E-Commerce Letters 电子商务评论, 2025, 14(10), 669-679
Published Online October 2025 in Hans. https://www.hanspub.org/journal/ecl
https://doi.org/10.12677/ecl.2025.14103194

Hans汉斯

电商平台服装畅销影响因素研究

——基于Logistic回归与文本挖掘

詹佳敏

贵州大学数学与统计学院，贵州 贵阳

收稿日期：2025年9月8日；录用日期：2025年9月22日；发布日期：2025年10月11日

摘　要

为探究影响电商服装畅销的因素，本研究结合Logistic回归与文本挖掘方法，分析了淘宝平台8800款商品数据及2400条评论。Logistic回归分析表明，“七天退换”、“正品保障”、“天猫认证”和“有视频”显著促进服装畅销，而“性别”(男装)、“价格”及“折扣”则抑制服装畅销；文本挖掘进一步验证了产品质量与体验的核心地位，揭示了性别差异：女性消费者更关注服装细节与款式，男性则更看重功能实用性与购物效率。本研究为商家提供了基于实证的运营优化建议。

关键词

电子商务，服装销售，Logistic回归，文本挖掘

Research on the Influencing Factors of Clothing Sales on E-Commerce Platforms

—Based on Logistic Regression and Text Mining

Jiamin Zhan

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: September 8, 2025; accepted: September 22, 2025; published: October 11, 2025

Abstract

To investigate the factors influencing the sales of clothing on e-commerce platforms, this study employed logistic regression and text mining methods to analyze data from 8800 product listings and 2400 customer reviews on Taobao. The Logistic regression analysis revealed that “7-day return and exchange”, “authenticity guarantee”, “Tmall certification” and “product video” significantly promoted

文章引用: 詹佳敏. 电商平台服装畅销影响因素研究[J]. 电子商务评论, 2025, 14(10): 669-679.
DOI: 10.12677/ecl.2025.14103194

sales, whereas factors such as "gender" (men's clothing), "price" and "discount" had inhibitory effects. Text mining further validated the core importance of product quality and user experience, highlighting gender-specific differences: female consumers placed greater emphasis on clothing details and aesthetics, while male consumers prioritized functional utility and shopping efficiency. This study provides empirically supported operational recommendations for e-commerce merchants.

Keywords

E-Commerce, Clothing Sales, Logistic Regression, Text Mining

Copyright © 2025 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

随着数字经济对消费市场的逐步渗透，电商平台已成为服装零售的重要渠道。线上渠道不仅重塑了服装消费的场景与习惯，改变了传统的服装销售服务，还因购物环境缺乏实体触感，使得消费者更加依赖商品描述、用户评论、服务保障等多元信息来降低感知风险(Bhatnagar & Ghose, 2004) [1]。同时，服装品类的销售具有强主观性、快时尚性与体验依赖性，消费者在电子平台购买服装的决策既受价格、销售保障等因素影响，也会因款式描述、面料评价、尺码反馈等文本信息中隐含的信息而改变，如何识别这些多维度元素对销量的影响，成为电商平台优化运营策略、服装商家提高销量的关键问题。

从目前国内外的研究来看，学界对电商平台商品销量影响因素的探讨主要有两个方向：一个是基于因子分析或聚类分析，通过市场调研获取信息，将影响因素划分为不同的类别，通过对比各聚类的得分快速定位"高销量聚类"的核心元素组合，如雷晶等(2014) [2]以江苏某品牌女装做的市场细分研究；另一类则围绕用户评论、商品详情页等文本数据，采用文本挖掘方法获得消费者对产品质量、服务体验的态度倾向，如崔志超(2025) [3]根据评论设计的情感分析系统及 Agarap (2020) [4]使用双向循环神经网络(RNN)进行的情感分类。但是鲜少有对商品畅销的影响因素给出一个明确的量化结果。

理论上，Engel 等提出的消费者决策模型(1995) [5]指出，购买行为是信息处理、评价与决策的多阶段过程，线上信息构成关键输入。DeLone 等提出的信息系统成功模型(2003) [6]则强调，系统质量、信息质量与服务质量共同影响用户满意度与使用意愿。这为理解电商环境中商品信息、服务保障与用户反馈如何共同影响销售提供了坚实框架。

