

基于 LDA 模型的旅游住宿接待能力评价

——以济南市为例

张莹莹, 陈恒宇, 张梦迪

(山东大学中泰证券金融研究院, 济南 250100)

摘要: 对济南市旅游住宿设施进行了评价与需求预测研究, 利用 Python 爬虫技术, 从携程、去哪儿网和美团等平台抓取数据。应用 LDA(latent Dirichlet allocation)主题模型揭示文本数据中的潜在主题结构, 从而有效分类住宿地标签。包括 XGBoost、CatBoost、LightGBM 和随机森林(random forest)在内的机器学习方法被用来筛选影响住宿地评分的核心指标。通过结合层次分析法(AHP)和熵权法来构建评估体系, 以确定权重。通过可视化分析, 揭示不同主题类别、住宿地类型、区域类型以及交通、商业区、学校和景点的接近程度在接待能力上的差异。基于酒店主题、评分分布、地理位置、评论数量和价格提出优化建议, 旨在提高服务质量和监管, 从而增加接待能力和用户满意度。

关键词: LDA 模型; 机器学习; 层次分析法; 熵权法; python 爬虫

中图分类号: F719.2; F590.8 **文献标志码:** A **文章编号:** 1671-1807(2025)06-0204-11

随着中国步入大众旅游的新时代, 2016 年政府工作报告^[1]的提出标志着旅游产业的重要转型。在这一背景下, 国内旅游市场在疫情后迎来了显著的复苏, 国民出游的热潮推动了对旅游住宿设施需求的激增。消费者对旅游体验质量的日益关注, 使得住宿业不仅需求增长, 而且对服务质量的要求也不断提升。

济南市, 作为山东省的省会城市, 不仅是政治、经济和文化的重要枢纽, 更凭借其丰富的旅游资源和便捷的交通网络, 成为备受欢迎的旅游目的地。根据统计局的最新统计公报^[2], 2023 年济南市接待国内游客数量达到了 10 631.1 万人次, 同比增长 61.9%; 国内旅游收入高达 1 132.9 亿元, 增长了 65.5%, 这些数据充分展示了济南市旅游业的蓬勃发展。

随着旅游热潮的兴起, 学术界对旅游服务管理行业的研究正日益增多, 众多学者将文本挖掘技术应用到旅游领域, 如识别产品销量影响因素、研究游客满意度、构建游客的评价指标体系等。范宁^[3]基于在线旅游网站评论, 分析了顾客关于民宿满意度的影响因素。李爽等^[4]利用携程网旅游评论数据, 挖掘了西安和成都两个城市游客的满意度。范

博雷^[5]通过文本挖掘技术对提高国家公园旅游满意度水平进行研究。刘阳^[6]依托飞猪网中东南亚旅游产品的信息, 据此分析了消费者对产品的真实所需, 为商家提供了更加有效的增加产品销售量及竞争力的措施。郭佳琪和胡彬^[7]以江苏省南京市的网红景区为例, 基于扎根理论构建游客满意度评价模型, 并通过高频词汇和网络语义分析各因素的影响程度, 从“旅游吸引物”“人文景观”“网红特质”和“配套服务设施”等维度评估游客满意度。

文献显示, 旅游住宿服务质量管理愈发受到广大学者的关注, 而能够衡量住宿服务质量好坏的服务质量评价指标体系构建问题渐渐地成为学者关注的重点。杨瑜^[8]在服务评价的多学科多研究方向的基础上, 通过对现有的顾客满意度评价体系进行精细的有针对性的改进, 构建了新的住宿顾客满意度测评模型并进行了实证研究。陈玉莲^[9]基于大数据挖掘技术, 从住宿条件、周围环境、服务质量以及个性化体验等维度, 研究肇庆市酒店和民宿旅游住宿的游客体验差异及其对满意度的影响。翟姗姗等^[10]基于价值共创理论, 利用爱彼迎平台北京市的房东和房客生成内容数据, 从文本资源的视角, 探讨共享住宿平台信息服务价值共创的影响机制。

收稿日期: 2024-10-08

作者简介: 张莹莹(2002—), 女, 山东德州人, 研究方向为金融统计; 陈恒宇(2001—), 男, 浙江金华人, 研究方向为金融统计; 张梦迪(2001—), 女, 山东烟台人, 研究方向为金融统计。

关于利用文本情感分析技术建立旅游相关评价指标体系,司育^[11]基于 LDA 模型,利用国内三大旅游资讯平台上的山西热门景区在线评论数据,从景区维护、游玩体验、人文历史感知等九大维度,构建旅游目的地评价体系,并对山西省十大热门景区进行综合性评价和满意度动态分析。董爽和汪秋菊^[12]基于隐含狄利克雷分布(LDA)主题模型和自然语言处理技术,利用旅游网站游客在线评论信息,从游客感知的角度识别了国家矿山公园的游客感知维度及其构成因子。

基于以上讨论,利用线上预订平台和企业信息查询软件的住宿设施信息,运用机器学习算法,如 XGBoost、CATBoost、LightGBM 和随机森林,结合 LDA(latent Dirichlet allocation)主题模型进行文本分析,通过层次分析法(analytic hierarchy process, AHP)和熵权法来确定评价指标的权重,构建一个全面的评价体系,以科学地评估和预测旅游住宿地的接待能力。

1 理论基础

1.1 LDA 主题模型

LDA 主题模型是一种文本挖掘中的主题识别算法。它通过以下步骤工作:首先,对文档进行清洗、分词、去除停用词等预处理操作,构建词袋模型;然后,初始化每个文档的主题分布和每个主题的词分布;其次,迭代更新文档的主题分布和主题的词分布,直到模型收敛;最后,从模型中提取主题,通常是基于每个主题中最具代表性的词汇。

LDA 模型的核心在于它假设文档是由多个主题的混合生成的,每个主题则是由一系列词汇的概率分布定义的。模型通过贝叶斯统计方法,使用狄利克雷分布作为先验分布,来推断文档中的主题分布和主题下的词分布。这种方法允许模型在面对新文档时,也能够有效地预测其主题分布。

LDA 模型广泛应用于文本分类、信息检索、推荐系统等领域,它能够帮助我们理解大量文本数据中的主题结构,从而为进一步的分析和应用提供支持。在实际应用中,LDA 模型的训练和参数估计通常使用成熟的库,如 Python 中的 gensim 库来进行。通过这些库,可以方便地对文档集合进行主题建模,提取主题关键词,并对新文档进行主题预测。

1.2 机器学习模型

XGBoost 是一种基于梯度提升决策树的优化算法,它通过引入正则化项有效控制模型复杂度,减少过拟合。XGBoost 支持并行处理,提高了大规

模数据集的处理效率。它的优势在于能够自动处理各种类型的数据,包括稀疏数据,并且可以自定义优化目标和评估准则,使其在许多机器学习竞赛中表现突出。

CATBoost 是一个高效的梯度提升库,专门优化了类别特征的处理。它能够自动处理缺失值,无需额外的数据预处理。CATBoost 通过对称性原理和基于统计测试的特征选择,提高了模型的泛化能力。此外,它支持多类别分类和回归任务,使得 CATBoost 在处理具有大量类别特征的数据集时尤为有效。

