

齊魯工業大學

本科毕业设计（论文）

基于情感分析的网易云音乐个性化歌曲推荐系统设计与实现

学部（学院） 计算机科学与技术学部

专业班级 数据科学 21-2

学生姓名 任宇然

学号 202107090151

导师姓名 肖雨腾

2025 年 6 月 3 日

本科毕业设计（论文）

基于情感分析的音乐个性化歌曲推荐系统设计与实现

学部（学院） 计算机科学与技术学部

专业班级 数据科学 21-2

学生姓名 任宇然

学号 202107090151

导师姓名 肖雨腾

2025 年 6 月 3 日

齐鲁工业大学本科毕业设计（论文）

原创性声明

本人郑重声明：所提交的毕业设计（论文），是本人在指导教师的指导下独立研究、撰写的成果。设计（论文）中引用他人的文献、数据、图件、资料，均已在设计（论文）中加以说明，除此之外，本设计（论文）不含任何其他个人或集体已经发表或撰写的成果作品。对本文研究做出重要贡献的个人和集体，均已在文中作了明确说明并表示了谢意。本声明的法律结果由本人承担。

毕业设计（论文）作者签名： 任宇然

2025年6月3日

齐鲁工业大学本科毕业设计（论文）

使用授权说明

本毕业设计（论文）作者完全了解学校有关保留、使用毕业设计（论文）的规定，即：学校有权保留、送交设计（论文）的复印件，允许设计（论文）被查阅和借阅，学校可以公布设计（论文）的全部或部分内容，可以采用影印、扫描等复制手段保存本设计（论文）。

指导教师签名： 肖雨腾

2025年6月3日

毕业设计（论文）作者签名： 任宇然

2025年6月3日

目 录

摘 要	I
ABSTRACT	II
第 1 章 绪论	1
1.1 研究背景与意义	1
1.2 国内外研究现状	1
1.2.1 音乐系统的研究	1
1.2.2 推荐算法的研究	2
1.3 研究内容	3
1.4 论文结构	4
第 2 章 相关技术介绍	5
2.1 自然语言处理技术	5
2.2 评论与歌曲情感判定规则	5
2.2.1 单条评论情感判定规则	5
2.2.2 评论与歌曲情感映射规则	6
2.3 相关机器学习算法	6
2.4 情感匹配规则	9
2.5 本章小结	9
第 3 章 系统分析	10
3.1 需求分析	10
3.1.1 用户需求分析	10
3.1.2 功能需求分析	10
3.2 可行性分析	11
3.2.1 经济可行性分析	11
3.2.2 操作可行性分析	11

3.2.3 技术可行性分析	12
3.3 本章小结	12
第 4 章 系统设计与实现	13
4.1 系统总体架构	13
4.1.1 设计思路	13
4.1.2 技术栈选择	14
4.2 核心功能模块实现	14
4.2.1 数据采集模块	14
4.2.2 情感分析模块	15
4.2.3 歌曲推荐模块	16
4.2.4 用户交互模块	16
4.3 关键技术难点与解决方案	19
4.4 本章小结	20
第 5 章 测试与评估	21
5.1 评价指标	21
5.1.1 情感分析算法评价指标	21
5.1.2 推荐系统评价指标	21
5.2 情感分析模型评估	21
5.2.1 模型训练与测试数据	21
5.2.2 评论情感分析	22
5.2.3 歌曲情感分类模型对比与分析	23
5.3 系统总体评估	26
5.3.1 推荐系统性能	26
5.3.2 系统优势与不足	26
5.4 本章小结	27
第 6 章 结论	28

6.1 总结	28
6.2 展望	28
参考文献	30
致 谢	32
附 录	33

摘 要

在数字化音乐快速发展的今天，各大音乐平台之间的竞争日益激烈，用户对个性化音乐推荐功能的要求越来越高。目前大多数音乐平台只是根据用户日常搜索记录和常听曲目推送歌曲，难以感知用户当下的情绪状态，无法满足用户在不同情绪下的音乐需求。比如当用户心情低落时，推荐一首舒缓的慢歌往往比摇滚乐更容易引发情感共鸣。为突破这一局限，本研究利用机器学习技术深入分析网易云音乐用户评论中的情感倾向，构建基于情感分析的个性化歌曲推荐系统。