基于此，本研究以国内主流服装电商平台淘宝的销售数据为研究对象，系统解析服装畅销的影响因素。具体而言，研究首先选取价格、折扣、七天无理由退换等 12 项结构化指标，构建 Logistic 回归模型，以"商品是否畅销"为因变量，量化各因素对销量分类结果的影响方向与显著程度；其次，通过文本挖掘进一步验证及挖掘畅销商品评论中的隐形影响因素。

2. 数据说明及描述

2.1. 数据说明

本研究使用的数据来源于淘宝电商平台的服装商品销售记录，经过数据清洗与预处理后，最终用于分析的有效样本共有 8800 个商品，其中男装样本和女装样本各 4400 个，包含 12 个变量(11 个特征变量，

1 个目标变量)，涵盖商品属性、价格策略、服务保障、产地信息等多个维度。

本研究的数据具有一定的局限性，由于样本仅来源于淘宝平台，结论普适性可能受限；此外，为了后续研究方法的使用，对于“畅销”的定义为销量超过样本均值，这是一个相对标准，且不区分“爆款”与“普通畅销”间的差异。

各特征变量的名称与具体定义如表 1 所示：

Table 1. Description of feature variables
表 1. 特征变量描述

特征名称	特征描述
季节	0 = 秋冬装；1 = 春夏装。
性别	0 = 女装；1 = 男装。
价格	商品出售的实际价格：0 = 0~49.99 元；1 = 50~99.99 元；2 = 100~149.99 元；3 = 150~199.99 元；4 = 200 元以上。
原价格	商品原本标注的价格：0 = 0~99.99 元；1 = 100~199.99 元；2 = 200~299.99 元；3 = 300~399.99 元；4 = 400 元以上。
折扣	价格与原价格的比值，比值越小，商品减免力度越大，比值为 1 则代表实际价格与原价格一致：0 = 小于 0.10；1 = 0.11~0.40；2 = 0.41~0.70；3 = 0.71~0.99；4 = 1。
包邮	0 = 否；1 = 是。
产地	商品生产地址：0 = 浙江；1 = 江苏；2 = 上海；3 = 广东；4 = 福建；5 = 其他地区。
天猫认证	0 = 否；1 = 是。
七天退换	0 = 否；1 = 是。
正品保障	0 = 否；1 = 是。
有视频	0 = 否；1 = 是。
是否畅销	销量大于 8800 个商品的平均销量的服装为畅销，销量小于或等于平均销量的服装为不畅销：0 = 否；1 = 是。

2.2. 数据描述

对本研究各变量的分布占比进行统计描述，结果如表 2 所示：

Table 2. Proportion of feature variables
表 2. 特征变量比例

	季节	性别	价格	原价格	折扣	包邮	产地	天猫认证	七天退换	正品保障	有视频	是否畅销
0	81.85%	50.00%	41.89%	22.93%	3.85%	3.75%	30.39%	48.22%	57.41%	56.74%	66.18%	76.00%
1	18.15%	50.00%	39.53%	42.50%	53.32%	96.25%	16.25%	51.78%	42.59%	43.26%	33.82%	24.00%
2			10.35%	17.41%	29.80%		5.49%					
3			5.43%	8.42%	4.74%		29.61%					
4			2.80%	8.74%	8.30%		9.97%					
5							8.30%					

通过对样本的描述性分析，可以得到以下结果：

从市场特征上看，数据中秋冬装的占比较高，有 81.85%；价格小于 100 的服装占比 81.42%，说明中低价位服装占服装市场的主导地位，商家普遍提供包邮服务和高力度折扣，一定程度上反映了激烈的市场竞争环境。

从服务与认证上看，天猫认证、七天退换和正品保障的占比分别为 51.78%、42.59%和 43.26%，服务的覆盖率仍有提升空间，这可能是影响服装是否畅销的关键因素。

从内容营销上看，商品视频的采用率较低，说明图片和文字依旧是商家展现服装的主要手段，视觉营销仍有发展空间。

从产销地上看，商品产地高度集中于浙江和广东，分别占服装产地的 30.39%和 29.62%，说明该两地是电商平台服装产业生产和营销的主要地区，体现了中国服装产业的集群效应。