LightGBM 是一种基于梯度提升框架的高效机器学习算法,它采用基于直方图的算法来优化训练速度和内存使用。LightGBM 支持大规模数据和高维度特征,同时保持了高效的训练速度。它还支持类别特征的原生处理,无需额外的转换,使得模型训练过程更加简洁。此外,LightGBM 提供了灵活的配置选项,包括各种损失函数和评估指标,适用于各种预测任务。

随机森林是一种集成学习方法,通过构建多个决策树并将其预测结果进行汇总,以提高模型的准确性和稳定性。每棵树都在数据集的一个随机子集上进行训练,这增加了模型的多样性,从而减少过拟合的风险。随机森林能够提供特征重要性的评估,这对于理解模型的决策过程和特征选择非常有用。它适用于分类和回归任务,并且在处理高维数据时表现出色。

将这些机器学习模型综合应用构建指标体系,数据经过预处理、特征工程、特征选择与评估,以及使用 4 种不同的机器学习算法(random forest、LightGBM、CATBoost、XGBoost)进行模型训练与验证。最后,根据模型性能评估结果,进行下一步操作。

1.3 层级结构划分法

层级结构划分法是一种组织和管理复杂信息的方法,它通过将信息分解成不同的层级和子层级来简化理解和分析。陈亮^[13]基于旅游客源地的空间距离、经济发展水平、人口规模、旅游吸引力和交通水平等因素,运用层次结构划分法,将旅游客源节点从旅游流量、时间距离和城镇居民人均年可支配收入 3 个维度进行分类研究。在构建旅游住宿服务质量的评价指标体系时,这种方法尤其有效,因为它允许将广泛的服务质量概念细化为可操作和可衡量的指标。如图 1 所示为运用层级结构划分法构建旅游住宿服务质量的评价指标体系。

1.4 层次分析法(AHP)

AHP是一种多准则决策方法,通过将决策问题分解为目标、准则和方案3个层次,构建判断矩阵并计算权重。何文宏和赵乾卉^[14]基于AHP建立城市文化街区活力评价指标体系,目标层为城市文化街区活力评价指标,准则层为空间规划度、设施建设度、文化承载度、社会交往度、商业活力度,方案层划分了14个子指标。高原^[15]基于KANO-AHP构建包含27个指标的京津冀旅游文创产品需求评价体系,利用AHP得出各指标综合权重,为后续的文创设计与改进提供了量化依据。

1.5 熵权法

熵权法是一种基于信息熵原理的权重确定方法,它通过计算各指标的熵值来衡量信息的有效性和差异性。在多指标决策分析中,熵权法能够客观地反映各指标之间的相对重要性,从而为决策者提供科学的依据。郭悦和白祥^[16]利用熵权法将上村寺乡村旅游发展绩效评估体系进行了权重的计算。李明峰^[17]以鼓浪屿5A级景区的旅游解说系统为研究对象,运用熵权法分析自导式解说、向导式解说和综合式解说3个关键因素,构建评价指标体系,并进行多级模糊评价。

2 构建旅游住宿接待能力评价指标体系

2.1 数据来源

选择济南市作为研究对象,在线上预订平台(如携程网、去哪儿网、飞猪旅行、同程旅游、途牛、美团、各大酒店官方网站等)和企业信息查询软件(如企查查、天眼查)中提取包含住宿设施名称、地址、经纬度、评价分数、房间类型等信息的记录。

基于Python爬虫技术^[18],经过数据爬取-数据去重-数据清洗-数据生成(图2)的流程后,最终生成超过3500条济南市旅游住宿地数据,指标涵盖评分、评分描述、客户评价、服务、环境、卫生、设施、星级类型、星级、距离市中心的公里数、是否靠近交通、是否靠近商圈、是否靠近学校、是否靠近景点与价格,基本满足了数据的全面覆盖、完整性且无重复。

2.2 构建指标体系

在构建住宿地接待能力指标体系表时,采用层次分析法(AHP)将决策问题分解为更小的部分:一级为目标层(A),即住宿接待能力;二级为指数层(B);三级为指标层(C);为了保证指标体系更为科学性,同时采用层级结构划分的方法,从宏观到微观

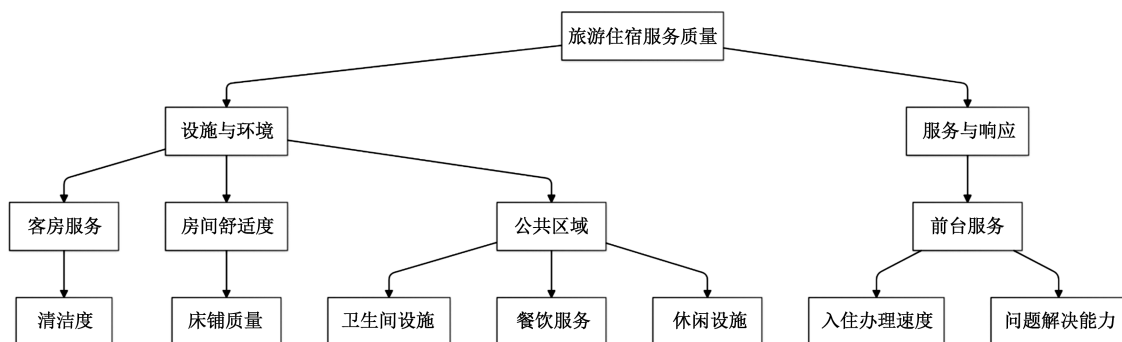


图1 层级结构划分法(以构建旅游住宿服务质量的评价指标体系为例)

