

基于文本分类的新能源汽车满意度研究

——以比亚迪与特斯拉为例

赵嘉麒, 罗鄂湘

上海理工大学管理学院, 上海

收稿日期: 2025年1月27日; 录用日期: 2025年2月20日; 发布日期: 2025年2月28日

摘要

量化评估新能源汽车满意度对于推动新能源汽车产业发展、未来能源转型具有重要意义。本文使用python爬取汽车之家上比亚迪与特斯拉两品牌的新能源汽车口碑评论数据, 对非结构化数据进行LDA主题建模分析; 对结构化数据提出一种基于Stacking集成学习的文本情感分类算法以了解各属性的满意度情况, 在对比近十种同质算法后, 选择精度最佳的Catboost、神经网络(ANN)和随机森林(RF)构建多基分类器, 并利用逻辑回归(LR)作为元分类器, 实现对汽车结构化维度的情感分类, 最终准确度高达95%。研究发现, 购车体验、车型选择与噪音问题是结构化数据中被忽视的用户关注维度。而就满意度来看, 比亚迪的整体满意度高于特斯拉, 且均在内饰方面表现较差。未来应关注内饰部分的设计改进, 以提升用户体验并改善用户满意度。

关键词

新能源汽车, 满意度, 集成学习

A Study on Satisfaction of New Energy Vehicles Based on Text Categorization

—Taking BYD and Tesla as an Example

Jiaqi Zhao, Exiang Luo

Business School, University of Shanghai for Science and Technology, Shanghai

Received: Jan. 27th, 2025; accepted: Feb. 20th, 2025; published: Feb. 28th, 2025

Abstract

Quantitatively assessing the satisfaction of new energy vehicles is of great significance for promot-

文章引用: 赵嘉麒, 罗鄂湘. 基于文本分类的新能源汽车满意度研究[J]. 建模与仿真, 2025, 14(2): 532-542.

DOI: 10.12677/mos.2025.142173

ing the development of new energy vehicle industry and future energy transition. In this paper, we use python to crawl the review data of new energy vehicles of BYD and Tesla brands on Autohome.com, and perform LDA topic modeling analysis on the unstructured data; and propose a text sentiment classification algorithm based on Stacking integrated learning for the structured data to understand the satisfaction situation of each attribute. After comparing nearly ten homogeneous algorithms, Catboost, neural network (ANN) and random forest (RF) with the best accuracy are selected to construct a multi-base classifier, and logistic regression (LR) is used as a meta-classifier to realize sentiment classification of structured dimensions of cars, with a high final accuracy of 95%. It was found that car-buying experience, model selection and noise problem are neglected user concern dimensions in structured data. In terms of satisfaction, BYD's overall satisfaction is higher than Tesla's, and both perform poorly in the interior. In the future, attention should be paid to the design improvement of the interior part to enhance user experience and improve user satisfaction.

Keywords

New Energy Vehicles, Satisfaction, Ensemble Learning

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

随着能源和环境问题日益凸显,新能源汽车的发展备受关注。国务院发布的《新能源汽车产业发展规划(2021~2035年)》可视为汽车产业的重要推动力。但目前我国新能源汽车市场仍处于发展阶段,发展模式正从政策主导转向市场主导,存在着缺乏核心技术导致的质量不稳定[1]、基础建设与配套设施不足、相关服务体系(例如售后服务)尚未完善[2]、市场竞争缺乏差异化、产业未形成规模优势[3]等问题,而这些问题的存在不仅影响新能源汽车市场的宣传与推广,还制约我国从汽车大国向强国的转型以及绿色发展战略的实施。同时用户满意度是用户忠诚的重要影响因素,是反映用户感受程度的重要指标,如果需要对用户满意进行实际深入的研究就需要对满意度进行量化。随着社交媒体的普及,消费者通过网络发表的产品反馈(用户生成内容)日益增多。研究开始将传统的满意度模型与用户生成内容相结合,通过在线评论提供商家有效信息从而提升用户满意度。对于消费者来说,这些信息在很大程度上都能够帮助其做出更有效的决策并且降低购物的感知风险;对于商家而言,也能从中受益获取关键的商业价值[4]。

基于此,本文从新能源汽车口碑评论中挖掘顾客情感并以此作为评估新能源汽车满意度的基础,通过对这些文本的分析,能够帮助新能源汽车企业了解用户在购买新能源汽车之后对产品的满意程度,从而帮助其优化产品功能,提升其顾客满意度,进一步优化顾客忠诚度,用户满意度的提升是该领域面临的重要问题之一。

2. 研究现状

文本情感分析(Text Sentiment Analysis)是文本挖掘的一种具体应用,它利用文本挖掘的技术和方法来实现对文本情感的分析,又称意见挖掘,是对带有情感色彩的主观性文本进行分析、处理、归纳和推理的过程[5],旨在识别和理解人们在书面表达、社交媒体帖子、评论等文本中所表达的情感倾向。由于情感分析可以有效地对文本的情感偏好和观点进行挖掘,因此被广泛应用于需要依据用户观点进行判断和分析的领域。