从目标变量的分布上看，畅销服装占比 24%，远低于不畅销服装的 76%，这表明大部分的服装的销量都没有达到本研究所服装销量的平均值，该现象符合电子商务中的“长尾理论”[7]，即少数爆款商品贡献了大部分销售额。

3. Logistic 回归分析

Logistic 回归模型是广义线性模型(Generalized Linear Model, GLM)的一个重要分支，主要用于处理因变量为二分类情况的建模问题。该模型通过联结函数(Link Function)将自变量与因变量之间的非线性关系转化为线性关系，进而估计事件发生的概率(Hosmer & Lemeshow, 2013) [8]。在本研究中，该模型被用于分析影响服装是否畅销的关键因素，具有重要的理论价值与应用意义。

3.1. 模型解释

设因变量 Y 为一离散二值变量，其中 $Y=1$ 表示事件发生(本文中指服装畅销)，$Y=0$ 表示事件未发生(本文中指服装不畅销)。令 $p=P\left(Y=1|X\right)$ 表示在给定自变量 X 的条件下事件发生的条件概率。

Logistic 回归通过 Logit 联结函数建立自变量与因变量之间的关系[9]，模型的基本形式为：

$$\ln\left(\frac{p}{1-p}\right)=\beta_0+\beta_1X_1+\beta_2X_2+\cdots++\beta_kX_k \tag{1}$$

其中，β_0 为截距项，$\beta_1,\cdots,\beta_k$ 为自变量的回归系数，$k=1,2,3,\cdots$。通过 Logit 反函数(即 Sigmoid 函数)，可得到事件发生的概率估计：

$$p=\frac{\exp\left(\beta_0+\sum_{j=1}^{k}\beta_jX_j\right)}{1+\exp\left(\beta_0+\sum_{j=1}^{k}\beta_jX_j\right)} \tag{2}$$

该表达式确保概率预测值严格约束于区间[0, 1]内，从而克服了一般线性模型在分类问题中的局限性。

在 Logistic 回归中，系数 β_j 表示在其他变量保持不变的情况下，自变量 X_j 每变动一个单位，所导致的对数发生比(Log-Odds)的边际变化。为便于实证解释，通常进一步计算优势比(Odds Ratio, OR)：

$$OR_j=\exp\left(\beta_j\right) \tag{3}$$

若 $OR_j>1$ 表明 X_j 为风险因素，其取值增加将提升事件发生概率；若 $OR_j<1$ 则说明 X_j 为保护因素，其值增大会抑制事件发生；若 $OR_j=1$ 则表示 X_j 对事件发生无显著影响。

3.2. 似然比检验

模型似然比检验用于对整体模型有效性进行分析。此处模型检验的原定假设为：是否放入自变量(季

节，性别，价格，原价格，折扣，包邮，产地，天猫认证，七天退换，正品保障，有视频)时模型质量均一样，检验结果如下表 3 所示：

Table 3. Likelihood ratio test results
表 3. 似然比检验结果

模型	−2 倍对数似然值	卡方值	df	p
仅截距	9699.007			
最终模型	7254.391	2444.615	11	<2e^(−16)

检验结果的 p 值小于 0.05，因而说明拒绝原定假设，即说明本次构建模型时，放入的自变量具有有效性，本次模型构建有意义。

3.3. Logistic 回归结果与分析

将季节，性别，价格，原价格，折扣，包邮，产地，天猫认证，七天退换，正品保障，有视频为自变量，而将是否畅销作为因变量进行 Logistic 回归分析，所得结果如下表 4 所示：