本研究首先利用 Python 爬虫技术获取网易云音乐评论数据，接着运用自然语言处理技术对评论文本进行清洗、分词、去除停用词等预处理操作，结合 SnowNLP 情感分析、情感词典匹配和情感短语识别的混合策略进行评论情感标注。在模型选择环节，本研究借助混淆矩阵对支持向量机、朴素贝叶斯、随机森林三种机器学习算法在评论情感分类任务上的性能进行对比，并在系统中展示歌曲情感分类结果。依据机器学习划分出的情感分类结果，结合统计计算得到每首歌的情感分布比例，构建三层匹配机制进行歌曲推荐，达成根据用户情绪状态和需求的个性化音乐推荐。还运用可视化技术展示分析结果与评论词云使系统变得更加直观清晰。实验结果显示，该系统能有效按照用户的情感需求来实现歌曲推荐，为音乐平台增添新的推荐策略，提高了用户与平台的互动体验。

关键词：音乐推荐；情感分析；机器学习；自然语言处理

ABSTRACT

Today, with the rapid development of digital music, the competition among major music platforms is becoming increasingly fierce, and users' demands for personalized music recommendation functions are getting higher and higher. At present, most music platforms only push songs based on users' daily search records and frequently listened tracks, making it difficult to perceive users' current emotional states and failing to meet their music needs under different emotions. For example, when a user is in a low mood, recommending a soothing slow song is often more likely to evoke emotional resonance than rock music. To break through this limitation, this study utilizes machine learning technology to deeply analyze the emotional tendencies in the user comments of NetEase Cloud Music and construct a personalized song recommendation system based on emotional analysis.

This study first uses Python web crawler technology to obtain the comment data of NetEase Cloud Music, and employs natural language processing technology to perform preprocessing operations such as cleaning, word segmentation, and removal of stop words on the comment texts. Combined with a hybrid strategy of SnowNLP sentiment analysis, sentiment dictionary matching, and sentiment phrase recognition, comment sentiment annotation is carried out. In terms of model selection, this study compared the performance of three machine learning algorithms, namely Support Vector Machine, Naive Bayes, and Random Forest, in the task of comment sentiment classification through the confusion matrix, and presented the classification results of song sentiment in the system. Based on the emotion classification results divided by machine learning, combined with statistical calculation to obtain the emotion distribution ratio of each song, a three-layer matching mechanism was constructed for song recommendation, achieving personalized music recommendation according to the user's emotional state and needs. Visualization technology is also employed to display the analysis results and comment word clouds, making the system more intuitive and clear. The study results show that this system can effectively implement song recommendations according to the emotional needs of users, inject a new recommendation strategy into the music platform, and enhance the interactive experience between users and the platform.

Key words : Music Recommendation; Sentiment Analysis; Machine Learning; Natural Language Processing

第 1 章 绪论

1.1 研究背景与意义

随着互联网技术不断发展以及数字音乐平台日益普及，人们获取音乐资源的方式出现显著变化。网易云音乐、QQ 音乐等音乐平台存有海量音乐资源，使得用户面临着信息过载状况，短时间内很难找到符合自身需求的音乐作品。音乐推荐系统作为解决信息过载的重要技术手段，受到学术界与产业界广泛关注。传统音乐推荐系统主要依据协同过滤方法^[1]或者基于用户常听曲目来开展推荐^[2]，虽说能在一定程度上满足用户发现新音乐的需求，但有着明显不足：它们主要关注用户的历史行为数据或者音乐的低层次特征，像是节拍、旋律、音高等，却忽略音乐作为情感载体的核心属性以及用户情感需求这一重要方面。

