

“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 18 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

工作范围扩大与教程开发

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。

卷皮网研究情况
000000

工作范围扩大与数据开发
0000000000

财务状况
0

后续工作安排
00

商品营销分析

- 数据来源：卷皮网数据分析工程师提供。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，同品类商品价格离散程度较小。
 - 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。

卷皮网研究情况
000000

工作范围扩大与数据开发
0000000000

财务状况
0

后续工作安排
00

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，高频词分析。
 - 特征提取，用户情绪分析。
 - 评论关联度分析。

卷皮网研究情况
000000

工作范围扩大与数据开发
0000000000

财务状况
0

后续工作安排
00

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对评论关注程度低，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。

卷皮网研究情况
000000

工作范围扩大与数据开发
0000000000

财务状况
0

后续工作安排
00

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，消费者价格感知程度低。

卷皮网研究情况
○○○○○●

工作范围扩大与教程开发
○○○○○○○○○

财务情况
○

后续工作安排
○○

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，并进行了系统性的描述。
- 卷皮网用户群像单一，数据量较小，特征偏好明显。

卷皮网研究情况
○○○○○○○

工作范围扩大与教程开发
●○○○○○○○○○

财务情况
○

后续工作安排
○○

工作范围扩大

```
graph LR; A[卷皮网项目] --> B[尝试更多行业，了解用户画像的特殊性]; A --> C[使用多种研究方法并进行对比];
```

卷皮网研究情况
○○○○○○○

工作范围扩大与教程开发
○●○○○○○○○

财务情况
○

后续工作安排
○○

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。

卷皮网研究情况
○○○○○○○

工作范围扩大与教程开发
○○●○○○○○○○

财务情况
○

后续工作安排
○○

例子：中国电信用户流失画像

- 研究路径：
 - 数据导入及预处理
 - 数据建模（采用决策树模型）
 - 总结和思考
- 数据类型：
 - 用户月套餐余额
 - 每月套餐外通话时长
 - 每月套餐外流量
 - 办理的是否是集团业务
 -

卷皮网研究情况
○○○○○○

工作范围扩大与教程开发
○○○○●○○○○

财务状况
○

后续工作安排
○○

例子：中国电信用户流失画像

研究结论：

- 高价值用户的重点关注更有利于企业的发展。
- 维护老用户比获取新用户效用更高。
- 优惠福利和竞争对手的销售策略都会对业绩产生显著的影响。

卷皮网研究情况
○○○○○○

工作范围扩大与教程开发
○○○○●○○○○

财务状况
○

后续工作安排
○○

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，项目推进困难。
- 身边许多老师同学对此领域有很浓厚的兴趣。

卷皮网研究情况
○○○○○○

工作范围扩大与教程开发
○○○○●○○○○

财务状况
○

后续工作安排
○○

网页

用户画像与机器学习实践指南

userprofileguide.github.io

卷皮网研究情况
○○○○○○

工作范围扩大与教程开发
○○○○●○○○○

财务状况
○

后续工作安排
○○

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。

卷皮网研究情况
○○○○○○○

工作范围扩大与教程开发
○○○○○○○●○

财务情况
○

后续工作安排
○○

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

卷皮网研究情况
○○○○○○○

工作范围扩大与教程开发
○○○○○○○●●

财务情况
○

后续工作安排
○○

即将上线

技术：

- 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？

案例：

- 社交网络用户情绪分析
- 购物网站商家决策方式与影响

以上教程将在一个月线上线

卷皮网研究情况
○○○○○○○

工作范围扩大与教程开发
○○○○○○○●○

财务情况
●

后续工作安排
○○

财务情况

财务情况统计

类别	金额	占比
书籍资料	828.81	76%
网页维护	210	19%
打印费	55	5%

卷皮网研究情况
○○○○○○○

工作范围扩大与教程开发
○○○○○○○●○

财务情况
○

后续工作安排
●○

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。

疫情期间研究情况 ○○○○○○○	工作范围扩大与教程开发 ○○○○○○○○○	财务状况 ○	后续工作安排 ●●
---------------------	--------------------------	-----------	--------------

感谢聆听!

Q&A