在进行文本情感分析时有部分研究基于情感词典的方法。例如早期研究中, Maite 等(2011)提出了一种基于词典的 SO-CAL 方法, 创建手动词典, 对影评等文本提取情感, 最终实现 78% 的准确率[6]。由于其准确性在跨域应用时会受到词典质量的影响, 因此卫青蓝等(2023)提出了一种基于语义规则的领域自适应中文情感词典自动构建算法, 实现了词典的自动扩充, 消除了跨域情感歧义性, 在酒店与中文情感分析语料库两个数据集上提高了 7.4% 的准确率[7]。还有研究使用机器学习的方法进行文本情感分类, 不同的机器学习分类器在处理情感倾向分类的任务中具有不同的准确性与有效性。例如尹丽春等(2018)为研究影响读者满意度的因素, 挖掘在线商品评论数据并使用朴素贝叶斯分类器进行情感识别, 最终实现 85% 的精确度[8]。集成学习模型通过组合多个机器学习模型的预测力, 往往能够提供比单一模型更为精准的性能, 从而在许多情况下实现了对传统机器学习模型的超越。例如康雁等(2020)提出一种新的集成学习模型并使用中英文数据集进行验证, 研究发现该模型可以有效提取文本中蕴含的更深层次的情感信息, 且准确率高达 83.2%, 优于朴素贝叶斯、支持向量机等传统机器学习模型[9]。此外, 在文本情感分类的研究中为了实现更加详细的语义关联会使用到深度学习。主流的深度学习模型以卷积神经网络与循环神经网络为主, 其中 LSTM (长短期记忆网络)属于循环神经网络的一种, 是目前研究的主要方向之一。例如闵洁等(2023)使用 BiLSTM 模型以提取旅游、酒店、购物等领域的文本情感分类信息, 实现了 92.5% 的准确性, 相比逻辑回归模型, 提升了 8% 的 F 值[10]。从上述研究中发现, 深度学习模型被广泛应用于非结构化文本数据的情绪识别, 旨在从评论文本中提取更为精细的情感维度信息。

目前对于情感倾向分类的研究领域主要集中在酒店、商品等领域。而在新能源汽车行业下, 为了更好地了解用户对于新能源汽车的看法与态度, 有研究使用不同平台收集到的数据对新能源汽车进行情感分析以帮助消费者和企业更好地做出决策。目前评论的类型分为结构化数据与非结构化数据两种, 对结构化数据的研究中, 范黄健等(2022)爬取三大汽车网站下的新能源汽车口碑评论, 利用词云图及语义网络图发现用户对新能源汽车的不满主要体现在续航、内饰及性能等方面[11]。对非结构化数据的研究, Shu 等(2022)为研究电车消费市场中顾客的感知风险, 对从社交媒体上爬取关于汽车评论的非结构化数据进行多维度、细粒度的 NLP 分析, 并在 BERT 模型下实现 0.91 的准确度[12]。

综合来看, 相比情感词典的方法而言, 使用机器学习和深度学习进行文本情感分类时具有更好的分类效果; 而机器学习与深度学习各有优势, 深度学习模型在处理非结构化文本数据能发挥更好的性能。由于本文在进行情感分类时主要针对结构化数据且出于数据量、计算资源以及模型调试等限制原因的考虑, 最终选择机器学习的方法以实现文本情感分类, 并对比选出最优分类模型。而目前的研究针对不同领域均有涉及, 但较少集中在新能源汽车行业, 因此本文将从新能源汽车消费者情感角度出发, 采用文本挖掘方法并利用集成学习模型对不同文本类型的消费者网络评论展开满意度研究。拓宽了现有新能源汽车领域的研究范围, 丰富了新能源汽车研究方法, 充实了该领域的理论文献。

3. 实证分析

目前在线评论分为结构化与非结构化类型, 为使本研究更具有稳健性和普适性, 本文将分别研究这两种类型的评论。对于非结构化数据进行 LDA 主题建模, 以探索其中隐藏的潜在主题; 对于结构化数据进行基于机器学习的情感倾向分类, 以发掘各个属性的具体的满意度情况。具体而言, 首先, 使用一组训练集样本对机器学习分类算法进行训练, 建立情感分类模型。然后, 使用测试集样本对情感分类模型的性能进行测试。所有模型通过性能测试后, 选择最优模型作为最终的情感分类器进行文本情感分类。最后, 输入待分类文本进行情感分类, 得到情感分类的结果, 并将正向占比作为最终的满意度情况。