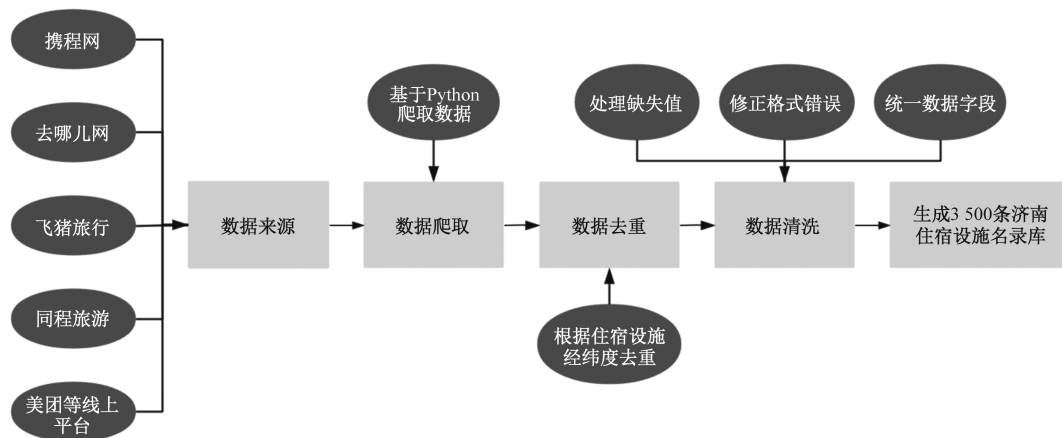


图2 济南市的旅游住宿地数据生成流程

逐级细化。一级指标包括总体评价、服务质量、环境与设施、酒店等级、地理位置与便利性以及价格与费用。

在这些指标下,进一步细化了三级指标,如评分、服务、环境、星级、地理位置和价格等,以深入理解客户体验。最终,济南市住宿地接待能力指标体系表包括 6 个一级指标、15 个二级指标。表 1 展示了济南市旅游住宿地接待能力指标体系。这种系统的评价体系不仅为住宿业者提供了改进的方向,也为消费者提供选择住宿的参考标准,促进旅游住宿业的整体发展。

表 1 住宿地接待能力指标体系

目标层	指数层	指标层	中文解释
住宿地接待能力	总体评价 (B ₁)	score(C ₁₁)	评分
		score_desc(C ₁₂)	评分描述
		comments(C ₁₃)	客户评价
	服务质量 (B ₂)	service(C ₂₁)	服务
	环境与设施 (B ₃)	environment(C ₃₁)	环境
		hygiene(C ₃₂)	卫生
		facilities(C ₃₃)	设施
	酒店等级 (B ₄)	star_type(C ₄₁)	星级类型
		star(C ₄₂)	星级
	地理位置与便利性 (B ₅)	POI(C ₅₁)	距离市中心的公里数
		near_traffic(C ₅₂)	是否靠近交通
		near_shop0(C ₅₃)	是否靠近商圈
		near_school(C ₅₄)	是否靠近学校
		near_scenic(C ₅₅)	是否靠近景点
	价格与费用 (B ₆)	price(C ₆₁)	价格

3 住宿地文本分类建模与分析

3.1 基于 LDA 的住宿地标签的分类主题模型

3.1.1 最优 LDA 主题数的确定

在 LDA 主题模型中,困惑度和一致性是常用的评价指标,用于衡量模型的性能。

(1) 困惑度。

困惑度(Perplexity)是衡量语言模型好坏的标准之一,用来评估 LDA 模型在新文本上的表现。它反映了模型对数据的拟合程度,困惑度越低,模型越好。LDA 中的困惑度公式为

$$\text{Perplexity}(D) = \exp \left[-\frac{\sum_{d=1}^M \ln P(w_d)}{\sum_{d=1}^M N_d} \right] \quad (1)$$

式中: D 为文档集; M 为文档数; w_d 为第 d 个文档的词序列; $P(w_d)$ 为根据 LDA 模型生成第 d 个文档的概率; N_d 为第 d 个文档中的词数。

(2) 一致性。

主题一致性(Coherence)是另一种衡量 LDA 主

题模型质量的指标,反映主题词之间的语义相关性。常见的主题一致性衡量标准之一是基于对主题词对共现关系的统计。其基本公式为

$$C_w(V^{(k)}) = \sum_{m=2}^M \sum_{l=1}^{m-1} \ln \frac{D(v_m^{(k)}, v_l^{(k)}) + \epsilon}{D(v_l^{(k)})} \quad (2)$$

式中: $V^{(k)} = (v_1^{(k)}, v_2^{(k)}, \dots, v_M^{(k)})$ 为第 k 个主题的前 M 个高频词; $D(v_m^{(k)})$ 为包含单词 $v_m^{(k)}$ 的文档数; $D(v_m^{(k)}, v_l^{(k)})$ 为同时包含单词 $v_m^{(k)}$ 和 $v_l^{(k)}$ 的文档数; ϵ 为平滑项,通常取一个非常小的值以避免分母为零。一致性越高,说明主题中的高频词语之间的语义相关性越强,主题的可解释性更好。

gensim 模块计算的困惑度为单词的困惑度,困惑度越小,说明 LDA 模型的主题挖掘效果越好。对于一致性而言,一致性越高,说明 LDA 模型的主题挖掘效果较好。基于不同主题数所构建的 LDA 主题模型,其一致性与困惑度如图 3 所示。

当主题个数为 3、4、5 的时候,其一致性得分相差不大。此外,当主题数为 5 的时候,困惑度的下降速度明显趋缓。为此,选择主题个数为 5 进行 LDA 主题模型的构建。

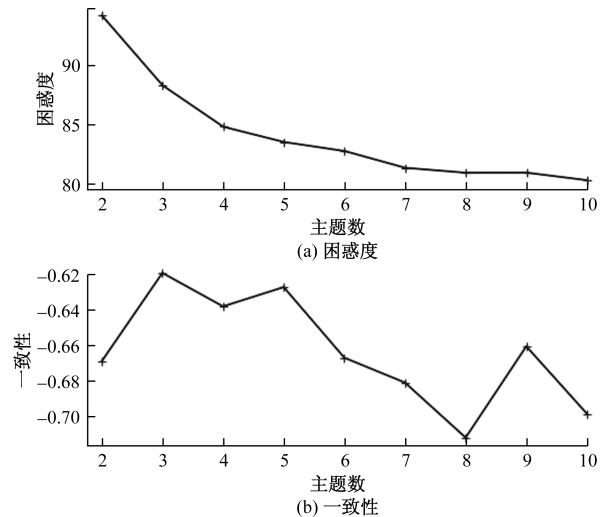


图 3 LDA 主题模型的一致性与困惑度

3.1.2 构建住宿地标签的 LDA 主题分类模型

围绕 3 531 个居住地数据,选取主题个数为 5 的情况构建 LDA 主题模型,可以得到如图 4 所示中的 LDA 模型主题分布。图 4 左侧利用 LDA 主题模型展示了 5 个主题的分布及其相互关系。边际分布图揭示了各主题在文档中的频率,指示了它们的普及和重要程度。多维缩放技术生成的主题间距离映射则展现了主题的相似性与差异性,其中主题间距离的远近代表了它们之间的相似度。图 4 右侧展示了通过 LDA 主题模型分析得到的旅游住宿

