

关于京东化妆品的分析及建模

刘之汇 刘歆怡

云南大学 云南昆明 650000

摘要: 随着网络技术的发展,人们越来越重视网络贸易与实际生活的联系。本文利用所爬取到的京东综合排名前100页的化妆品各项指标数据来对其在销量进行分析并提出建议。在本次研究中,选取价格、品牌、销量数以及其他指标数据等,分析出对商品销量影响高的因素,研究不同商品客单价与其销量、销售额的关系,给出具体的分析结果。基于分析结果,本文对京东化妆品提出合理化建议,从而提高商品的吸引力和网友喜爱度来增加各化妆品的销售以影响力。

关键词: 梯度提升树; Python; 分析建议

Analysis and modeling of Jingdong Cosmetics

Zhihui Liu, Xinyi Liu

Yunnan University;Kunming,Yunnan,650000

Abstract: With the development of network technology, people pay more and more attention to the connection between network trade and real life. This paper makes use of the index data of cosmetics in the top 100 pages of Jingdong Comprehensive ranking to analyze its sales volume and put forward suggestions. In this study, price, brand, the number of sales, and other index data are selected to analyze the factors that have a high impact on the sales volume, study the relationship between the unit price of different commodity customers and their sales volume, and sales volume, and give specific analysis results. Based on the analysis results, this paper puts forward reasonable suggestions for Jingdong cosmetics to improve the attractiveness of the products and the popularity of users to increase the sales of various cosmetics to influence.

Keywords: Gradient lift tree; Python; Analysis and Suggestion

1. 引言

本文通过建立五种预测模型:全变量线性回归、回归树、随机森林、梯度提升树以及支持向量回归,并分别计算其均方误差来选取最优销量预测模型。根据预测模型均方误差以及模型的拟合优度的计算,本文通过对比以上五个模型的结果,最终选取梯度提升树回归预测模型为最佳销量预测模型。

1.1 研究背景

现今网络市场发展迅速,电商、直播带货等新型模型的出现促进了网络购物的飞速发展。作为数字经济新业态的典型代表,网络零售继续保持较快增长,成为推动消费扩容的重要力量。其作为打通生产和消费、线上和线下、城市和乡村、国内和国际的关键环节,在构建新发展格局中不断发挥积极作用^[1]。

1.2 研究目的和意义

由于时代的发展和人们对时间的重视,绝大部分人愿意在像天猫、淘宝、京东等线上购物APP进行消费。化妆品逐渐从奢侈品转变成了日常常见的生活用品。因此,如何在广大消费者中使用京东电商平台来推广化妆品吸引消费者是每一个化妆品公司下面面临及解决的重要问题。通过对该话题的研究和建议,可以对化妆品线

上销售、推广提供策略,品牌营销人员可以根据研究结果和建议对线上品牌宣传、销售等方面进行合理的规划和改进。

1.3 研究现状

在淘宝、京东等网购平台的销量研究中,大部分研究分为两个方向:一个是对影响店铺产品销量因素的研究。学者叶欣(2018)对影响店铺产品销量因素研究^[2]得出的结论是卖家信用、好评率、描述相符和售后率对销量有着正向影响。还有部分学者通过研究,得出价格对店铺销量有着显著负向影响,而信誉、服务质量对销量有显著正向影响^[2];国内学者多以淘宝店商品的影响因素来对淘宝店铺销量情况进行分析与研究。

另一类是基于文本分析研究店铺的销量。学者朱蜀江(2018)对淘宝网服装类的在线评论进行研究^[3],得到研究结果表明:评论字数和有图评论数对销量影响最大。国外电子商务相关研究较少,大部分研究都基于亚马逊等国外电子商务网站来进行。

国外学者的研究大部分认为服务质量对商品的销量影响最大,这与国内的研究基本一致。

1.4 改进与创新

通过阅读大量文献,我们发现大部分对电商的销量

研究，都是从只商品的价格、好评率和服务质量等方面进行^{[4][5]}，很少对别的因素进行分析和研究。因此，为了探究其他因素对网络商品销量的影响，本次从其他角度着手，通过对商品品牌^[6]、营销方式^[7]、商品种类和产地等指标进行研究分析，观察这些指标对商品销量的影响。综上，本文会填补现有研究对商品品牌、商品种类和产地等指标对商品销量研究的空白。