3.1. 数据准备

使用 Pycharm 爬取汽车类网站排行榜榜首网站“汽车之家”上的新能源汽车口碑评论作为数据源进

行文本情感分类。具体爬取的汽车车型如表 1 所示。爬取的原始数据包括评论的用户名、车型、发表时间、汽车信息(行驶里程、购入时间、购车地点、裸车价格等)、各类评分以及具体评论内容等。

Table 1. Crawling new energy vehicle models

表 1. 爬取的新能源汽车车型

比亚迪			特斯拉		
车系	车型	级别	车系	车型	级别
宋 PLUS	SUV	紧凑型	MODEL X	SUV	中型
唐	SUV	中型	MODEL Y	SUV	中大型
元 PRO	SUV	小型	MODEL 3	轿车	中大型
护卫舰 07	SUV	中型	MODEL S	轿车	中型
汉	轿车	中大型			
海鸥	轿车	小型			
海豚	轿车	小型			
海豹	轿车	中型			

3.2. 数据预处理

使用 Python 对数据进行预处理。首先，需要去除文本中的特殊符号、标点符号和数字等噪声数据以及重复值与缺失值。其次，结合哈工大停用词表删除评论中停用词，以减少干扰。然后，使用 Jieba 工具的精确模式对文本进行分词，将文本划分为单词或短语。最后，使用 TF-IDF 将文本数据向量化，转换为数值型向量，以便后续使用机器学习算法进行分类。

3.3. 高频词提取

在文本数据中，高频词汇是揭示文本核心主题和关键信息的重要线索，也是理解用户观点和情感倾向的基础。通过对高频词汇的提取与分析，可以快速定位用户关注的焦点内容。本研究分别对用户最满意和最不满意的评论文本进行预处理后，对处理后的评论文本进行词频统计，并按照词语出现的频率生成词频表，选取排名前十的高频词，如表 2 所示。这些词频表能够直观地呈现用户在不同情感倾向下的关注点，为进一步的情感分析和主题挖掘提供有力支持。

Table 2. Word frequency for comment text

表 2. 评论文本词频统计表

比亚迪				特斯拉			
满意文本		不满意文本		满意		不满意	
高频词	词频	高频词	词频	高频词	词频	高频词	词频
满意	18500	不满意	15052	特斯拉	623	座椅	360
外观	12114	地方	7418	驾驶	551	不满意	342
空间	11989	空间	5590	加速	483	内饰	277
车子	11474	内饰	5244	空间	471	后排	273
动力	9676	新车	2825	动力	458	续航	263

续表

比亚迪	8189	后排	2782	满意	436	硬	254
设计	6630	座椅	2499	充电	384	充电	244
油耗	6492	味道	2425	外观	377	做工	240
续航	6123	隔音	2412	续航	343	异响	233
配置	5974	异味	2265	操控	318	底盘	200

在比亚迪的评论文本中，用户最为满意的是车辆的空间和外观设计，同时对动力表现、新能源配置和续航能力也持较高评价，反映了用户对比亚迪车型的宽敞空间、引人注目的外观设计以及动力性能等方面的高度满意。不过，用户也表现出一些不满意的意见，主要关注的焦点例如空间、内饰、续航和座椅等，突出了用户对比亚迪车型的后排空间、内饰设计以及异味味道等方面的不满意之处。

在特斯拉的评论文本中，满意的方面，用户普遍对特斯拉的驾驶体验、加速性能、空间设计和动力表现表示满意，特别是对充电便利性和外观设计的喜爱。词频较高的词汇有驾驶、加速、空间、动力等，反映了用户对特斯拉车辆整体性能和驾驶体验的积极评价。然而，在不满意方面，用户关注的焦点主要集中在座椅舒适度、续航能力、后排空间、内饰品质等方面，同时也提到充电体验和特斯拉的做工问题。词频较高的词汇包括座椅、续航、后排、内饰等，突显出用户对特斯拉车型在座椅舒适度、续航能力和内部设计等方面存在一定的不满意情绪。

3.4. LDA 模型

LDA (Latent Dirichlet Allocation)主题模型是一种基于概率的文本建模方法，用于从文本数据中发现潜在主题结构。它假设每篇文档由多个主题混合而成，每个主题包含多个特征词。通过分析词语分布，LDA 能够生成文本的主题分布和每个主题的特征词分布。在应用 LDA 时，主题数的选择至关重要，因为它直接影响模型效果。主题一致性通过评估同一主题下词语的语义关联性来确定最优主题数：一致性越高，主题内词语关联性越强，模型效果越好。本研究先将所有非结构化数据(最满意和最不满意评论)合并后进行 LDA 建模，通过一致性得分确定最优主题数为 12，如图 1 所示。