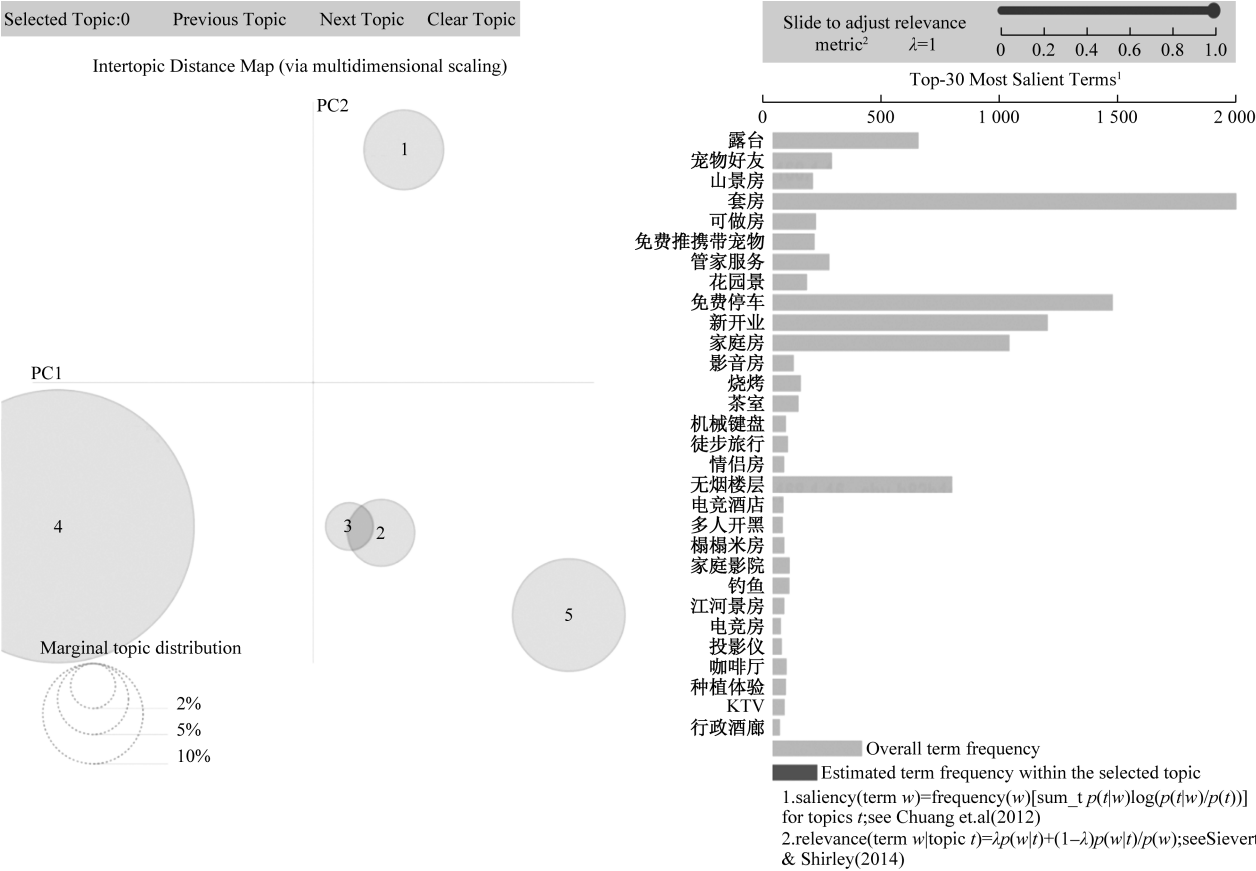


图 4 济南市各住宿地标签的主题可视化分布

地文本数据中的关键词显著性。图中列出了与特定主题最相关的前 30 个词汇,这些词汇的显著性是通过计算词汇频率和词汇在特定主题下的条件概率来确定的。词汇的显著性越高,表示它与主题的关联性越强。

利用 LDA 主题模型对济南市各住宿地标签主题类型进行分类,共识别出 5 个核心主题,每个主题都由其高频词所定义。提取每个主题中相对应的前 4 名特征词,形成了济南市各住宿地标签主题类型分类(表 2)。

主题 1“宠物友好与休闲娱乐结合的住宿类型”包含宠物友好、可做饭、免费携带宠物、影音房等高频词,主题 2“享受自然景观和户外活动的生态型住宿”包含山景房、花园景、徒步旅行、榻榻米房等高频词,主题 3“针对年轻群体的电竞与娱乐主题住宿”包含机械键盘、情侣房、电竞酒店、多人开黑等高频词,主题 4“商务与家庭出行的高档舒适型住宿”包含套房、免费停车、新开业、家庭房等高频词,主题 5“度假与亲子娱乐为主的特色住宿”包含露台、管家服务、烧烤、茶室等高频词。

这 5 个主题反映济南市住宿地的不同特色和服

表 2 济南市各住宿地标签主题类型

序号	主题类型	高频词
1	宠物友好与休闲娱乐结合的住宿类型	宠物友好,可做饭,免费携带宠物,影音房
2	享受自然景观和户外活动的生态型住宿	山景房,花园景,徒步旅行,榻榻米房
3	针对年轻群体的电竞与娱乐主题住宿	机械键盘,情侣房,电竞酒店,多人开黑
4	商务与家庭出行的高档舒适型住宿	套房,免费停车,新开业,家庭房
5	度假与亲子娱乐为主的特色住宿	露台,管家服务,烧烤,茶室

务,从宠物友好型到自然生态型,再到电竞娱乐和高档舒适型,以及亲子娱乐特色,为不同需求的游客提供了多样化的选择。

3.2 基于机器学习的住宿地类别分析的重要特征筛选

基于机器学习领域的 4 种先进算法——XG-Boost、CatBoost、LightGBM 和随机森林对济南市住宿接待场所的数据进行分析。本文的目标是通过特征选择和分类评价,揭示影响住宿地接待能力的关键因素。在特征选择阶段,利用这些算法的内

置功能来识别最能代表住宿地接待能力的特征。

通过对比不同算法的特征重要性排名,能够洞察每种算法在处理此类数据时的独特视角和优势。XGBoost 以其出色的预测精度和对不平衡数据集的鲁棒性脱颖而出;CatBoost 在处理类别特征和缺失值方面展现了其优势;而 LightGBM 则以其高效的训练速度和较低的内存使用率在大规模数据集上表现优异;随机森林算法则通过集成多个决策树提供了稳定性和准确性的良好平衡。

通过分析不同模型对特征重要性的评估(图 5),可以得出一些关键结论。特征“评分描述”(C₁₂)在

XGBoost 和 CatBoost 模型中显示出极高的重要性,这表明它对分类结果有显著影响。此外,“客户评价”(C₁₃)和“价格”(C₆₁)在所有模型中都具有较高的重要性,说明这两个因素在预测中扮演着核心角色。特别地,对于“距离市中心的公里数”(C₅₁),LightGBM 模型赋予它极高的重要性,这可能意味着 LightGBM 在处理地理位置信息方面更为敏感。总体来看,“评分描述”在各个模型中都极为重要,而“评论数”和“价格”在多数模型中也显示较高的影响力。这些发现提供了深入理解模型如何根据特定特征做出决策的视角,并有助于优化模型性能。