2、数据分析

2.1 化妆品销量分析

由于人们对不同商品的喜爱程度不相同，因此对商品的销量进行分析，通过分析销量靠后的产品特点来给出相应的营销和改进建议。

首先对销量较后的商品进行可视化分析。同样，由于商品数据集数量庞大，排名最后几个商品销量数非常低，甚至有些并没有超过 50。销量后十位的产品分别是白云山祛痘膏，雪柔雅防晒，南京同仁堂美白祛斑霜，大宝肌妍抗皱套装，欧莱雅礼盒，仁和匠心祛斑套装，MAMEW 晚樱精华套装，仁和匠心 7898 等，销量范围在 10~50 之间，这些产品的购买用户较少。

2.2 化妆品销售额分析

我们发现商品的销量高但销售额不一定高，因此接着来分析销售额高和销售额低的产品，观察出销量和销售额均低的商品，并根据商品本身性质提出建议。对销售额后 10 名的商品进行分析可以发现白云山祛痘膏、妙倩唇刷、古迪 0226 礼盒、自然堂套装、同仁堂祛斑霜、仁和匠心 7898 礼盒、古迪双眼皮贴、玛丽佳人黑胶和雪柔雅的男士防晒霜这十个化妆品的总销售额最低，且不高于 3000 元。

我们来分析销售额和销量均非常小的商品，并建议白云山的凝胶，雪柔雅的男士防晒霜，同仁堂的祛斑霜，仁和的 7898 礼盒和韩倩的唇线笔需要下架整改。

2.3 化妆品产地分析

不同省份的商品的销量也有所不同，其商品品牌数也不一样，因此对每个省份的销量在地图上展现和比较，如下图 1。

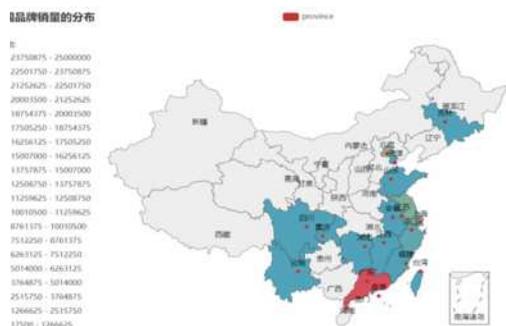


图 1 省份销量图

由于各个省份的品牌数不一样，因此不能仅凭借商品的总销量来判断商品产地的好坏。

2.4 化妆品品牌分析

不同品牌之间的顾客销量和喜爱度也不相同，因此在京东电商平台上综合排序前 100 页的商品中，销量是品牌热度最好的展现。对此，将对上架的对前 100 页的品牌销量数进行进行分析，并对销量靠后的品牌的销量和客单价进行可视化分析，由分析结果可以看出，MAMEW、韩倩、JEAN MISS、姬存希、希芸、纪梵希、梵贞 和 DI-YAMAN 这八个品牌的销量小于 1000，但 MAMEW、JEAN MISS 和希芸的客单价偏高，这可能是造成销量少的原因。

3、数据建模

3.1 销量预测

3.1.1 模型对比

本文通过建立全变量线性回归、回归树、随机森林、梯度提升树以及支持向量回归，并分别计算其均方误差来选取最优销量预测模型，计算结果如下表 1。

表 1 模型对比表

预测模型	均方误差	拟合优度
全变量线性回归	66953422765.465880	0.1126
回归树	27062063432.585920	0.6413
随机森林	17412599458.031593	0.7692
梯度提升树	15833339281.486557	0.7902
支持向量回归	78792641808.560490	-0.0443

因此从分析结果可以看出，全变量线性回归、回归树、随机森林、梯度提升树以及支持向量回归五种预测模型中，预测能力最差的模型是支持向量回归模型，能力最好的模型是梯度提升树回归模型。因此，本文选择建立梯度提升树回归模型作为预测模型。