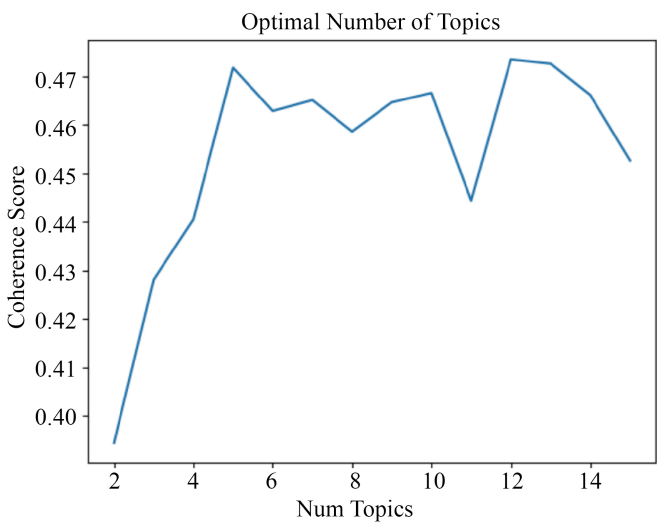


Figure 1. Diagram of trends in thematic coherence
图 1. 主题一致性变化趋势图

对收集的非结构化数据进行 LDA 主题建模，最终得出的结果如下表 3 所示，其中包含 12 个主题，每个主题下包含权重靠前的 10 个特征词。以 Topic 0 为例，相关评论主要与座椅的舒适性、隔音减震效果以及驾驶体验，这些特征词都与舒适性相关，因此将该主题归纳为舒适性。类似地将 Topic 1-Topic 11 分别归纳为：操作感受、车辆性能、操纵体验、性价比、购车体验、车型选择、噪音问题、内饰、外观、续航、空间。

Table 3. Overall subject headings matrix

表 3. 总主题词矩阵

主题		特征词
Topic 0	舒适性	座椅 硬 效果 功能 驾驶 隔音 减震 空调 颠簸 异响
Topic 1	操作感受	偏硬 刮 蹭 太长 太久 后续 一是 二是 按钮 应急
Topic 2	车辆性能	车子 动力 地方 这款 满意 表现 高 这辆 油耗 特别
Topic 3	操纵体验	车身 方向盘 操控 车漆 漆 动力 踏板 驾驶 转向 车子
Topic 4	性价比	价格 优惠 耐脏 转弯半径 高 一点 贵 不太好 便宜 后期
Topic 5	购车体验	不满意 提车 时间 地方 说 新车 买 发现 月 一点
Topic 6	车型选择	特斯拉 比亚迪 薄 电池 新能源 希望 品牌 车型 避震 选择
Topic 7	噪音问题	高速 噪音 唐 底盘 发动机 胎噪 不满意 隔音 说 声音
Topic 8	内饰	不满意 地方 内饰 车子 异味 做工 味道 一点 新车 空间
Topic 9	外观	内饰 设计 外观 喜欢 颜色 好看 满意 轮毂 风格 整体
Topic 10	续航	续航 充电 跑 公里 桩 里程 高速 电 冬天 空调
Topic 11	空间	空间 后排 第三排 坐 后备箱 座椅 乘坐 一点 稍微 舒服

这些主题维度涵盖了用户在购车过程中所注重的几个方面，首先就车辆自身而言，舒适性是一个重要的考虑因素；操作感受和操纵体验是评价一辆车的重要标准；噪音问题是评价的指标之一；内饰和外观是车辆整体质感和设计的重点；续航能力是选购新能源车型所需考虑的关键指标；同时期望车辆内部空间布局合理、宽敞舒适。其次就购车过程而言，购车体验也是重要的一环，用户希望能够获得专业、完善的售前及售后服务，使整个购车过程更加愉快和顺利。同时价格问题及性价比一直是购车者的关注焦点，更偏向能在合理的价格范围内购买到性价比较高的车辆。

LDA 主题建模可以帮助我们从大量的文本数据中提取并总结关键主题，为进一步地分析和决策提供更全面的信息和视角。这种对比分析有助于我们更好地理解 and 解释数据。通过与原维度的对比可以发现新增购车体验、车型选择与噪音问题等方面，这些主题在原有结构化维度中并没有被充分关注或观察到。购车体验指的是购车者在购买新能源汽车过程中所获得的服务和体验。提供专业的购车咨询和完善的售后服务将使购车者更加满意，并增加其购买意向。随着新能源汽车市场上涌现出各种新的车型和品牌，用户需要考虑自身需求，一个正确的车型选择能够提升用户的满意度。新能源汽车通常采用电动驱动系统，相对于传统燃油车来说，它们往往具有更低的噪音水平，能够提升用户的驾驶体验与舒适性。因此，在新能源汽车市场中，购车体验、车型选择和噪音问题的重要性不可忽视。

