女鞋；销售
中图分类号：TP301.6 文献标识码：A 文章编号：1001-828X(2017)016-0322-01

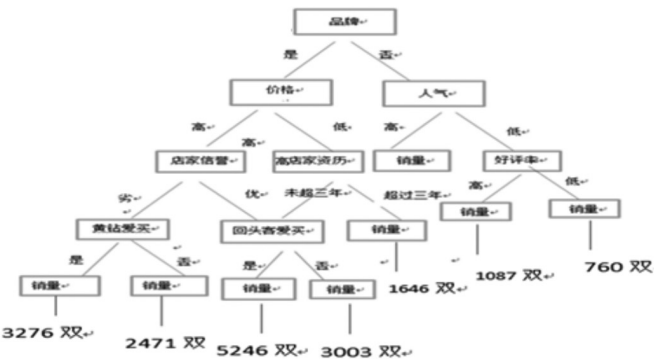
一、对抽样数据进行分析

本文分析使用的数据是近一个月淘宝女鞋的销售数据，影响销售的基本因素的数据包括：
基本因素：价格、服务态度、商品人气、卖家的资历、好评率、回头客爱买店铺、黄钻爱买店铺。
消费者相关数据：年龄、工作、教育、收入等。
一般来说销售数据及相关数据对我们预测相当重要，而消费者的基本数据例如年龄、工作、教育、收入等都会对销量产生一定的影响，所以进一步的分析是必要的。

在模型建置前，首先要了解数据的组成。通过统计数据，我们可以察觉数据的异常、极端值。以好评率栏位为例，我们可以通过最大、最小及平均值，来观察有无异常分布。好评率在 80%~90% 属于正常分布，所以不需要做特殊处理。其它栏位可以通过同样方式检验，以增加对电商上女鞋买家消费生成的数据的了解

二、构建决策树

本文决策树模型是依据影响销量的价格、人气、店家信誉、资历、好评率、回头客爱买店铺、黄钻爱买店铺这七种因素的数据量分类而构建的。下图是依据影响销量的价格、人气、店家信誉、资历、好评率、回头客爱买店铺、黄钻爱买店铺这七种因素的数据量分类而构建的决策树模型。



三、使用决策树方法进行分析的原因

通过和淘宝有关方面的商品质量审核人员及商务运营人员进行访谈，初步得出了可能影响商品销量的几个关键性的因素。在商品方面上，包括商品的质量好坏、价格的高低、原价的高低、折扣的多少、商品历史销售情况好坏、品牌的影响力、商品包邮与否、是否提供闪电发货与配送、是否提供商品相关的质检报告等；店铺方面，店铺的开店时长、星级、信用程度、商品的丰富度等。电子商务平台购物与一般的线下购物相比，其模式会在一定程度上刺激顾

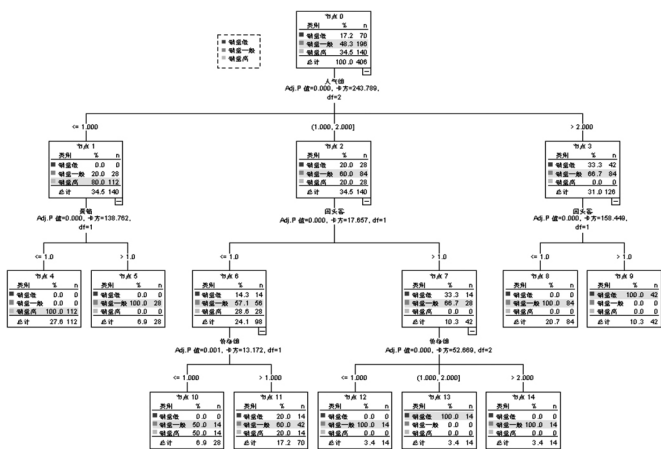
客的购买冲动，这里指的是累计的购买人数、超低的价格等因素。

决策树分类规则如下：

IF 品牌 = 是
THEN
IF 价格 = 低
IF 好评率 = 高
THEN 销量高

四、决策树生成

将数据录入到 spss 中对人气组，黄钻，回头客，价格组，品牌，资历，好评这些自变量进行分类，根据自变量的具体数据的大小分为高中低三等，同时也把销量分成高中低三等，选择的增长方法为 CHAID，通常，我们考察总的样本量大小，本文中样本数据量是 200，所以父节点最小个案数是 200，子节点最小个案数是 100，当然如果设定的太小，树会非常茂盛，得到很多非常小的细分市场。



从上述决策树中看出，基于目标变量（独立变量）自我分层的树状结构，根结点是因变量，预测变量根据卡方显著性程度不断自动生成父节点和子节点，卡方显著性越高，越先成为预测根结点的变量，程序自动归并预测变量的不同类，使之成为卡方显著性。最后每一个叶结点就是一个细分市场。我们可以得出对销售量影响最大的商品的价格，其次是商品的好评率和人气，回头客和黄钻爱买店铺占的比重最小。

作者简介：赵雅菲 (1993-)，女，汉族，河北邯郸人，硕士研究生在读，现就读于河北大学管理学院，管理科学与工程专业。