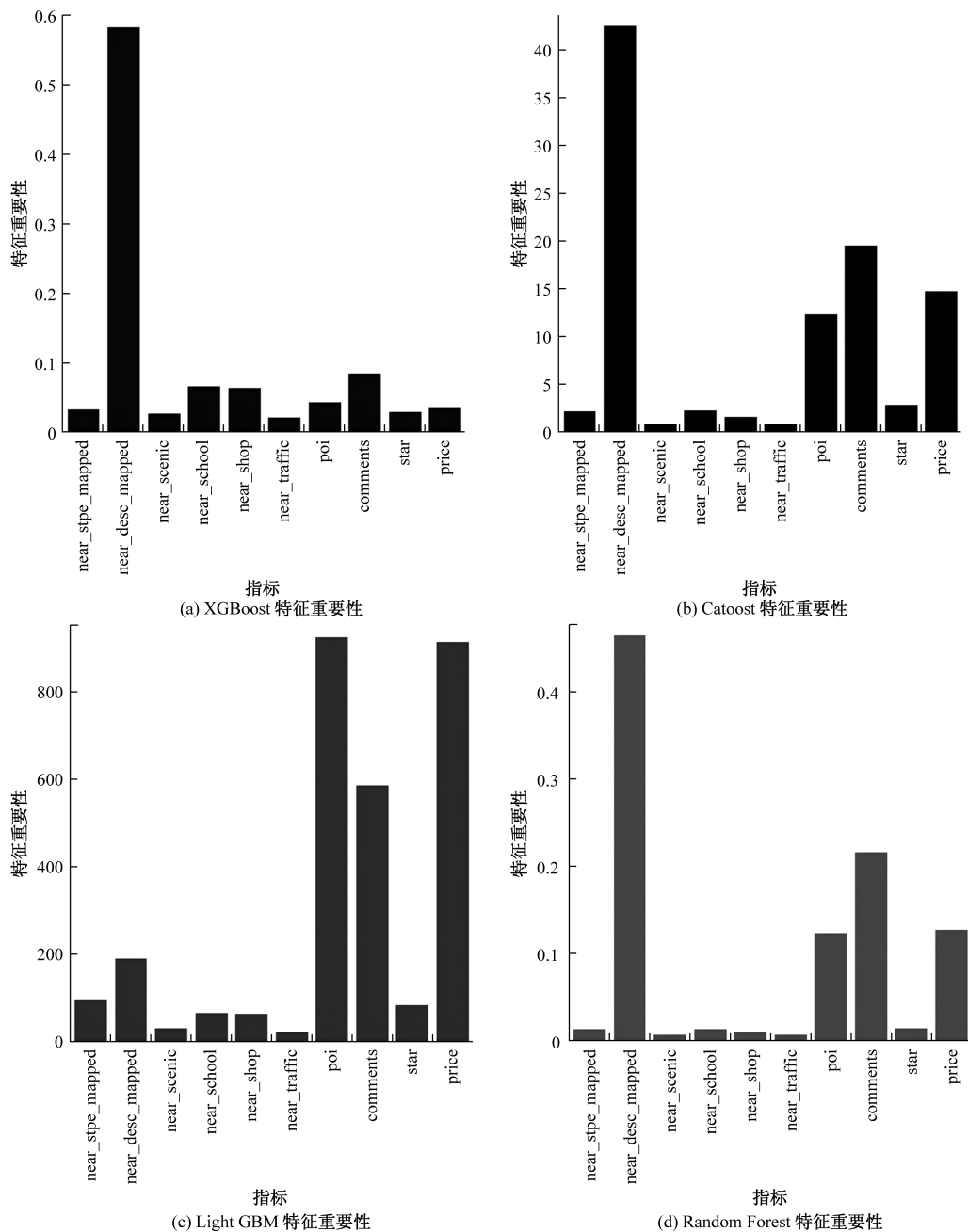


图 5 模型重要特征筛选结果示意图

3.3 济南住宿地接待能力综合评价计算

为了确定住宿地评价指标的权重,采用层次分析法和熵权法两种方法。层次分析法通过构建评价指标的层次结构模型,利用专家打分和成对比较的方式,确定各指标的相对重要性。熵权法则是基于信息熵原理,通过计算各指标的熵值来反映其信息的有效性和差异性,从而客观地确定权重。将客户评论、价格与距离市中心的距离、交通便利性与学校邻近度、酒店星级及其类型映射,以及环境、卫生、服务和设施等进行权重分析,以确保评价体系的科学性和合理性。

为了获得更为科学、客观的指标权重,将层次分析法(AHP)和熵权法计算的权重结果进行综合。具体方法是对两种方法所得的权重进行算术平均(表 3),这样既结合了 AHP 的主观判断,又考虑了熵权法的客观数据分布特点,从而得到更为平衡和合理的权重。通过这种融合,可以提高指标权重的可靠性和有效性,准确评估住宿地的接待能力。

表 3 层次分析法与熵权法的平均权重

指标名称	平均权重
score	0.003 2
Environment	0.003 5
Hygiene	0.003 4
Service	0.002 9
Facility	0.003 7
price	0.133 6
star	0.029 9
comments	0.238 1
poi	0.136 8
near_traffic	0.091 0
near_shop	0.057 2
near_school	0.096 9
near_scenic	0.083 4
score_desc_mapped	0.081 0
star_type_mapped	0.035 5

3.4 住宿接待能力评价结果分析

3.4.1 各指标与综合评价得分的关系分析

可以对住宿地的综合评价得分与其相关因素之间的关系进行分析(图 6~图 9)。首先,经纬度与综合评价得分的关系图显示,住宿地的地理位置,尤其是经度,可能对其评价得分有一定影响。例如,位于特定经度范围内的住宿地可能因为接近某些旅游景点或商业区而获得更高的评分。

其次,兴趣点(POI)^①与综合评价的关系图揭示

①POI,即“point of interest”,中文意为“兴趣点”。在地理信息系统(GIS)中,POI 指的是地图上的点类数据,它们通常与人们的日常活动密切相关,如餐馆、超市、银行、公园等。这些点数据不仅包含名称、地址、坐标等基本信息,还可能包括类别、电话号码、网站链接等更详细的属性。