接着对最满意与最不满意细分，分别进行 LDA 主题建模，试图寻找满意与不满意在具体维度上的差异，生成的主题一致性趋势变化图如图 2 所示。

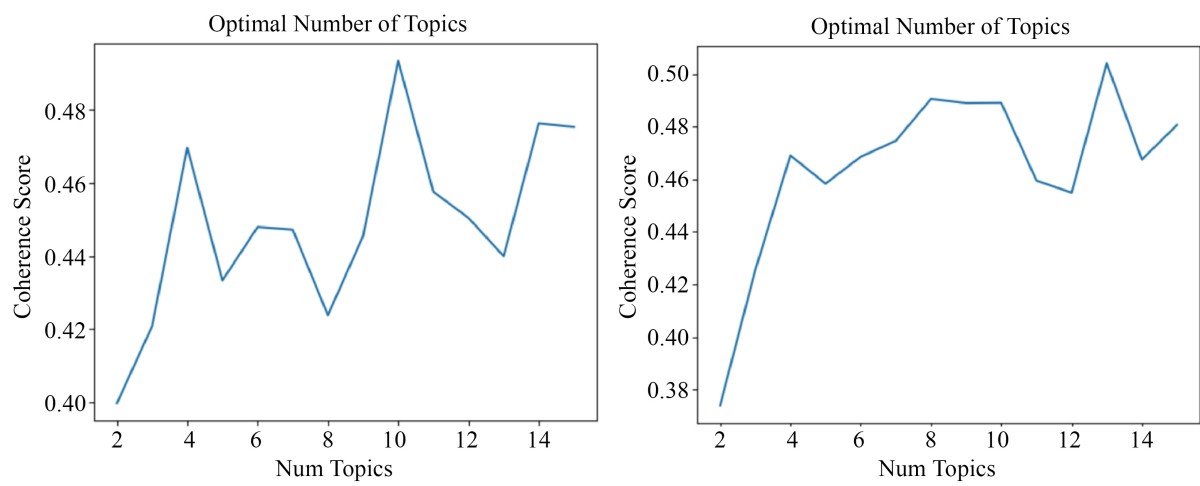


Figure 2. Diagram of dissatisfaction & satisfaction theme consistency trend
图 2. 不满意&满意主题一致性趋势图

最不满意文本选择 10 作为最优主题数，最满意文本选择 13 作为最优主题数，生成主题词及特征词如下表 4 所示。

Table 4. Satisfaction & dissatisfaction subject matter matrix
表 4. 满意&不满意主题词矩阵

主题		特征词
最满意	智能	功能 驾驶 辅助 自动 配置 手机 系统 智能 座椅 科技
	加速	唐 加速 模式 秒 动力 满意 四驱 电机 底盘 悬挂
	设计	空间 外观 满意 设计 颜值 车子 高 喜欢 这辆 七座
	停车	空间 停车 特别 露营 倒 喜欢 不用 小巧 上下班 满意
	品牌	比亚迪 特斯拉 新能源 品牌 配置 内饰 高 选择 汽车 满意
	续航	续航 充电 公里 跑 油耗 电 满意 纯电 高速 油车
	外观	外观 设计 满意 喜欢 这款 造型 车身 特别 时尚 颜值
	价格	买 满意 新能源 这款 油耗 用车 低 价格 高 成本
	动力	动力 满意 车子 提速 空间 这款 外观 不错 地方 加速
	宽敞	空间 后备箱 后排 宽敞 满意 坐 放 乘坐 东西 储物
	消费者	熟悉 能量 消费者 改色 外观 经历 满意 中心 全国 车能
	试驾	动力 说 试驾 提车 买 销售 加速 感受 电车 月
	操控	驾驶 车子 座椅 满意 操控 感受 动力 不错 空间 舒服
最不满意	充电槽	手机 槽 停车 充电 吐 支持 豪华 转弯半径 只能说
	底盘	车子 底盘 不满意 地方 颠簸 减震 硬 一点 效果 稍微
	座椅	座椅 轮胎 后排 空调 调节 异响 玻璃 功能 夏天 说
	空间	空间 后排 后备箱 不满意 第三排 坐 座椅 地方 一点 乘坐
	续航	续航 新车 公里 异味 跑 不满意 味道 一点 空调 通风
	高速噪音	高速 噪音 不满意 隔音 地方 车子 动力 声音 发动机 胎噪
	车漆	车漆 薄 漆 好看 划痕 黑色 太 烤 贴 天窗
	充电	充电 价格 优惠 不满意 高 特斯拉 桩 新能源 车子 比亚迪
	提车时间	时间 不满意 提车 月 长 地方 等车 唐 一点 销售
	内饰	内饰 不满意 设计 味道 喜欢 太 地方 做工 新车 塑料