音乐作为一种情感表达的艺术形式，有相当强烈的情感属性。研究表明，人们选择想听的音乐容易受到当前情绪状态的作用^[3]，有时候期望借助音乐强化现有情绪，比如聆听悲伤的音乐释放悲伤情绪，而有时又会希望借助音乐调节情绪，像是依靠欢快的音乐摆脱抑郁状态。所以把情感因素融入音乐推荐系统，达成情感驱动的个性化推荐，有着重要的理论价值和实践意义。

本研究设计并实现基于评论情感分析的个性化音乐推荐系统，结合 SnowNLP 情感分析、情感词典匹配和情感短语识别的混合策略对评论进行情感标注，统计得到歌曲情感分布特征，并使用机器学习算法实现对歌曲的情感分类。在机器学习分类歌曲的基础上，融合用户情绪状态、需求以及多样性因子创建三层匹配机制，从而实现依据用户情绪状态和需求对歌曲的个性化推荐。这一研究丰富了音乐推荐系统的理论体系，为音乐平台提供可以提升用户体验的新思路和技术手段。

1.2 国内外研究现状

近年来，人工智能技术飞速发展，推荐系统已成为学术界与工业界的研究热点，在多个领域都有着广泛应用。音乐平台作为信息传播与用户情感的重要载体，个性化音乐推荐技术成为音乐信息检索与推荐领域的关键研究方向。国内外众多音乐平台、高校以及科研机构都积极参与其中并取得丰富的成果。接下来从音乐系统以及推荐算法这两个方面论述国内外研究现状。

1.2.1 音乐系统的研究

国内音乐系统的研究重点在于将中文音乐数据与本地文化因素相结合，强化用户行

为分析与多模态数据融合^[4]。网易云音乐借助整合用户播放历史、常听曲目以及使用时间、设备类型等场景化数据构建立体用户画像，并结合音乐信息检索技术对旋律节奏进行情感分类来实现场景化推荐，比如说在深夜向用户优先推荐节奏较为舒缓、旋律相对轻柔的音乐；通勤时段则会推送活力动感的流行曲目。QQ 音乐创建了歌曲知识图谱，分析歌曲的流派、歌手、音色、创作背景等特征，并结合用户的听歌历史、收藏偏好、播放时段等行为数据，构建出个性化的音乐推荐模型。针对国内特有的音乐类型，像“国风”“民谣”等开发专门的标签，借助对音乐元素进行细致标注达成精准的本土化推荐。拿“国风”音乐来说，系统会根据歌曲的编曲特色、歌词风格等维度，为用户推送符合其喜好的作品。酷狗音乐在垂直音乐领域深入耕耘，围绕用户多样化的音乐需求打造“听看唱玩”一体化的推荐生态，借助直播、K 歌、音乐社区等功能提高用户互动性，结合用户在不同场景下的行为数据优化推荐策略来满足用户多样化的音乐娱乐需求。

对比来说，国外音乐系统的研究更注重技术创新与多模态融合。Spotify 把用户听歌历史、搜索关键词、播放列表等文本数据跟音乐旋律、节奏、和声等音频特征相结合，构建起复杂的多模态推荐模型。精准捕捉用户个性化音乐偏好，为用户提供更贴合需求的音乐推荐。TALK-PLAY 系统将音乐推荐任务重新定义为大语言模型的下一个词预测任务，借助扩展词表融合音频、歌词等多模态信息，在对话中动态调整推荐策略提升上下文理解能力。Apple Music 凭借整合用户收听历史、收藏喜好、歌曲评分等数据，以及专家音乐评分与用户社交网络数据来构建混合推荐模型，从专业评价与社交影响这两个方面优化推荐结果。另外国外音乐系统还积极探索跨领域技术融合，像把音乐推荐与虚拟现实技术结合^[5]，为用户打造沉浸式音乐体验，通过场景化的音乐适配进一步提升用户参与度与情感共鸣。

这些研究从系统架构层面描述了国内外在音乐推荐领域的创新实践，不过在情感识别上国内外音乐系统虽可以根据用户行为和音乐特征进行歌曲推荐，但对用户实时情感状态的捕捉仍有着很大的不足。目前多数系统没有充分考虑用户在不同情绪下对音乐的动态需求，这为后续研究指明了方向。