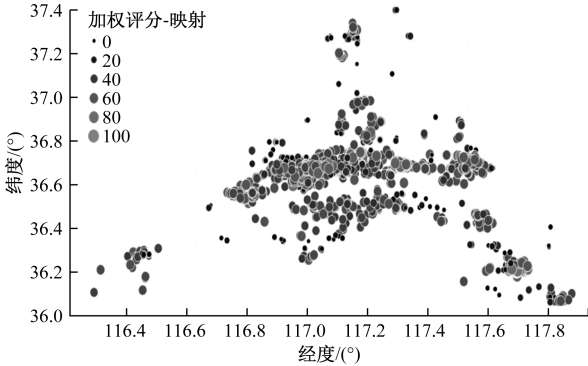


图 6 经纬度与综合评价得分的关系

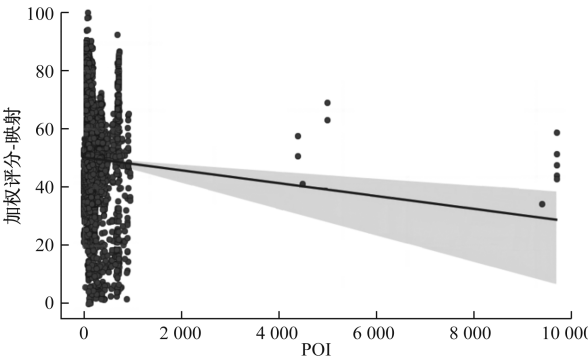


图 7 POI 与综合评价得分的关系

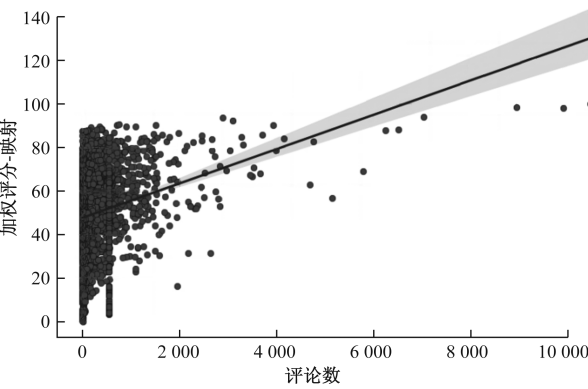


图 8 评论数与综合评价得分的关系

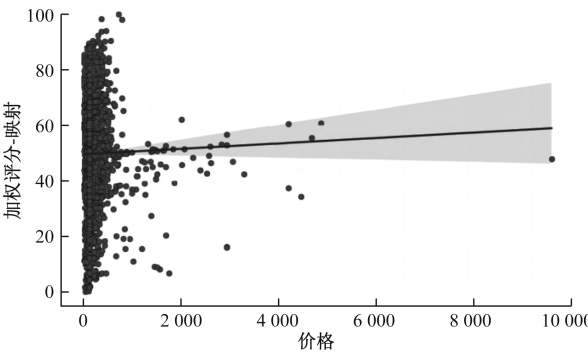


图 9 价格与综合评价得分的关系

了住宿地附近的景点、餐饮和娱乐设施对游客选择和满意度的影响。住宿地周边丰富的 POI 可能会提高其吸引力,从而提升综合评价得分。这表明,为了提高住宿地的评价,经营者可以考虑增加或改善周边的便利设施和服务。

再次,评论数与综合评价的关系图表明,顾客的在线评论数量与住宿地的综合评价得分之间存在一定的正相关性。这可能是因为高评论数反映住宿地的受欢迎程度和顾客的参与度,同时也意味着更多的顾客愿意分享他们的住宿体验。住宿地可以通过积极的客户服务和优质的住宿体验来鼓励顾客留下正面评论,从而提高其综合评价。

最后,价格与综合评价的关系图显示,价格可能是影响顾客满意度的一个关键因素。一般来说,价格合理的住宿地可能会获得更高的评价,因为顾客可能期望物有所值的服务。然而,价格并非唯一决定因素,高质量的服务和设施同样重要。住宿地应该寻求在保持价格竞争力的同时,提供超出顾客期望的服务和体验。

3.4.2 基于不同区域下的综合评价均值分析

图 10 展示了不同区域的住宿地接待能力的综合评价均值。洪家楼/山东大学:平均得分最高,接近 70 分,表明该区域的住宿地接待能力评价最高。大学周边可能由于环境优美、设施完善而受到游客青睐。大明湖/趵突泉:平均得分次高,大约 65 分,这些都是济南的知名景区,说明景区附近的住宿地接待能力较好。济南西站/山博国际会展中心:约 60 分,说明该交通便利区域的住宿地受到较高评

价。英雄山千佛山/舜耕会展中心区:平均得分接近 55 分,显示此区域的住宿地接待能力中等偏上。火车站/长途车站:接近 50 分,得分中等,表明火车站周围的住宿地评价较为中肯,可能由于交通便利,但服务和设施一般。西客站:平均得分略高于 40 分,但仍属于低评分区域,住宿质量和服务有待提升。

3.4.3 基于交通、商圈、学校、景区分类的住宿地接待能力差异分析

图 11 展示靠近交通枢纽的住宿地接待能力显著优于不靠近交通枢纽的住宿地。交通便利性提升了用户的满意度,带来了较高的评分。靠近商圈的住宿地评分明显高且分布集中,证实商圈的便利性对住宿地的接待能力有显著提升作用。学校附近的住宿地得分较高且分布均匀,显示出较强的接待能力,尤其适合学生和家庭游客,这类位置的住宿地对顾客满意度提升效果显著。靠近景区的住宿地评分偏高且相对集中,反映其接待能力较强。这些住宿地因位置优越,通常能够吸引更多游客,提升整体评价。

4 提升济南市住宿接待能力的对策建议

4.1 根据济南市住宿接待能力的综合评分分布情况提出的建议

济南市文化旅游局应重点关注万象新天体育中心和南部风景区的住宿服务质量提升,这两个区域的综合评价均值低于 40 分。对于这些低分区域,建议提供专项基金支持,用于基础设施改造和引入智能服务技术,以提升服务质量。