Table 4. Logistic regression results
表 4. Logistic 回归结果

项	回归系数	标准误	z 值	Wald χ^2	p 值	OR 值
季节	−0.238	0.078	−3.038	9.23	0.002	0.789
性别	−0.586	0.066	−8.878	78.82	<2e^(−16)	0.557
价格	−0.321	0.052	−6.231	38.831	9.09e^(−11)	0.725
原价格	−0.13	0.039	−3.294	10.851	0.001	0.878
折扣	−0.207	0.045	−4.598	21.143	3.66e^(−05)	0.813
包邮	0.465	0.239	1.949	3.799	0.779	1.592
产地	−0.005	0.018	−0.296	0.087	0.767	0.995
天猫认证	0.656	0.064	10.175	103.535	<2e^(−16)	1.927
七天退换	2.029	0.064	31.726	1006.545	<2e^(−16)	7.604
正品保障	1.254	0.06	20.925	437.844	<2e^(−16)	3.504
有视频	0.784	0.061	12.936	167.342	<2e^(−16)	2.19
截距	−2.965	0.27	−10.967	120.273	<2e^(−16)	0.052

根据分析结果，可建立 Logistic 回归模型形式如下：

$$\begin{aligned}\ln\left(\frac{p}{1-p}\right) = &-2.965 - 0.238\times\text{季节} - 0.586\times\text{性别} - 0.321\times\text{价格} - 0.130\times\text{原价格}\\ &- 0.207\times\text{折扣} + 0.465\times\text{包邮} - 0.005\times\text{产地} + 0.656\times\text{天猫认证}\\ &+ 2.029\times\text{七天退换} + 1.254\times\text{正品保障} + 0.784\times\text{有视频}\end{aligned} \tag{4}$$

其中，p 表示商品“畅销”的概率，$1-p$ 表示“不畅销”的概率。

回归结果显示：

季节、性别、价格、原价格和折扣的回归系数均为负数，p 值均小于 0.05，故它们对服装是否畅销存在显著的负向影响，其优势比(OR)分别为 0.789、0.557、0.725、0.878 和 0.813，即意味着春夏装和男装相对于秋冬装和女装会使畅销幅度分别减少 0.789、0.557 倍，价格、原价格和折扣每增加一个单位，畅销幅度分别减少 0.725、0.878 和 0.813 倍。

天猫认证、七天退换、正品保障和有视频的回归系数均为正数，p 值均小于 0.05，对服装是否畅销存在显著的正向影响，其 OR 值分别为 1.927、7.604、3.504 和 2.190。即意味着提供天猫认证、七天退换、正品保障和有视频可以使畅销幅度分别增加 1.927、7.604、3.504 和 2.190 倍。

包邮和产地的 p 值均大于 0.05，说明这两个变量对是否畅销的影响在统计上不明显，对服装是否畅销影响较小。

4. 文本挖掘

本研究采用分层抽样方法分别选取平台上 6 个畅销男装商品和 6 个畅销女装商品，从每个服装商品的近期评论中随机抽取 200 条有效评论，最终构成一个总量为 2400 条评论的样本集，其中畅销男装评论 1200 条，畅销女装评论 1200 条，平衡性别品类以便于后续进行对比分析。

4.1. 总评论词云图分析

词云图是一种将文本数据中高频出现的关键词进行视觉化展示的数据可视化技术。首先，对收集到的 2400 条非结构化评论文本进行数据清洗，去除无意义的符号、数字及统一大小写；其次，运用 Jieba 分词库对评论进行分词处理，并滤除通用停用词表(如“的”、“了”、“吗”)以及领域特定停用词(如“衣服”)；然后，计算所有词汇在语料库中的出现频次；最后，基于词频统计结果，抓取 100 个高频词利用 WordCloud 库生成词云图如图 1 所示，其中词汇的字体大小与其出现频率成正比。

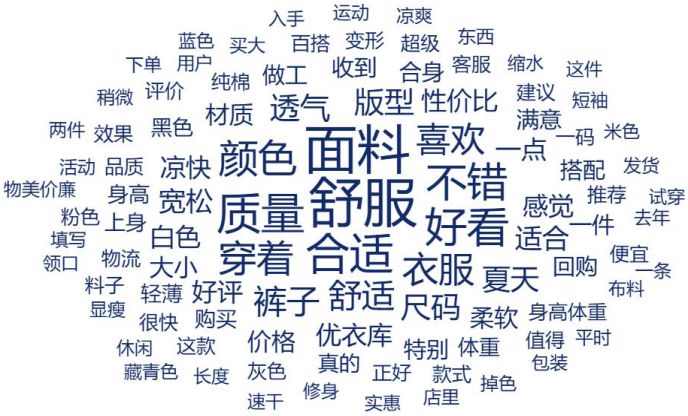

Figure 1. Word cloud of total reviews
图 1. 总评论词云图

“舒服”和“面料”在图中的占比最大，说明消费者在畅销服装的评论中最为看重的是服装的上身体验和材质。通过对词云中 100 个显性词汇的归纳，可识别出消费者评论关注的 4 个主要方向：