1.2.2 推荐算法的研究

当下的推荐算法主要有基于内容的推荐算法、基于协同过滤的推荐算法以及混合推荐算法。基于内容的推荐算法会对音乐的节奏、旋律、音色，歌词文本语义和歌手风格等属性展开分析，计算出音乐之间的相似度，以此来为用户推荐与已收听音乐特征相近的曲目；基于协同过滤的推荐算法依据用户群体的行为数据，去挖掘有相似音乐偏好的

用户，然后把相似用户喜爱的音乐推荐给目标用户；混合推荐算法结合了前两者的优势，借助加权融合、级联等方式弥补单个算法的局限性，从而提升推荐的准确性和多样性。

国内在推荐算法上注重混合模型与前沿技术的充分融合。杨建等人提出了一种混合算法，将协同过滤与内容过滤相结合，凭借对用户行为数据与歌曲特征的交叉分析来提升推荐多样性^[6]。这样做可依据用户历史行为推荐热门曲目，还可以挖掘出小众且符合用户潜在兴趣的音乐，为用户带来更多元化的音乐体验。QQ 音乐的深度神经网络模型借助多层特征提取与高维相似度计算，使得推荐准确率大大提高，同时引入注意力机制聚焦用户兴趣偏好，强化了推荐的精准度与个性化程度。国内研究还尝试把知识图谱与推荐算法结合^[7]，挖掘音乐实体间的关联关系，为用户提供更具解释性的推荐结果，提升推荐系统的用户体验。最后，值得一提的是生成式人工智能技术在推荐算法中的引入，这一举措带来了突破性的变革。凭借强大的内容生成能力，该技术可深度挖掘用户的个性化需求，并依据推荐曲目的风格特点创作独具匠心的专属歌单介绍，提升推荐服务的交互性与趣味性^[8]。

国外对推荐算法的研究侧重于先进技术架构的探索与跨领域算法的融合应用^[9]，目的在于突破传统推荐算法的限制。谷歌研究团队把图神经网络运用到音乐推荐中，借助其强大的图结构数据处理能力有效建模用户与音乐之间复杂的交互关系，大幅提升推荐性能。Sakurai Keigo ,Togo Ren 等学者提出结合深度强化学习与知识图谱的音乐推荐框架^[10]，解决传统推荐算法在冷启动和稀疏数据场景下的局限性。此项研究依靠构建一个包含歌曲声学特征的知识图谱，将音乐内容的语义信息转化为结构化的图数据。国外研究人员还把联邦学习用于音乐推荐^[11]，在保障用户数据隐私的情况下实现多源数据协同建模，探索推荐算法在数据安全环境下的优化路径。

不过在国内外现有的推荐算法体系中，评论情感资源仍未被充分发掘。拿网易云音乐、Spotify 等头部音乐平台来说，用户每天平均产生的评论数据量有数百万条之多，这些评论包含了对歌曲旋律、歌词的直观评价，也蕴藏着用户的深层情感倾向。遗憾的是，现阶段的主流推荐算法仍将重心放在用户行为数据与音乐元数据的分析上，像情感极性分析、语义情感挖掘这类自然语言处理技术，并没有系统性地融入推荐模型架构之中，导致大量承载用户真实喜好与情感共鸣的评论数据未能转化为提升推荐精准度与个性化程度的有效信息，需进一步研究。

1.3 研究内容

本文的主要研究内容如下：

（1）爬取网易云音乐的用户评论数据

设计并实现网易云音乐的 API 接口访问机制，涵盖了参数加密处理以及构建稳定的爬虫系统，采用设置请求频率来预防 IP 封禁这一潜在风险；之后爬取前 100 首热门歌曲以及每首歌各 200 条评论数据。

（2）歌曲评论数据预处理与情感分析

对用户评论进行预处理，包括分词、去停用词、特殊字符处理等，结合 SnowNLP 情感分析工具、情感词典匹配和情感短语识别的混合策略对随机抽取的部分评论进行情感标注，构建的情感词典包含着音乐领域常见的情感表达词汇。然后统计每首歌曲中不同情感评论的占比，获取歌曲的情感分布特征^[12]。