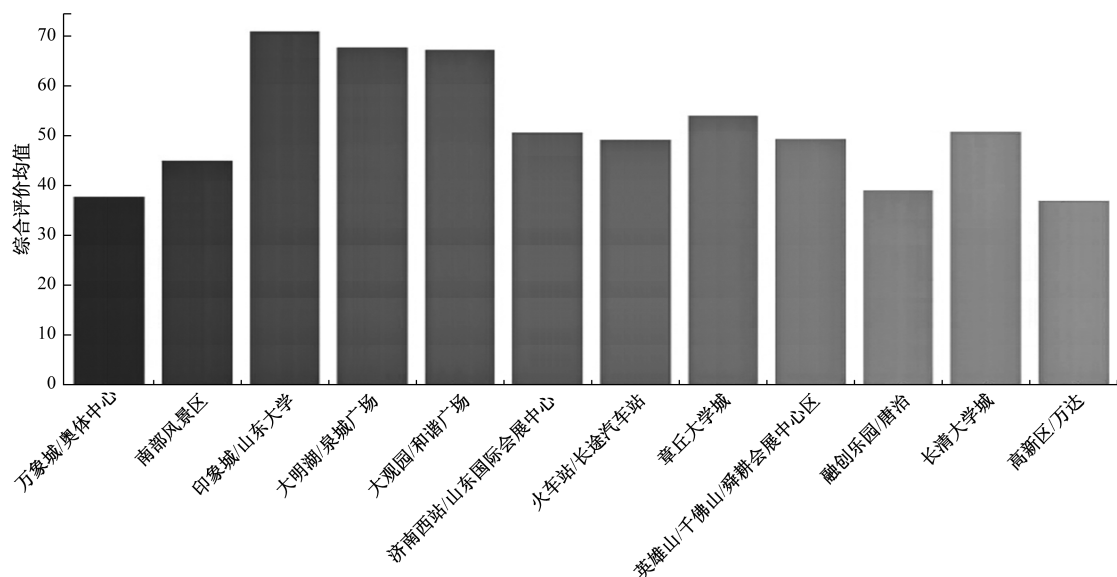


图 10 不同区域下的综合评价均值

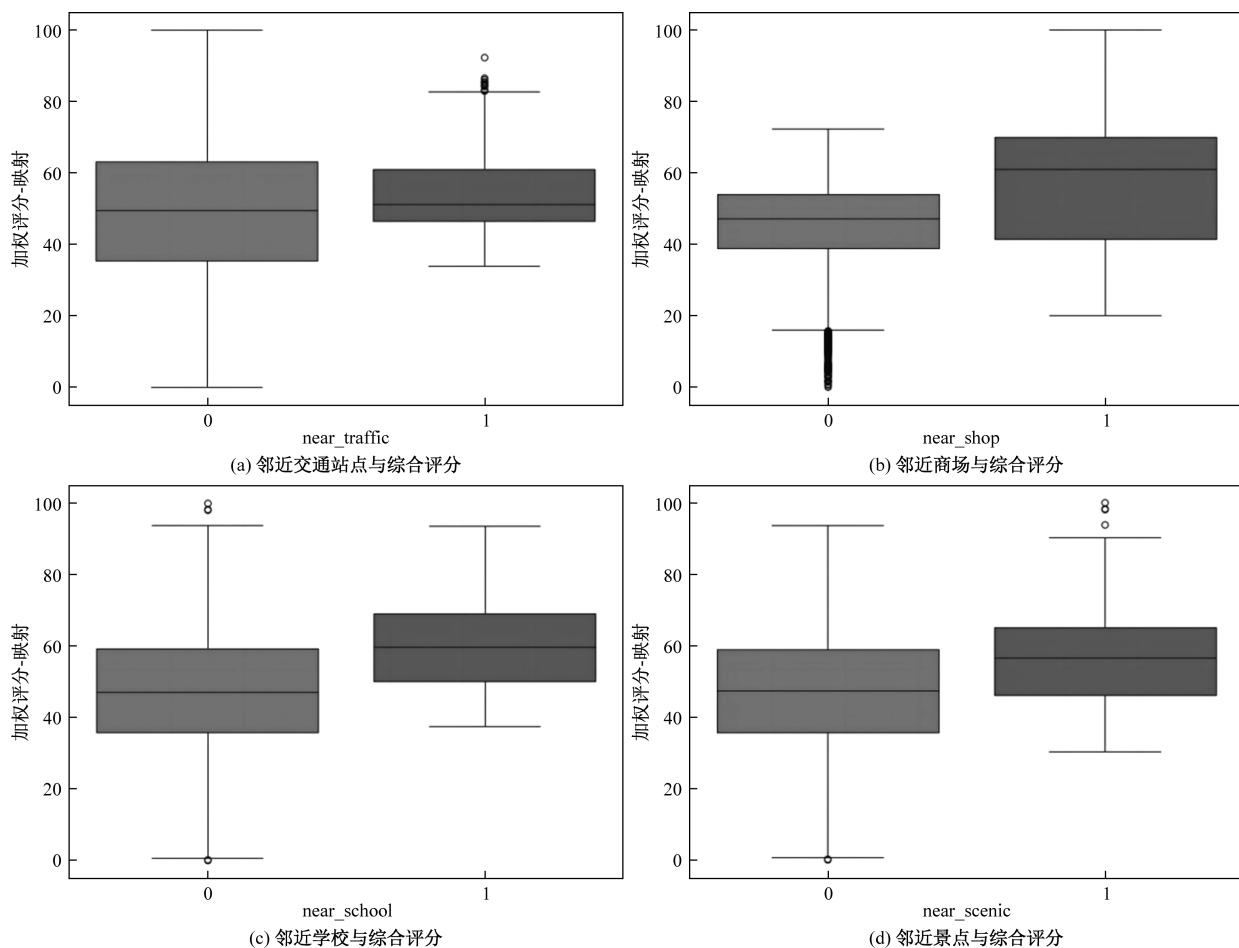


图 11 不同邻近区域的综合评价得分

对于评分在 40~70 分的中等区域,如大明湖/泉城广场、洪家楼/和谐广场等,实施奖励机制,鼓励这些住宿地开发独特的文化体验和特色服务,以增强顾客满意度。同时,支持这些区域参与行业交流和培训,提升服务人员的专业技能。

对于评分超过 70 分的高分区域,如印象城/山东大学和济南西站/山东国际会展中心,将其培育成为济南市的标志性品牌。通过参与国际旅游展会、开发特色文化旅游产品和加强线上营销,进一步提升其在国内外的知名度和吸引力。这些措施将有助于提升济南市整体的旅游住宿服务质量,吸引更多游客。

4.2 根据济南市住宿地理位置与评价得分的关系提出的建议

印象城/山东大学、大明湖/泉城广场、大观园/和谐广场等市中心及热门区域的住宿地综合评价较高。建议济南市文化旅游局针对这些区域,实施定期的服务质量评估和监管,确保服务标准的持续符合性。对于评分较低的偏远地区,如高新区/万

达,建议文化旅游局提供基础设施改善计划,如增加公共交通设施、政府提供资金援助。

4.3 济南市根据评论数与评价的关系提出的建议

顾客评论数与住宿地评价得分之间存在高度的正相关性,建议住宿地通过提供优质服务 and 顾客体验,激励顾客留下正面评论。此外,文化旅游局可以通过社交媒体开展活动,鼓励顾客分享住宿体验,以提高住宿地的在线可见度和口碑。

4.4 根据济南市价格与评价得分的关系提出的建议

由图 9 可知,价格与综合评价得分之间存在一定的正相关关系,但并非所有高价位住宿地都能获得高评价,表明价格并不总是反映服务质量。建议济南市文化旅游局针对印象城/山东大学、大明湖/泉城广场等热门区域,实施价格与服务质量的匹配评估,确保价格合理。同时,对于价格不透明的住宿地,如高新区/万达,要求其明确列出服务内容和费用,提供透明的预订和取消政策,以保护消费者权益,提升市场透明度和竞争力。通过这些措施,

可以促进住宿地提供与价格相称的高质量服务,增强消费者信任。

4.5 根据济南市不同区域综合评价均值的对比提出的建议

4.5.1 高评分区域(大学、景区)

建议济南市文化旅游局进一步推广洪家楼/山东大学区域的住宿地,强调其学术氛围和优美环境,吸引更多游客和商务人士。大明湖/趵突泉作为知名景区,济南市文化旅游局应充分挖掘该区域的旅游价值,加强住宿地与景区的联动,将景区特色融入住宿地,如住宿地推出景区-住宿一体优惠套餐、济南特色文化体验等。