3.1.2 梯度提升树回归

3.1.2.1 CART 回归树^[7]

CART 回归树的原则是最小均方差 (MSE)。即对任意划分特征 A，对应的任意划分点 s 两边划分成的数据集和，求出使和各自集合的均方差最小，同时和的均方差之和最小所对应的特征和特征值划分点。表达式为：

$$\min_{A,s} [\min_{c_1} \sum_{x_i \in D_1(A,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in D_2(A,s)} (y_i - c_2)^2] \quad \text{(公式 (5))}$$

其中，为数据集的样本输出均值，为数据集的样本输出均值。

3.1.3 结果

将建立好的梯度提升树回归预测模型对测试集进行预测，并将预测的结果与原测试集结果进行可视化对比，通过对比我们可以看出，梯度提升树回归预测模型对数据的预测准确度较高。

通过对模型中解释变量的重要程度进行可视化分析来体现整体模型的情况，变量重要性分析结果从高到低为价格、营销方式、品牌、产地、类型、使用人群、省份、地址。

从上图可以得出，价格、营销方式和品牌是影响化

妆品销量的重要因素，消费者最看重化妆品的价格。重要性最小的变量是使用人群，可见消费者对化妆品所适用的人群关注度不高。

4、结论建议

对于京东上综合排名前 100 页的化妆品品牌的各项指标数据的分析，为了提高其在顾客间的关注度和吸引力从而提高产品销量，我们建议：

(1) 对于销量高的商品，商家可以通过广告、知名明星代言等方式加大商品的宣传，以便吸引新顾客消费。对于销量低，且销售额不高的商品，商品价格、质量等各方面进行整改或完善。

(2) 通过对价格的分析，顾客更偏向于价格低且性价比高的商品，因此对于高价位商品，本文建议商家可以推出满减、折扣等优惠来降低价格刺激客户前来消费；对于 400 元以下的中低价位的商品，本文建议其商家可以通过赠送小样、赠品来吸引消费者，且能间接宣传该商品的其他类型产品。

(3) 根据品牌与产地的结合，对于国际性化妆品的商家，可以通过国内知名明星的代言来提高这些外国化妆品品牌的国内普及度。对于中国品牌，我们建议采取大力宣传、减免等优惠方法对其进行推广，并对销量低的商品进行下架整改或更新，加强宣传力度，多学习销量高品牌的营销方法。

(4) 从商品的销量、销售额和客单价的分析，对于销售额低而客单价偏高的商品，其销售少可能是品牌的普及度不高，本文建议商家可以根据奖励好评或多赠赠品、加强优惠力度等方法扩大其影响力，从而提高销

售量。

(5) 营销方式、价格和品牌是影响化妆品销量的重要因素，消费者最看重化妆品的营销方式。并且，消费者会根据化妆品的品牌和价格来决定是否购买该商品。因此营销人员可以通过明星广告，多平台宣传，制作成小电影等方式提高品牌的知名度，并改善营销方式以此来吸引更多的消费者。

参考文献：

[1] CCINC. 第 49 次《中国互联网络发展状况统计报告》. <http://www.cnnic.net.cn>.

[2] net.cn/hlwfzyj/hlwxzbg/hlwtjbg/202202/t20220225_71727.htm.2022.5.19

[3] 叶欣. 淘宝集市店铺手机销量影响因素分析 [D]. 中国社会科学院研究生院, 2018.

[4] 朱蜀江. 在线评论对服装类产品销量的影响 [D]. 云南大学, 2018.

[5] 谢琦. 电商促销活动下销量的影响因素研究 [D]. 上海交通大学, 2018. DOI :10.27307/d.cnki.gsjtu.2018.001765.

[6] 岳大川. 在线评论对市场竞争替代品销量的影响研究 [D]. 哈尔滨工业大学, 2017.

[7] 刘丽娜, 齐佳音, 张镇平, 曾丹. 品牌对商品在线销量的影响——基于海量商品评论的在线声誉和品牌知名度的调节作用研究 [J]. 数据分析与知识发现, 2018,2(09):10-21.

[8] 银河的猪猪. 随机森林回归算法. https://blog.csdn.net/qq_40783326/Article/details/103165660.2022.05.20