

基于消费者反馈数据的 美妆品类营销内容选题预测

——以彩棠的抖音评论数据为例

 答辩学生：侯勇青  指导老师：龙思薇、刘杉

21广告学（计算广告双学士学位复合型人才培养项目）

选题的背景意义

国货美妆成为美妆消费品类的新趋势

产品创新

- 传统成分与现代科技相融合
- 个性化定制彩妆、产品

市场拓展

- 海外成功赢得关注
- 国内进一步扩大市场份额

消费群体变化

- 年轻一代消费者成为国货美妆消费主力军，重视文化认同和社交需求

美妆品牌重视在短视频平台的内容营销

短视频平台影响消费者决策

- 缩短决策路径，激发即时消费
- 社交互动加强消费者决策信任

美妆品牌的短视频内容营销布局

- 内容多元化，达人分层化
- 线上线下融合联动

研究问题

构建选题热度预测模型

以抖音平台的视频评论为原始数据，为彩棠品牌提取消费者的核心关注点，洞察消费者真实需求，并结合数据构建可以用于预测未来可能的营销内容选题热度的模型

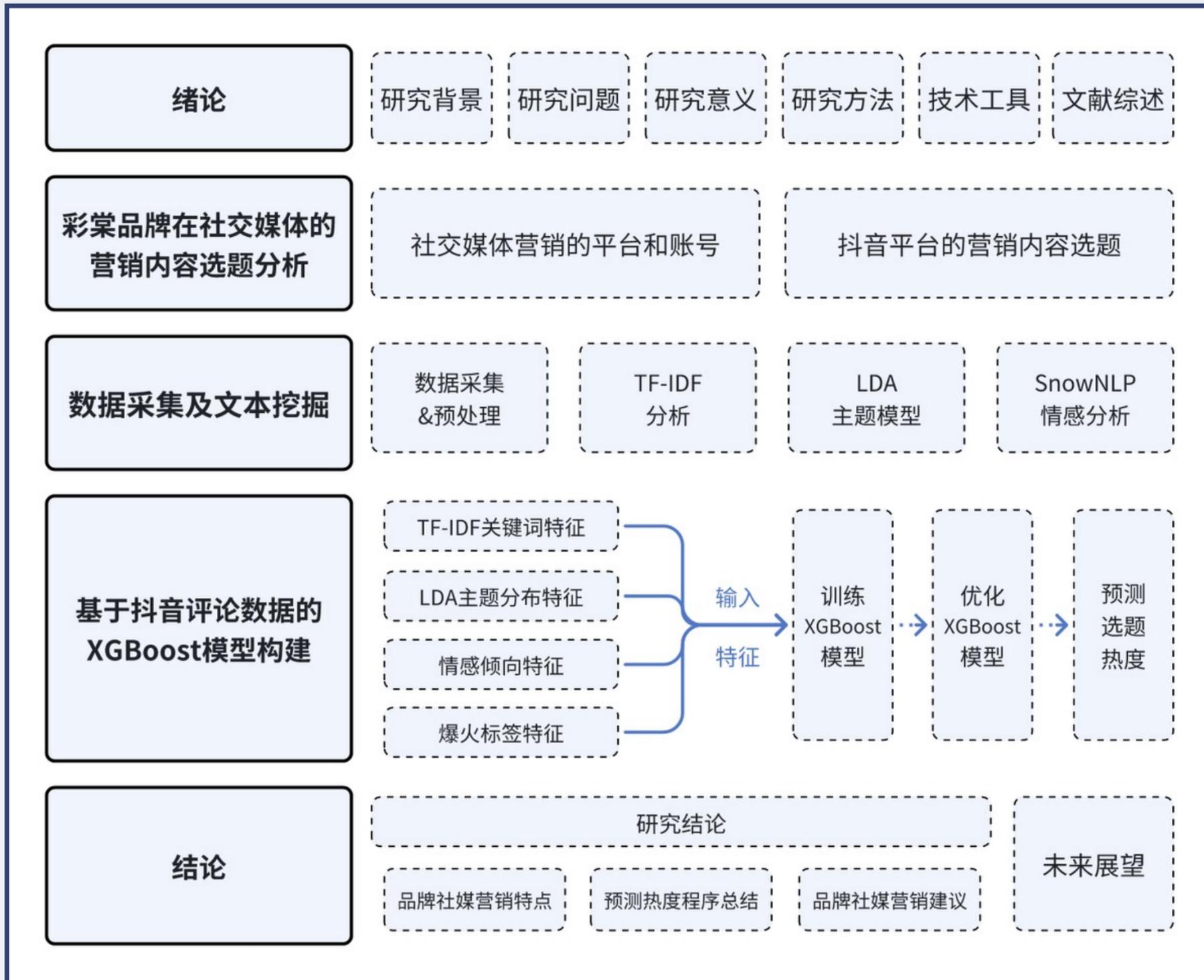
分析

彩棠现有的社交媒体营销内容

研究结论

结合前两个问题的研究成果
为彩棠之后的营销策略提供参考与建议

研究方法及框架



研究方法

- 文本分析法

技术工具

- 数据获取: python爬虫
- 文本挖掘: TF-IDF、LDA、SnowNLP
- 模型构建: XGBoost

研究过程——文本挖掘

01

数据采集&预处理

- 通过“彩棠产品名称”“彩棠化妆教程”等，获取不同时间发布、不同互动数据的视频，通过爬取共获得10811条评论数据；
- 每条数据包含评论昵称、评论时间、评论IP地址、评论内容。

02

TD-IDF关键词

- 计算前20的高频词：彩棠、粉底液、好用、我用、修容、姐姐、不好、推荐等；
- 主要聚焦于品牌的核心产品，情感态度呈两极分化。

03

LDA主题模型

- 挖掘4个主题；
- 从主题关键词看，消费者更关注产品适用肤质、使用体验、社交互动。

04

SnowNLP情感倾向

- 通过将每条评论映射为 0-1的情感得分，并划分为“正面”“中性”“负面”三类情绪；
- 最终获得整体评论的平均情感得分为0.6188，评论用户对品牌内容的整体情绪较为积极。