4.5.2 交通枢纽区域

火车站/长途车站区域/济南西站/山博国际会展中心区域作为济南市的重要交通枢纽,建议加强其与住宿地之间的接送服务,提供更多导向标识,提升旅客的出行体验。

5 结论

通过采用爬虫技术获取济南市住宿地数据集,结合 LDA 主题模型和多种机器学习算法构建评价体系,对旅游住宿设施的评价与需求预测进行深入研究。通过揭示文本数据中的潜在主题结构,并筛选影响住宿地评分的核心指标。采用层次分析法(AHP)和熵权法综合确定权重,综合考虑住宿地的接待能力及多种因素进行分析,发现影响旅游住宿设施评价的关键因素与服务质量、地理位置、顾客互动、价格策略以及区域特性密切相关。

(1)对于评分较低的万象新天体育中心和南部风景区,建议济南市文化旅游局提供专项基金,用于基础设施的改造和智能服务技术的引入,以显著提升这些区域的服务质量。

(2)对于评分在 40~70 分的中等区域,如大明湖/泉城广场/洪家楼/和谐广场,实施奖励机制,鼓励住宿地开发独特的文化体验和特色服务,同时支持服务人员参与专业培训,以增强顾客满意度。

(3)对于评分超过 70 分的高分区域,如印象城/山东大学,将其培育成为济南市的标志性品牌,通过参与国际旅游展会和开发特色文化旅游产品,进一步提升其知名度和吸引力。

(4)对于市中心及热门区域的住宿地,如大明湖/泉城广场,建议实施定期的服务质量评估和监管,确保服务标准的持续符合性。

(5)对于偏远地区如高新区/万达,建议提供基础设施改善计划,包括增加公共交通设施,并由政

府提供资金援助,以提升服务质量和吸引力。

(6)鼓励所有住宿地通过提供优质服务 and 顾客体验,激励顾客留下正面评论,同时通过社交媒体开展活动,提高住宿地的在线可见度和口碑。

(7)对于价格不透明的住宿地,如高新区/万达,要求其明确列出服务内容和费用,提供透明的预订和取消政策,以保护消费者权益,提升市场透明度和竞争力。

参考文献

- [1] 2016 年政府工作报告[OB/EB]. [2024-10-01]. https://www.gov.cn/guowuyuan/2016-03/17/content_5054901.htm.
- [2] 国家统计局济南调查队. 2023 年济南市国民经济和社会发展统计公报[EB/OL]. [2024-09-30]. http://jntj.jinan.gov.cn/art/2024/3/26/art_18254_4750937.html.
- [3] 范宁. 基于文本挖掘在民宿满意度中的研究[D]. 桂林: 广西师范大学, 2019.
- [4] 李爽, 张政, 刘娅娅. 基于文本挖掘的西部城市旅游满意度研究——以携程西安和成都为例[J]. 科技和产业, 2022, 22(4): 326-333.
- [5] 范博雷. 基于网络文本挖掘的国家公园旅游满意度评价研究——基于武夷山国家公园的数据分析[J]. 黑龙江生态工程职业学院学报, 2024, 37(2): 56-62.
- [6] 刘阳. 基于文本挖掘的在线旅游产品销量影响因素分析[D]. 北京: 首都经济贸易大学, 2018.
- [7] 郭佳琪, 胡彬. 基于网络文本分析的南京市网红景区旅游满意度研究[J]. 西部旅游, 2024(9): 50-52.
- [8] 杨瑜. 饭店服务质量顾客满意度测评研究[J]. 现代经济信息, 2012(14): 244-245.
- [9] 陈玉莲. 游客对不同类型旅游住宿的体验差异研究——基于旅游网络大数据分析[J]. 北方经贸, 2024(3): 144-149.
- [10] 翟姗姗, 郭致怡, 查思羽, 等. 融合房东生成内容与房客生成内容的共享住宿平台信息服务价值共创机制研究[J]. 情报科学, 2023, 41(6): 84-93.
- [11] 司育. 基于 LDA 主题模型的山西省热门景区综合评价体系研究[D]. 太原: 山西财经大学, 2024.
- [12] 董爽, 汪秋菊. 基于 LDA 的游客感知维度识别: 研究框架与实证研究——以国家矿山公园为例[J]. 北京联合大学学报(人文社会科学版), 2019, 17(2): 42-49.
- [13] 陈亮. 徽州古村落旅游客源网络层级结构分析——以手机信令数据为依据[J]. 黑河学院学报, 2022, 13(2): 44-46.
- [14] 何文宏, 赵乾卉. 基于熵权法(AHP)的城市特色文化街区空间活力评价研究——以合肥市为例[J]. 吉林艺术学院学报, 2024(1): 64-72.
- [15] 高原. 基于 KANO-AHP 的京津冀旅游文创产品评价指标体系构建[J]. 天工, 2024(19): 46-49.
- [16] 郭悦, 白祥. 基于 AHP 和熵值法的乡村旅游发展绩效

- 评估——以乌鲁木齐县上寺村为例[J]. 安徽农业科学, 2024, 52(6): 140-144.
- [17] 李明峰. 基于熵权法的鼓浪屿景区旅游解说系统评价研究[J]. 南京师范大学学报(自然科学版), 2023, 40(3): 117-123.
- [18] 李轩宇, 赵颖, 肖忠良, 李轩. 基于 Python 爬虫的旅游网站数据分析与可视化设计分析[J]. 电脑知识与技术, 2022, 18(33): 58-60.

Evaluation of Tourism Accommodation Reception Capacity Based on the LDA Model: A Case Study of Jinan City

ZHANG Yingying, CHEN Hengyu, ZHANG Mengdi

(Zhongtai Securities Institute for Financial Studies, Shandong University, Jinan 250100, China)

Abstract: A study on evaluating and forecasting the demand for tourism accommodation facilities in Jinan City was conducted. Data was crawled from platforms such as Ctrip, GoWhere.com, and Meituan using Python crawler technology. The LDA topic model was applied to uncover the latent thematic structure within the text data, enabling effective classification of accommodation place labels. Machine learning methods, including XGBoost, CatBoost, LightGBM, and RandomForest, were used to screen the core indicators that affect the ratings of lodging places. An evaluation system was constructed by combining the hierarchical analysis method (AHP) and the entropy weighting method to determine weights. Visual analysis was performed to reveal differences in reception capacity across various themes, accommodation types and areas, as well as proximity to transportation, business districts, schools and scenic spots. Suggestions for optimization are proposed based on the hotel's theme, rating distribution, geographic location, review count and price, aiming to enhance service quality and supervision, thereby increasing reception capacity and user satisfaction.

Keywords: LDA model; machine learning; analytic hierarchy process (AHP); entropy method; python web scraping