1) 材质与穿着体验：消费者高度关注服装的基础品质与上身感受。评论集中于面料的舒适度(如“纯棉”、“透气”)和版型的合身效果(如“显瘦”、“合身”)，这是影响消费者满意度的直接因素。

2) 外观与风格：消费者重视服装的样式和搭配实用性。多种颜色和外观的关键词(如“好看”、“百搭”、“休闲”、“蓝色”、“黑色”等)提示商家需提供丰富的色彩选择，并注重产品设计的美观性与

日常搭配的便利性。

3) 质量与性价比：消费者对品质可靠性和价格存在双重关注。正面词汇(如“物美价廉”、“性价比”)与潜在负面词汇(如“变形”、“缩水”)并存，说明产品质量是赢得口碑的基石，任何缺陷都会在评论中迅速暴露。

4) 消费体验：消费者关注购物流程与售后服务。高效的物流、良好的客服能显著提升用户体验，是促使回购和产生好评的重要驱动因素。

4.2. 性别差异分析

分别构建畅销女装和畅销男装评论词云图分别如图 2 与图 3 所示，以此直观展示不同性别消费者对于服装关注点的差异。

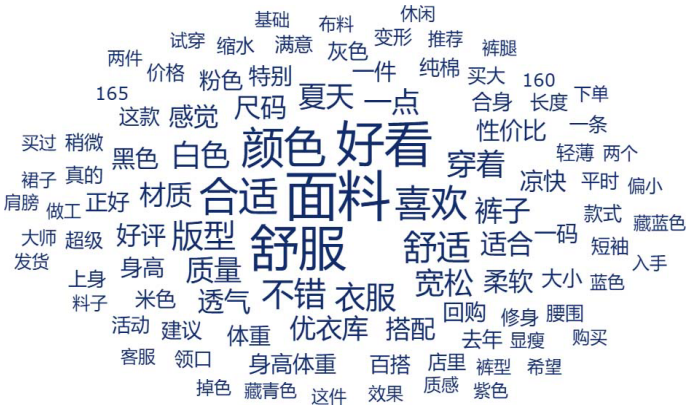

Figure 2. Word cloud of reviews for best-selling women’s clothing
图 2. 畅销女装评价词云图

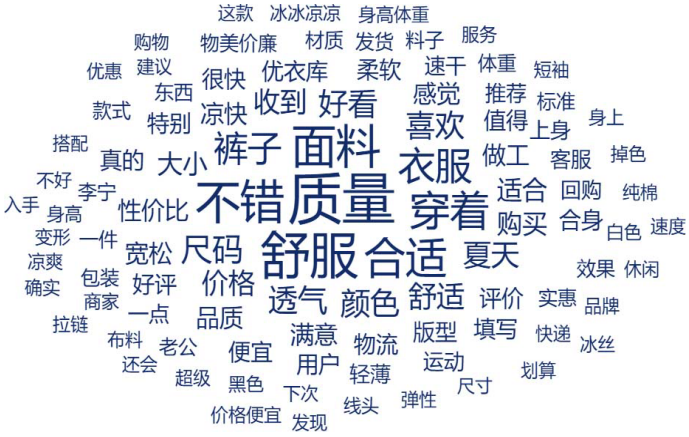

Figure 3. Word cloud of reviews for best-selling men’s clothing
图 3. 畅销男装评价词云图

可见，女装消费者和男装消费者同样注重服装的材质和穿着体验，但是相比于男装消费者，女装消费者更关注服装的外观。通过对其 100 个高频词的归纳分析，可以发现男女消费者在评论关注点上存在的差异具体表现在以下 3 个方面：

