

“卷皮网”用户画像及精准推荐实战

大学生创新训练项目中期答辩

答辩人：孙嘉轩 黄爽 龙海文 杨岳浩 许艳

指导老师：张千帆

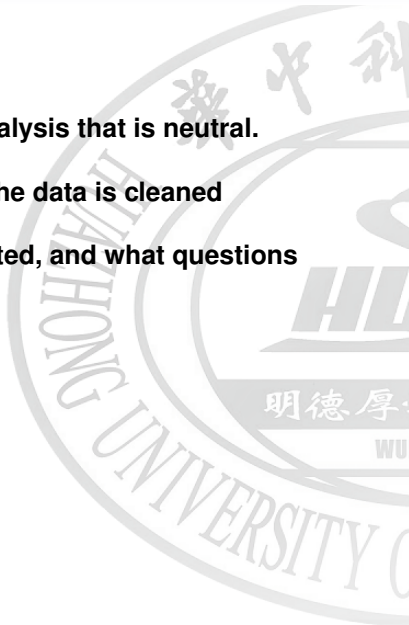
华中科技大学管理学院

2018 年 10 月 16 日

”There is nothing about doing data analysis that is neutral.

**What and how data is collected, how the data is cleaned
and stored, what models are constructed, and what questions
are asked – all of this is political.”**

Danah Boyd, NYU



目录

卷皮网研究情况

教程开发与编程社区建立

财务情况

后续工作安排



卷皮网概述

- 特点：主打低价商品的拼团类购物网站。
- 主营业务：日用品及百货的电子商务。
- 企业问题：重复购买客户数量少，客单价较低。



商品营销分析

- 数据来源：卷皮网数据分析员工提供。所有数据经过脱敏处理。
- 数据量：8种品类商品价格数据共2570组。
- 分析方法：统计、逻辑回归。
- 分析结论：
 - 各品类商品平均价格偏低，离散程度较小，商品质量与价格分布集中。
 - **99.8%**的商品以打折价出售，降价幅度较大，说明卷皮网消费者对商品打折信息较为敏感。
 - 不同时期商品浮动价格较大，一般以上商品不定期会有大幅度价格变动。

用户评论分析

- 数据来源：手机端APP评论数据+Python爬虫程序
- 数据量：8种品类商品，每种选取10个商品，爬取评论。平均每个商品约500条评论。
- 分析方法：文本分词，建立数据字典，分析情绪倾向性

用户评论分析

- 分析情况：
 - 评论少，无效评论较多，评论转化率低。
 - 用户评论字数偏少，差异性低。
 - 商品评论数量与评论倾向并不会显著影响商品销量。
- 结论：
 - 卷皮网面向的消费者评论关注程度低，且不会显著影响其购买决策。
 - 用户评论用词高度重复，差异性不够显著。

卷皮网项目结论

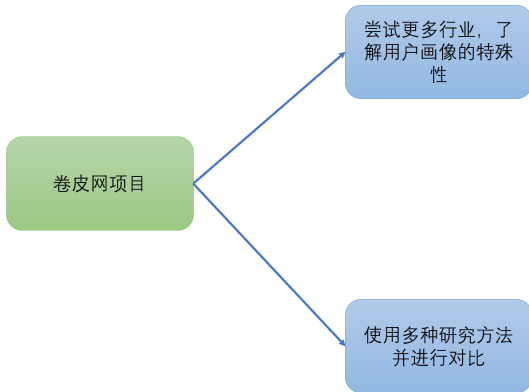
为什么卷皮网客单价低，用户重复购买意愿不强？

- 流量时效性强，其借助互联网渠道的转发难以维持固定客户。
- 商品质量参差不齐，山寨产品影响购物体验。
- 卷皮网消费者在网上发表评论意愿不强，重复度高。参考意义低。
- 营销模式偏低端，商品心里价格较低。

研究心得体会

- 研究目标用户群体的用户画像可以切实分析企业经营中的优势与问题。
- 机器学习的研究方法可以从另一个维度解读数据。
- 卷皮网的项目实践得出了对卷皮网目标用户群的画像，并得到了一定的结论。
- 卷皮网用户群像单一，数据质量不高，不具有很强的研究价值。