上表揭示了用户对新能源汽车在满意度和不满意度方面的具体关注点。在不满意的方面,用户对充电槽的设计和性能提出了批评,表达了对底盘稳定性和舒适性的期望未能满足,同时座椅的舒适度和调节功能也未达到预期。车辆内部空间,尤其是后排座椅和储物空间的局限性,对用户体验产生了负面影响。续航能力的不足、高速行驶时的噪音问题、车漆质量的担忧、充电过程中的各种不便、提车时间的延迟以及内饰设计和材质的不足,均是用户不满的重要因素。相较之下,在满意的方面,用户对新能源汽车的智能化特征给予了积极评价,包括车辆的自动化功能和用户界面。加速性能获得了认可,用户对车辆的快速响应和动力输出感到满意。设计方面,无论是内部还是外部,都得到了用户的青睐。停车便利性、品牌信誉、外观吸引力、合理的价格定位以及宽敞的车内空间,都是用户表达满意的关键点。此外,消费者对试驾体验和车辆操控性能的正面评价也凸显了新能源汽车的动态优势。

对比满意与不满意的维度,我们可以看到用户的期望是多方面的,用户赞赏新能源汽车在智能科技、性能和设计上的创新,这些都是新能源汽车品牌的强项。然而,日常使用中的实际体验,如续航、充电便利性和车内舒适度,仍有改进空间。这些反馈对于制造商来说是宝贵的,它们不仅揭示了产品的优势,也指出了需要关注和改进的领域,以满足消费者的全面需求并提升整体的客户满意度。

3.5. 情感分类

目前机器学习算法已成功应用于情感分析以及文本分类。机器学习分类器可以大致分为决策树(如C4.5、ID3 和随机森林)、基于感知器的方法(如人工神经网络)、统计学习方法(如朴素贝叶斯分类器)、基于实例的分类器(如k近邻算法)和支持向量机。目前朴素贝叶斯分类器、支持向量机和决策树被广泛用于文本分类问题[13]。

近年来,集成学习被认为是机器学习中最成功的技术之一。集成学习将几种机器学习技术结合到一个预测模型中,以便与单个模型(基本分类器)相比提高性能。大多数集成学习方法使用单一基学习算法来生成同构基学习器,即所有基学习器都基于同一种算法。也有一些方法使用异构基学习器,即基学习器由多种不同的算法生成[14]。组合基分类器可以通过不同的方式完成,包括Voting投票法、Bagging套袋法、Boosting提升法和Stacking堆叠法等。

在投票法中,每个分类器为一个类别投票,并通过考虑所有投票作出最终预测。在多数投票中,最终的预测结果等于由最多分类器选择的输出结果。多数投票的一个限制是它对所有模型都是一样的,即所有模型对预测的贡献都是一样的。为了解决这个问题,使用加权投票,因为模型在投票中的重要性是不同的。套袋法是一种同质集成方法,在训练集的不同随机子集上并行训练相同的基学习器。为了获得数据子集,采用了自举抽样。该方法在基分类器不稳定的情况下减少了方差,提高了稳定性。提升法是一种同套袋一样的同质集成方法,但与套袋相反,它是一个连续的过程,其中每个后续模型都试图纠正其前身的错误。为了提高模型的准确性,对前几轮被错误分类的样本赋予了更多的权重。增强方法减小了偏置误差,产生了强预测模型。堆叠法则是一种异构集成学习技术,通过元分类器将多个分类模型组合在一起。基础分类器基于完整的训练集进行训练,然后将基础级模型的输出作为特征对元模型进行训练。其优势在于,可以在分类任务上结合不同机器学习算法的优势,并做出比集成中任何单一模型都更好的预测,有研究认为是目前性能最好的集成学习模型[15]。

3.5.1. 评价指标

在机器学习中评价指标有四类:真正类,将样本预测为正类且预测正确(True Positive, TP)、去真类,将样本预测为负类但预测错误(False Negative, FN)、存伪类,将样本预测为正类但预测错误(False Positive, FP)、真负类,将样本预测为负类但预测正确(True Negative, TN)。

准确率(Accuracy)指所有预测正确数和占全部样本数的百分比。

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \tag{1}$$

精确率(Precision)也叫查准率，即指识别正确占识别为正样本的总数的比例，即预测为正/阳性的样本中有多少是真正的正样本。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

召回率(recall)也叫查全率，即有多少正性样本被正确识别出来，进行了识别且识别是正确的样本，占总正样本的比例。

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

F1 值作为 Precision 与 Recall 二者结合起来之后的评价指标来协调上述两方面。

$$\text{F1} = \frac{2\text{Precision}}{\text{Precision} * \text{Recall}} \tag{4}$$

3.5.2. 分类模型对比

本文将爬取的最满意和最不满意维度的评论分别作为正向情感评论和负向情并分别记为 1 和 0。将预处理过后的数据按照 2:8 的比例划分测试集与训练集，将训练集数据带入模型，用测试集输出预测结果，最终输出模型评估指标。模型方面，将采用贝叶斯、K 近邻、逻辑回归、决策树和人工神经网络算法作为基分类器，对比不同方法的集成模型情况，提高模型的泛化能力，并提升分类精度，以寻找最优模型。