1) 版型与细节：女装评论呈现出对合身性与细节的高度关注，高频词包括“尺码”、“合身”、“腰

围”、“裤腿”、“长度”等具体指标的出现频率显著高于男装评论，表明女装消费者对服装的合身度要求更为严格和细致。相比之下，男装评论关键词如“舒服”、“合适”、“宽松”、“尺码”等更强调整体舒适性与便捷性。

2) 材质功能：在材质方面，两类消费者都关注“纯棉”和“透气”等基础特性，但女装评论中，“柔软”、“轻薄”、“凉快”等描述触感与体感的词汇出现频率较高，反映出女装消费者对材质舒适性的细腻追求。男装评论中如“速干”、“弹性”、“冰冰凉凉”等词则频繁出现，体现了男装消费者对服装实用功能的重视。

3) 消费行为：女装评论中充满了情感表达与视觉审美相关的词汇，如“好看”、“颜色”(包含多种具体色彩名称)、“推荐”、“回购”、“喜欢”等，反映出女装消费者的购物过程具有较强的情感参与度和审美倾向。男装评论则体现出效率导向的特征，关键词包括“很快”、“发货”、“快递”、“服务”、“回购”、“划算”等，表明男装消费者更重视购物过程的效率和交易的便利性。

4.3. 总评论网络关系图分析

网络关系图是一种基于图论用于展示节点之间关系的可视化工具，每个节点就是一个关键词，连线说明两者在评价中存在共现关系，连线较粗的两个节点意味着它们在用户评价文本中共同出现的次数较多，关系更为紧密。本研究抓取词频最高的 20 个关键词，生成畅销服装评价网络关系图如图 4 所示：

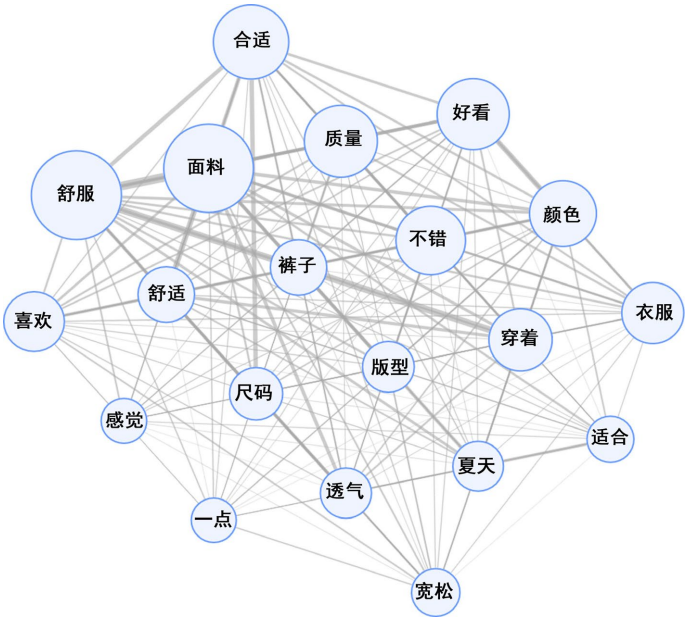

Figure 4. Network relationship graph of total reviews
图 4. 总评论网络关系图

图中“合适”“面料”“质量”“好看”“舒服”等节点较为突出，连线较多。这表明在服装评价相关文本中，这些词汇与其他词汇共同出现的频率较高。例如“合适”经常与“尺码”“舒适”等一起被提及，说明用户在评价服装时，合身程度是一个重要考量因素，且会连带描述穿着感受和尺码情况。“舒适”和“舒服”之间的连线暗示这两个词含义相近且在文本中常相伴出现，属于语义相近的高频词对；“裤子”与“版型”“面料”等有较多连线，说明在描述裤子时，版型和面料是主要的评价维度；“夏天”与“透气”“宽松”联系紧密，体现出夏季服装的特性需求。整个网络关系图呈现出较为复杂的网状

结构，说明消费者对服装的评价是综合考虑多个方面后得出的结论。

4.4. LDA 主题模型分析

为深度挖掘评论底层结构，本研究采用 LDA 主题模型对 2400 条评论进行无监督主题聚类。经过困惑度与一致性分数评估，发现当主题数为 3 时，运行后得到了三个可解释且区分较为明显的主题，主题与其重要词的权重关系如图 5 所示：