工作范围扩大



工作范围扩大

- 分析不同行业的数据和用户，全方位解读用户画像的企业应用场景。
- 对比不同编程语言及程序包的优劣。
- 通过建立网页版教程，系统归纳团队研究过程及心得，给入门用户提供学习的平台。
- 编写入门级机器学习实践案例，增强教程的实践性与互动性。

为什么要做教程？

- 在卷皮网项目开发过程中发现，相关技术资料不够系统，教程较少，推荐困难。
- 身边许多老师同学对此领域有很浓厚的兴趣。

网页

用户画像与机器学习实战指南

技术原理 案例分析 练习

前言

用户画像从何而来

在大数据、智能算法蓬勃发展的当今，传统业态正在面临新技术的冲击而重新洗牌，用户画像应运而生。人们开始关注如何从海量的数据中挖掘商业价值，切实的通过大数据的方法整合、优化用户体验，增加企业盈利能力，打造品牌价值。通过一个个数据完美的抽象出一个用户的信息全貌，这用户画像最原始的想法，也是企业应用大数据的根基。

什么是用户画像

用户画像是根据用户的属性、偏好、习惯、用户行为等信息抽象出的标签化用户模型，通俗地说就是给用户打标签，而标签是通过对用户信息分析而来的高度精练的特征标识。

如今，用户画像被广泛的使用和研究，现在已经产生的较为成熟的方法包括：

userprofileguide.github.io

Search in userprofileguide / userprofileguide.github.io

2,064 commits · 3 branches · 101 releases · 118 contributors · View license

Commit	Message	Time
jeanne remove test.md and update python.md	Latest commit asakaka 2 hours ago	
gFub	Change GitHub issue templates	14 days ago
_data	remove docs	3 days ago
_includes	Cache "static" includes to improve build performance [mistake#1874]	9 days ago
_layouts	Cache "static" includes to improve build performance [mistake#1874]	9 days ago
_seo	Make entire archive/feature item "clickable" [mistake#1864]	11 days ago
assets	Release 4.13.0	a month ago
practice	add main.md	3 days ago
samples	update samples link	4 hours ago
technical_resource	remove test.md and update python.md	2 hours ago
test	Cache "static" includes to improve build performance [mistake#1874]	9 days ago
中環區	上传中環區dev文档	22 hours ago

教程框架设计

- 技术原理：对概念和基础知识点的介绍、程序包的使用方法介绍、算法原理介绍等。
- 案例分析：全面记录团队项目开发过程和在其中使用到的研究方法、编程技巧、优化思路，并进行探讨比较。
- 入门练习：提供让入门读者快速上手实践并理解机器学习和用户画像的案例，数据与源代码。

已完成教程

- Python编程语言基础教程
- 机器学习简介
- 用户画像简介及企业实例
- 银行客户流失案例分析
- R语言基础及机器学习入门



即将上线教程

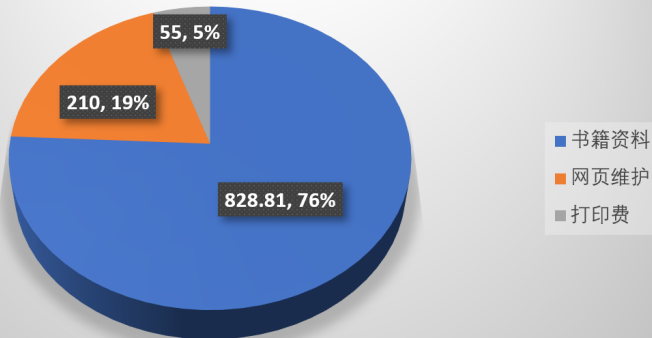
- 卷皮网项目开发历程，数据，代码，总结
- 机器学习主流论坛与资源获取站点
- 如何使用GitHub进行项目开发
- Kaggle的使用方法
- Stack Overflow上有哪些有用的项目实战经验？

以上教程将在一个月内上线



财务情况

财务情况统计



后续工作安排

- 拓宽研究领域，完成更多项目开发案例。
- 继续完善教程主体部分和延申部分，增强内容的完整度与可读性。
- 与企业进行合作，进一步探讨用户画像的商业价值与伦理学边界。
- 扩大影响力，打造可持续发展编程社区，不断引入更多优质内容，吸引越来越多的同行加入。