文本挖掘结果

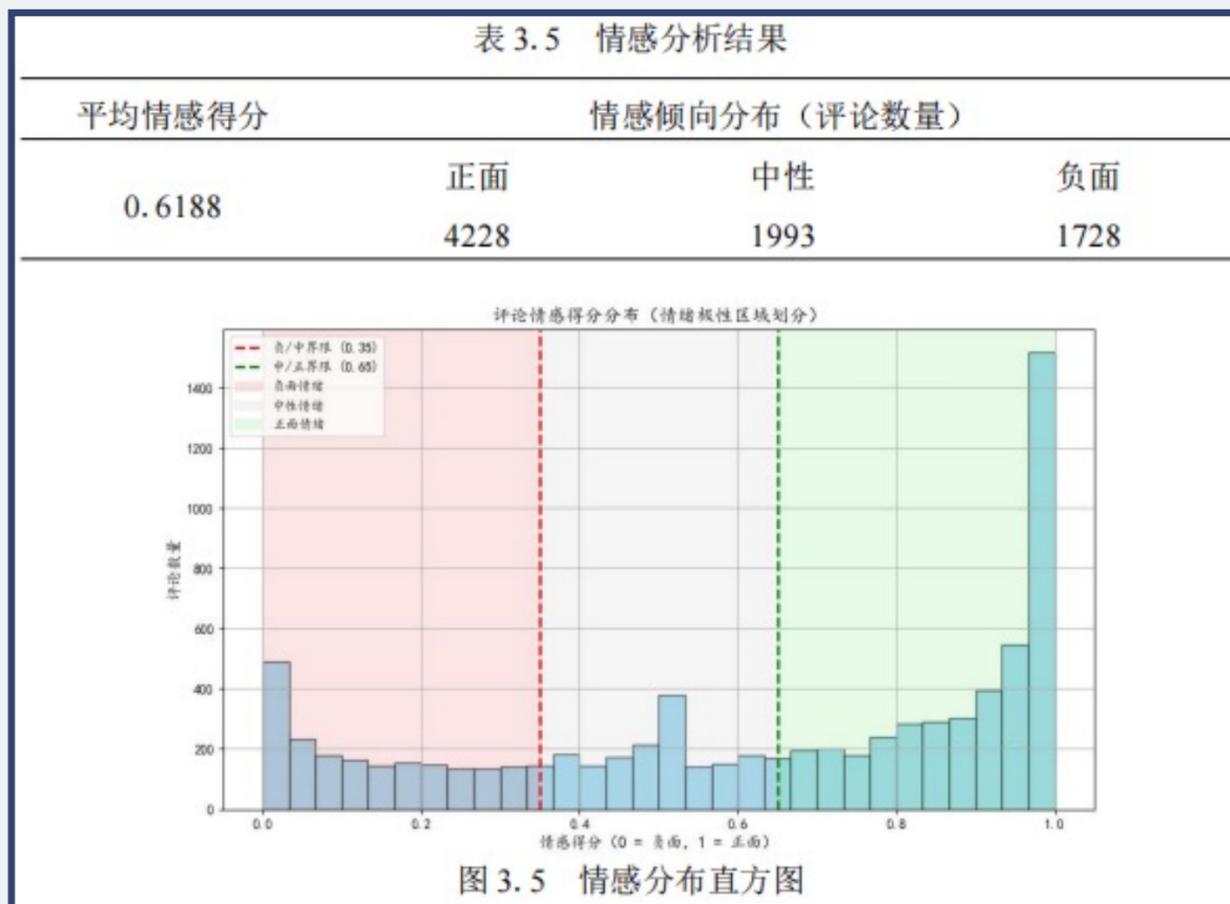
TD-IDF关键词

	word	tfidf
0	彩妆	251.755725
1	粉底液	203.558092
2	好用	131.979365
3	我用	103.322258
4	修容	101.303112
5	姐姐	93.234360
6	不好	89.437852
7	推荐	84.651690
8	粉底	80.126123
9	修容盘	78.771289

10	适合	78.147179
11	腮红	77.627636
12	妆前	77.312908
13	高光	73.976870
14	油皮	71.653557
15	干皮	70.791995
16	气垫	70.239060
17	喜欢	66.428455
18	圆管	64.215752
19	没有	59.628338