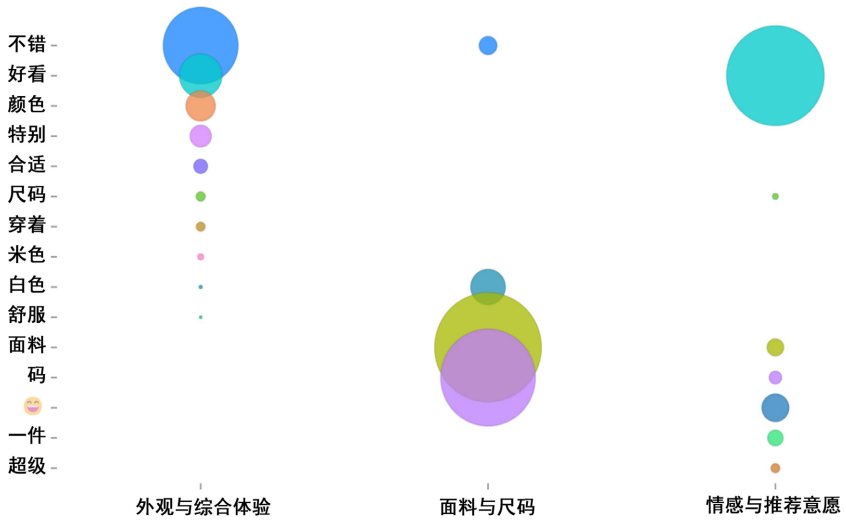

Figure 5. Relationship graph of topic and significant terms weight
图 5. 主题与重要词权重关系图

根据重要词为主题词命名如下：

主题 1 为外观与综合体验，其核心关键词包括好看、颜色、特别、合适等。这是最丰富、最综合的一个主题。它涵盖了服装的视觉审美(好看、颜色、特别)、合身度(合适)和基础舒适度(舒服)，消费者在这个主题下表达的是对产品整体上身效果的评价。

主题 2 为面料与尺码，其核心关键词为面料、码、白色等，该主题几乎完全由“面料”和“码”(指尺码，如 S/M/L)主导，附带提到了“白色”。这表明有大量消费者的评价核心紧紧围绕着材质的质感和尺码的准确性。

主题 3 为情感与推荐意愿，其核心关键词有不错、好看、表情、超级等。这个主题充满了积极的情感色彩，它通过“不错”、“好看”等词汇表达对产品整体的认可，并使用表情符号和“超级”等强调词来传递愉悦和兴奋的情绪。这个主题反映的是消费者在收到符合或超出预期的商品后产生的情感共鸣和推荐意愿。

随后，本研究根据从消费者评论中提取出的三个核心主题分析其分布情况，绘制饼图如图 6 所示。

外观与综合体验占比最高，为 83%，超过八成的消费者评论这个主题，这是消费者决策的核心和商家竞争的主战场，在线上环境中，消费者无法触摸实物，“外观”和“上身效果”是他们最重要的决策依据，所以在视觉呈现、版型设计、颜色准确性上的优化很可能获得最大的回报。

其次是情感与推荐意愿，占比 10%，该部分评论表达了强烈的情感(如“超级”喜欢)和推荐意愿(如“不错”、“买一件”)。这是客户口碑传播的体现，虽然占比不高，但是是产生复购和吸引新消费者的关键。

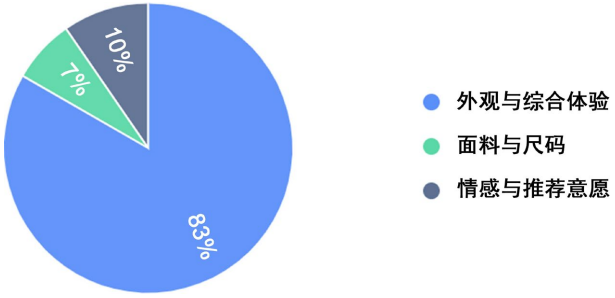

Figure 6. Topic distribution graph
图 6. 主题分布图

最后是面料与尺码，占比 7%，这并非意味着面料和尺码不重要，这个低占比可能暗示着消费者觉得面料和尺码符合描述是理所应当的，不会特别提及，都是一旦面料或尺码出现问题，负面评论会立刻出现，造成严重的负面影响。