Table 5. Model performance comparison
表 5. 模型性能对比

策略	分类器	Precision	Recall	F1-score	Accuracy
	决策树 DT	0.8683	0.8817	0.8749	0.8739
	逻辑回归 LR	0.9342	0.9509	0.9425	0.9419
	朴素贝叶斯(多项式)	0.8954	0.9486	0.9213	0.9189
	朴素贝叶斯(伯努利)	0.9366	0.8091	0.8682	0.8771
	K 近邻 KNN	0.9555	0.6735	0.7901	0.8210
	神经网络 ANN	0.9246	0.9173	0.9209	0.9212
bagging	随机森林 RF	0.9136	0.9493	0.9311	0.9298
Voting	RF、ANN、LR (hard)	0.9332	0.9539	0.9434	0.9428
	RF、ANN、LR (soft)	0.9329	0.9360	0.9344	0.9343
Boosting	GradientBoosting	0.9035	0.8741	0.8886	0.8903
	xgboost	0.9237	0.9245	0.9241	0.9240
	LightGBM	0.9190	0.9365	0.9277	0.9269
	CatBoost	0.9329	0.9313	0.9321	0.9322
	AdaBoost	0.8977	0.8861	0.8919	0.8925

续表

Stacking	① RF、ANN、Catboost ② LR	0.9376	0.9466	0.9421	0.9417
(调参&正则化)	① RF、ANN、Catboost ② LR	0.9438	0.9498	0.9468	0.9466
	① RF、ANN、LightGBM ② LR	0.9349	0.9471	0.9410	0.9405

注: stacking 策略下, ① 表示基分类器, ② 表示元分类器。

在本研究中, 对比不同集成方法在文本分类上的效果, 具体结果如上表 5 所示。其中除去单一模型外, 包含常见的五种 Boosting 模型; Bagging 选择随机森林作为代表; Voting 选择随机森林、神经网络、逻辑回归作为基分类器, 指定两种投票模式; Stacking 选择单一模型中表现较佳的 Catboost、神经网络 ANN、随机森林 RF 作为基分类器, 逻辑回归 LR 作为元分类器, 组合形成异构集成模型(经多次试验, 该组合分类器在 Stacking 策略下的性能最优)。由上表可以看出上述几种模型除了 KNN, 其余模型在四个评价指标上的得分都超过 0.8, 说明它们均具有较好的分类性能。

通过进一步比较各个性能的表现可以发现, 尽管集成学习的基础是单一模型, 但经过集成训练后, 性能并不是简单的各单一模型的均值, 而是优于每个单一模型。结果显示在进行参数调整与数据正则化后的 Stacking 模型(RF + ANN + Catboost + LR)具有最优的性能表现, 分类准确率最高能到 95%左右, 精确率、召回率及 F1 值也稳定在这个区间内, 说明该模型能够较好地实现文本情感分类。同时经过调整后的模型能够更好地防止数据过拟合问题的发生。因此采用该模型作为最优分类器, 对未标注的结构化评论数据进行情感分类, 分类后, 将正向评论的占比作为满意度得分。基于此, 比亚迪与特斯拉两品牌满意度情况如下图 3 所示。

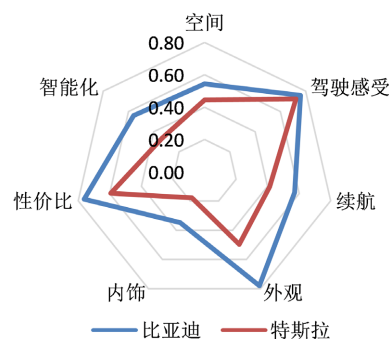


Figure 3. Satisfaction by dimension
图 3. 各维度满意度情况

该图展现了新能源汽车在空间、驾驶感受、续航、外观、内饰、性价比和智能化七个维度评论的情感倾向情况, 经过换算可以体现出消费者对于新能源汽车的各个方面的满意情况。越靠近中心点, 表明该维度的满意度越差。由上图可知, 两品牌在各维度上趋势大致相同, 但比亚迪的用户满意度整体高于特斯拉。比亚迪用户对新能源汽车的满意度情况为: 外观 > 性价比 > 驾驶感受 > 续航 > 智能化 > 空间 > 内饰; 特斯拉用户的满意度情况为: 驾驶感受 > 性价比 > 外观 > 空间 > 续航 > 智能化 > 内饰。说明比亚迪和特斯拉用户在对新能源汽车的满意度方面存在明显的差异, 比亚迪用户更注重外观和性价比, 而特斯拉用户更注重驾驶感受和性价比。同时, 智能化和内饰对于两个品牌的用户群体的满意