SnowNLP情感倾向

表 3.5 情感分析结果



文本挖掘结果

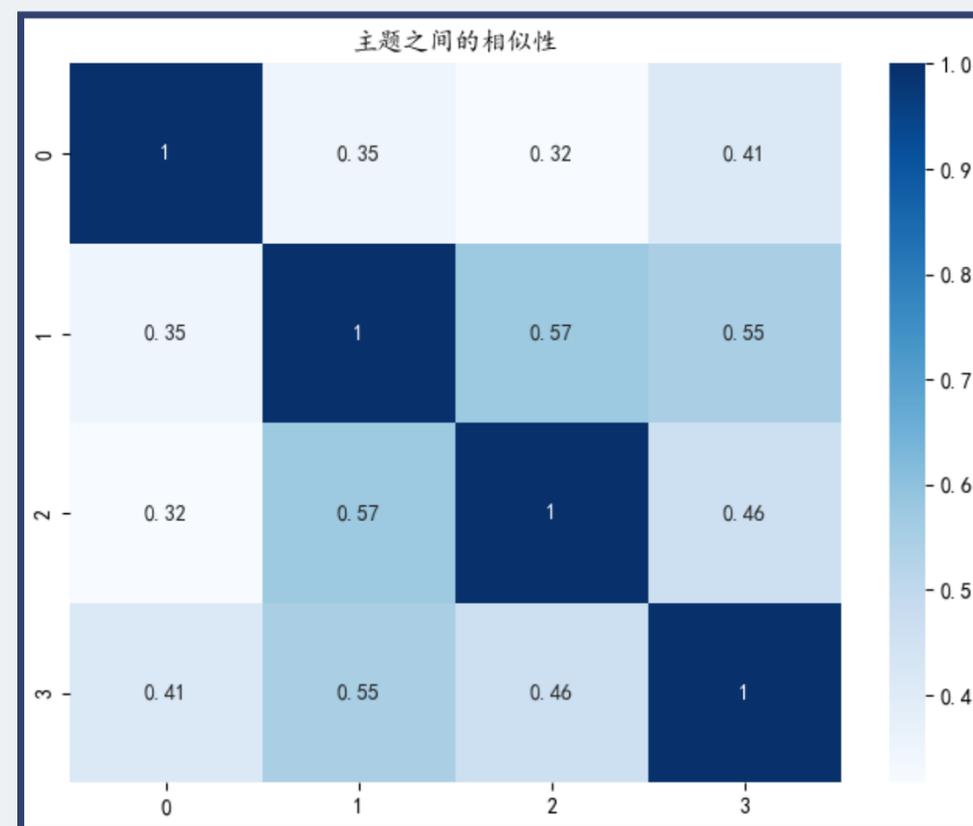
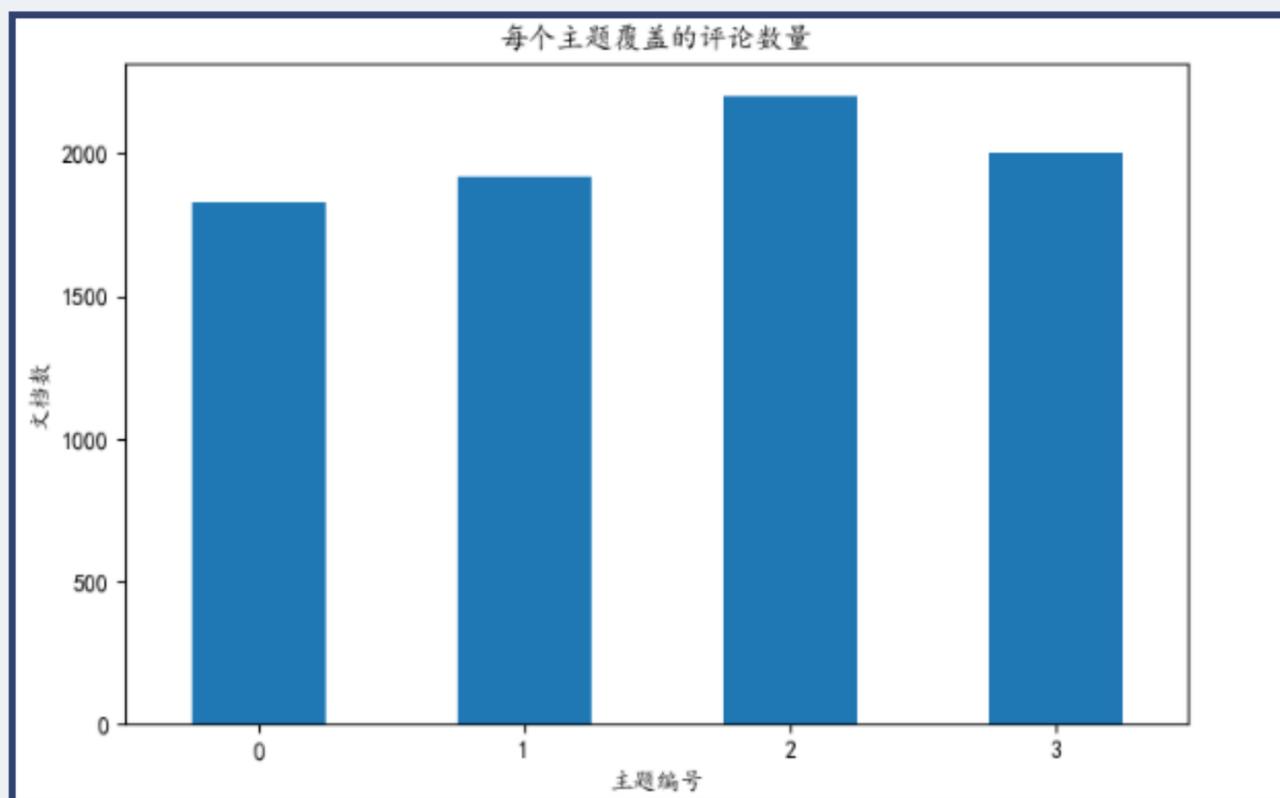
LDA主题模型