5. 结论与建议

5.1. 结论

1) Logistic 回归模型表明，服务保障、平台认证与性价比是影响服装畅销的重要因素。七天退换、正品保障与天猫认证对商品畅销具有极显著的正向影响。同时，商品提供视频能显著提升畅销概率，证明动态视觉营销可以弥补线上购物体验短板。反之，价格、原价格、折扣等因素呈显著负向影响，表明中低价位市场是畅销主体，消费者表现出价格敏感性。包邮虽普及率高，但对畅销无显著解释力，可能因同质化严重。

2) 文本挖掘表明，产品质量与穿着体验是服装畅销的基石。用户评论高频词分析表明，舒适、合身是消费者关注的核心。同时，性价比及客服等词的高频出现，与回归模型结论相互印证，证明了价值和服务是赢得市场的关键。而变形、缩水等质量问题会在评论中被迅速暴露并放大，并且通常会与综合体验关联，造成更严重的负面影响。

3) 文本挖掘的性别差异表明，消费需求存在显著的性别分化规律，需采用差异化策略。对比分析发现，女装消费者决策路径细致且情感化，高度关注颜色、版型、做工等细节与审美元素；而男装消费者决策路径则更为直接，更看重速干、弹性、透气等功能属性和交易效率。

5.2. 建议

1) 商家应主动提供“七天退换”和“正品保障”服务，获取平台认证(如天猫认证)，积极采用视频形式全方位展示商品，尤其注重展示服装的上身效果、面料质感和动态垂感，以弥补线上购物的体验短板，商品描述应该围绕“舒适”、“合身”、“透气”等核心体验展开，给消费者更优质的服务保障。

2) 商家可以聚焦中低价位市场，并通过定期的大力度折扣活动吸引价格敏感型客户。同时，建立严格的品控体系，从根本上杜绝质量问题，降低负面口碑风险。在品类规划上，可适当向女装和秋冬装倾斜资源，以契合市场需求。

3) 对女装提供精细化的尺码信息，如详细尺码、模特指标参考；提供多颜色对比实拍、局部做工特写。对男装突出功能性描述，如速干、冰凉感等，直接标注核心信息，优化购物流程，提供快捷的配送服务。

参考文献

[1] Bhatnagar, A. and Ghose, S. (2004) An Analysis of Frequency and Duration of Search on the Internet. *Journal of*

Business, **77**, 311-330. https://doi.org/10.1086/381277

[2] 雷晶, 李霞. 基于因子分析和聚类分析的市场细分研究——以江苏某电子商务品牌女装为例[J]. 南京邮电大学学报(社会科学版), 2014, 16(4): 49-54.

[3] 崔志超. 基于产品特征的中文评论情感分析系统设计与实现[D]. 石家庄: 河北科技大学, 2025.

[4] Agarap, A.F. (2018) Statistical Analysis on E-Commerce Reviews, with Sentiment Classification Using Bidirectional Recurrent Neural Network (RNN). arXiv: 1805.03687. https://arxiv.org/abs/1805.03687

[5] Engel, J.F., Blackwell, R.D. and Miniard, P.W. (1995) Consumer Behavior. 8th Edition, Harcourt Brace College Publishers.

[6] Delone, W.H. and Mclean, E.R. (2003) The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems*, **19**, 9-30. https://doi.org/10.1080/07421222.2003.11045748

[7] 王开洁, 王明. 电子商务市场长尾现象研究述评与展望[J]. 技术经济与管理研究, 2020(4): 80-86.

[8] Hosmer, D.W., Lemeshow, S. and Sturdivant, R.X. (2013) Applied Logistic Regression. 3rd Edition, John Wiley & Sons, 8-10. https://doi.org/10.1002/9781118548387

[9] 蔡俊娟. Logistic 回归模型分析应用[J]. 长春师范大学学报, 2013, 32(2): 8-10.