度普遍较低。

4. 结论与建议

本文使用比亚迪与特斯拉两品牌共计十二款车型的口碑评论数据进行满意度研究。经过模型对比, 最终使用集成学习 Stacking 框架, 融合 Catboost、神经网络 ANN 和随机森林 RF 构建多基分类器, 并利用逻辑回归 LR 作为元分类器, 进行文本情感分类, 模型的准确率达到 95%, 召回率、精确率等指标均达到了理想的预计值, 验证了模型的有效性和实用性, 且具有相对的稳定性。通过该分类模型得出的结果进一步研究新能源汽车各维度的用户满意度情况。研究发现用户对于目前新能源汽车的外观、性价比与驾驶感受比较满意; 但对于内饰与智能化的满意度较低。同时根据 LDA 主题建模的结果分析发现, 现有评论维度忽略了用户在购车体验、车型选择与汽车噪音方面的看法, 而这些维度对用户对于新能源汽车的满意度也存在一定的影响。综上, 本文基于上述研究结果给出以下建议。

鉴于用户对内饰与智能化的满意度较低, 新能源汽车企业应该注重提升内饰设计和智能化技术水平, 通过引入更先进的智能科技例如配置人机交互系统和高质量的内饰设计, 同时可以通过改进隔音设计、降噪技术和材料选择, 提供更安静舒适的驾驶环境, 以提升用户满意度。购车体验是影响用户对新能源汽车满意度的重要因素之一, 企业应优化销售流程、提升销售人员的专业水平, 以及提供更便利的购车渠道, 为消费者营造愉快、顺畅的购车体验。同时根据用户需求和市场趋势的不断变化, 企业应提供更多样化的车型选择, 包括不同尺寸、功能和配置的车型, 以满足不同消费者的需求。

参考文献

- [1] 张芳, 包先建. 我国新能源汽车市场推广问题与对策研究[J]. 价格理论与实践, 2011(5): 85-86.
- [2] 高仲敏. 中国新能源汽车产业存在的问题和解决策略[J]. 汽车维修技师, 2023(8): 110.
- [3] 郑舒允. 高质量绿色发展下中国新能源汽车产业发展现状及问题分析[J]. 科技和产业, 2022, 22(3): 132-137.
- [4] Lukyanenko, R., Parsons, J. and Wiersma, Y.F. (2014) The IQ of the Crowd: Understanding and Improving Information Quality in Structured User-Generated Content. *Information Systems Research*, **25**, 669-689. <https://doi.org/10.1287/isre.2014.0537>
- [5] 赵妍妍, 秦兵, 刘挺. 文本情感分析[J]. 软件学报, 2010, 21(8): 1834-1848.
- [6] Taboada, M., Brooke, J., Tofiloski, M., Voll, K. and Stede, M. (2011) Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, **37**, 267-307. https://doi.org/10.1162/coli_a_00049
- [7] 卫青蓝, 何雨, 宋金宝. 基于语义规则的自适应情感词典自动构建算法[J/OL]. 北京航空航天大学学报, 1-11. <https://doi.org/10.13700/j.bh.1001-5965.2023.0367>, 2024-12-24.
- [8] 尹丽春, 王悦. 基于在线评论的图书消费者满意度影响因素与作用机理[J]. 图书情报工作, 2019, 63(22): 106-117.
- [9] 康雁, 李浩, 梁文韬, 宁浩宇, 霍雯. 针对文本情感分类任务的 textSE-ResNeXt 集成模型[J]. 计算机工程与应用, 2020, 56(7): 205-209.
- [10] 闵洁, 吉秉戔. 基于主题相关性和深度学习的文本情感分析[J]. 信阳师范学院学报(自然科学版), 2023, 36(3): 490-494.
- [11] 范黄健, 叶楠, 张梦婷. 基于多源数据的新能源汽车用户满意度分析[J]. 经营与管理, 2024(2): 32-39.
- [12] Shu, T., Wang, Z., Lin, L., Jia, H. and Zhou, J. (2022) Customer Perceived Risk Measurement with NLP Method in Electric Vehicles Consumption Market: Empirical Study from China. *Energies*, **15**, Article 1637. <https://doi.org/10.3390/en15051637>
- [13] Ikonomakis, M., Kotsiantis, S. and Tampakas, V. (2005) Text Classification Using Machine Learning Techniques. *WSEAS Transactions on Computers*, **4**, 966-974.
- [14] Smagulova, K. and James, A.P. (2019) A Survey on LSTM Memristive Neural Network Architectures and Applications. *The European Physical Journal Special Topics*, **228**, 2313-2324. <https://doi.org/10.1140/epjst/e2019-900046-x>
- [15] 陈红阳, 黄正洪, 何盈盈, 周也力. 融合内容特征与传播特征的微博文本情感分类[J]. 重庆理工大学学报(自然科学), 2023, 37(7): 245-255.