Topic #0:
喜欢 姐姐 定妆 喷雾 粉饼 谢谢 脖子 修容 送心 彩棠
41.305 37.93 36.148 32.156 26.012 25.986 22.11 19.395 18.663 18.472

Topic #1:
彩棠 干皮 妆前 粉底液 适合 气垫 暗沉 刚买 修容 确实
94.589 54.437 43.907 41.361 33.118 28.989 28.78 28.692 28.652 27.92

Topic #2:
粉底液 好用 彩棠 我用 圆管 我刚 粉底 油皮 干面 千人
108.56 91.812 76.555 62.129 45.146 41.213 39.961 39.149 31.814 29.19

Topic #3:
高光 彩棠 腮红 不好 修容 推荐 修容盘 粉底液 姐姐 双拼
65.473 62.18 54.417 50.071 47.892 46.525 43.915 40.485 39.005 26.779



研究过程——预测模型

01

构造输入特征

- 对每条评论做TF-IDF、LDA、SnowNLP处理，得到对应输入特征；
- 构建爆火标签，标签评分 = 点赞数 × 0.5 + 评论数 × 0.3 + 分享数 × 0.2，将评分按前30%视频打标签为“爆火 (1)”，其余为“普通 (0)”，将每条评论继承其所属视频的标签，并作为之后模型的输出值。

02

训练并优化XGBoost模型

划分训练集、测试集，从XGBoost库中导入XGBClassifier，用于构建XGBoost模型；
使用scale_pos_weight/GridSearchCV优化模型

03

预测营销内容选题热度

所构建的模型支持对营销内容新选题进行热度预测，
输出其爆火概率与二元分类结果。

研究成果——预测模型

输入特征数据集

每条评论包含关键词、主题分布、情感得分、爆火标签四个维度的特征数据

表 3.6 输入特征数据示例

评论内容	情感得分	关键词 1 (好用)	...	主题 0	标签 (是否爆火)
我觉得粉底液挺好用的，已经用完了两瓶	0.8311	0.7950		0.6249	1
包难用的，一百多买了一把水枪，我请问呢	0.0877	0		0.25	0

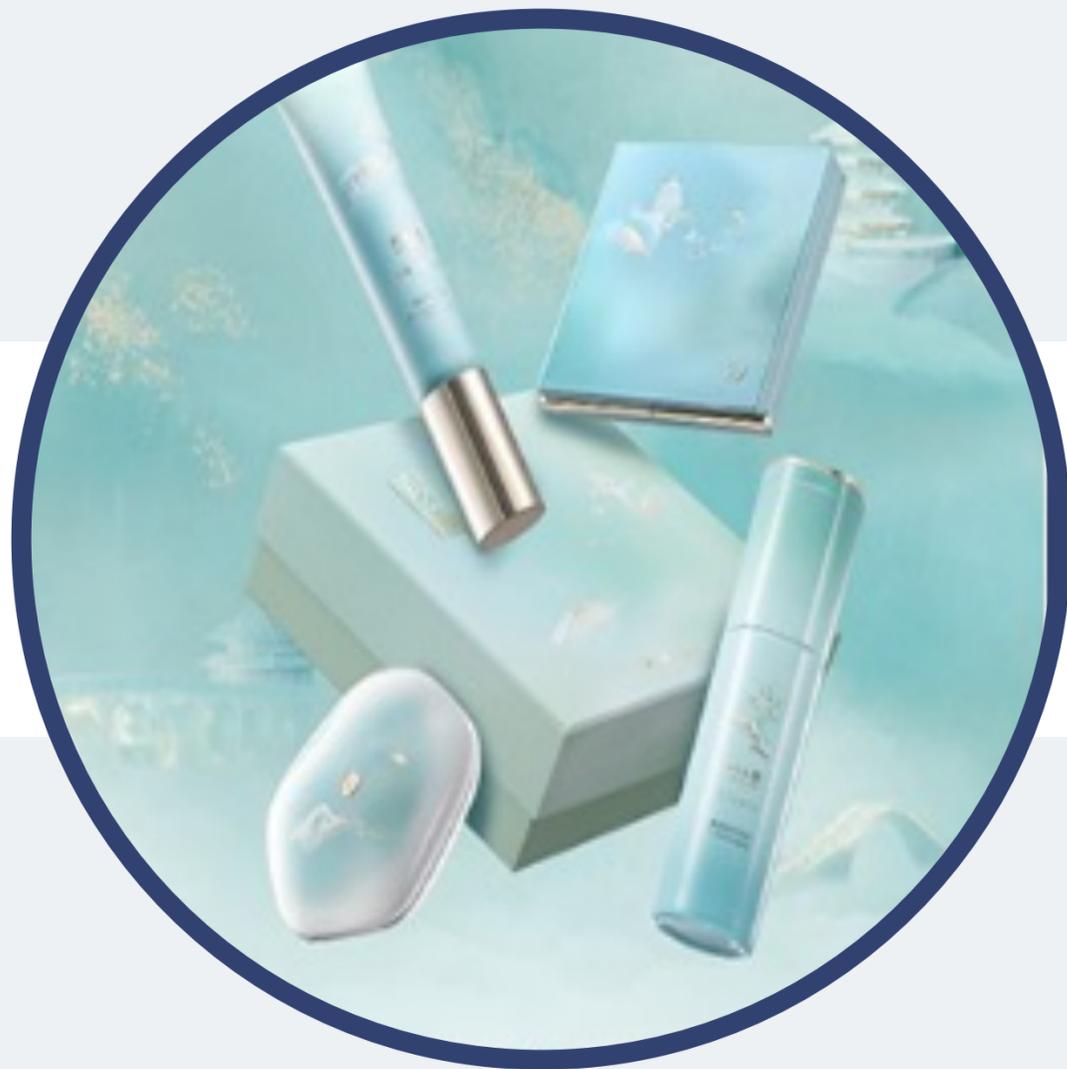
新营销选题预测

假设现有3个方向各不相同的新营销选题，将其放入模型中，判断是否能成为爆火内容

表 4.3 新选题预测结果

选题内容	是否爆火	爆火的概率
24 节气妆容图鉴——以传统时序重塑现代妆容美学	爆火	0.7458
新款粉底液测评合集，油皮干皮全覆盖	爆火	0.7800
彩妆的哲学：色彩与存在的辩证关系	未爆火	0.1584

研究结论



当前彩棠品牌在社交媒体的内容营销特点

- 多平台矩阵布局，以抖音为核心
- 三维价值策略助力品牌营销——专业产品+文化叙事+情感共鸣



内容选题预测程序总结

通过数据采集、文本挖掘、模型构建，了解消费者在评论中提及的重点内容，挖掘评论数据中的潜在主题，深入剖析消费者对彩棠品牌及产品的关注点，并分析情感倾向，最终实现营销内容选题的热度预测。



彩棠品牌社交媒体内容营销的优化建议

- 聚焦热门品类，通过数据实现精准的内容分发
- 打造实用内容，分层满足功能与情感需求
- 增强社交互动，建立完善的消费者反馈机制

恳请各位老师批评指正

THANK YOU FOR WATCHING

 答辩学生：侯勇青  指导老师：龙思薇、刘杉

21广告学（计算广告双学士学位复合型人才培养项目）