（3）机器学习模型评估与情感分类可视化

实现并评估朴素贝叶斯、支持向量机、随机森林三种经典机器学习算法在评论情感分类任务上的性能，借助混淆矩阵进行模型评估，并在系统中展示了依据最佳模型随机森林的分类结果^[13]。

（4）设计并实现推荐系统

按照 Flask 框架设计并实现 Web 应用程序，打造包含情感需求、情感强度以及多样性因子的三层匹配机制，根据用户当前情绪状态和需求，结合歌曲的情感类别开展个性化推荐。系统还实现了数据分析与可视化功能，像歌曲评论词云图、不同情感词云图等，使推荐结果更加直观清晰。

1.4 论文结构

本论文共分为六章，各章内容安排如下：

第一章为绪论，介绍研究背景与意义、国内外研究现状及主要研究内容。

第二章为相关技术介绍，详细介绍该音乐推荐系统所使用的关键技术，并深入讲解相关推荐算法的设计流程与原理，为系统实现提供理论支撑。

第三章为系统分析，对系统的用户需求和功能需求进行系统性梳理，为后续系统的设计与实现奠定需求基础。

第四章为系统设计与实现，对系统架构进行整体规划，明确各功能模块设计细节及交互逻辑，运用 Flask 框架搭建 Web 应用，设计直观的用户情绪输入界面与推荐结果展示页面保障系统的运行和良好的用户体验。

第五章为测试与评估，评估系统性能，分析不同算法的优缺点，并展示测试结果。

第六章为结论，总结研究成果，分析系统不足，提出未来改进方向。

第 2 章 相关技术介绍

本章主要介绍音乐情感分析与推荐系统所使用的关键技术，同时还涉及到这些技术详细的实现过程以及原理，为后续系统设计与实现提供了理论基础。

2.1 自然语言处理技术

为了正确判断各条评论数据的情感倾向，先对评论数据进行文本预处理，主要包括以下几个步骤：

（1）中文分词：采用 Jieba 分词工具，并融合自定义评论词典，纳入像“好听”“喜欢”等蕴含情感倾向的词汇，为后续的情感分析奠定了准确的词语基础。

（2）停用词过滤：设置了一份专属于音乐领域的停用词表，去掉“的”“了”等无实际意义且无情感倾向的常见高频词汇，从而提升文本特征的有效性与纯净度，使后续分析可以着力在真正有情感倾向的词汇。

（3）文本向量化：应用 TF-IDF 算法来计算每个词汇的权重，公式如下：

$$TF - IDF(t, d) = TF(t, d) * \log\left(\frac{N}{DF(t)+1}\right) \quad (2-1)$$

其中 N 为总文档数，DF(t)为包含词 t 的文档数，并通过参数控制（`max_features=2000`，`min_df=2`，`max_df=0.95`）来降低特征维度、减少稀有词和高频词的干扰。

2.2 评论与歌曲情感判定规则

2.2.1 单条评论情感判定规则

本研究采用 SnowNLP 情感分析、情感词典匹配和情感短语识别的混合策略对单条评论进行情感标注。判定规则如下：

- （1）使用 SnowNLP 获取评论的情感得分作为基础分数；
- （2）借助情感词典匹配识别评论中的情感词，统计积极词和消极词的数量；
- （3）结合情感短语识别，捕捉特定的情感表达模式，短语匹配的权重为词语匹配的 2 倍；

具体来说：

如果积极得分大于消极得分，或有积极词且 SnowNLP 得分 >0.6 ，判定为积极

如果消极得分大于积极得分，或有消极词且 SnowNLP 得分 <0.4 ，判定为消极

对于长度小于 10 个字且无明显情感词的评论，判定为中性，其他情况判定为中性

在此基础上，训练并评估了三种机器学习算法，后采用最佳的推荐算法对单条评论进行情感预测，具体步骤如下：

（1）特征提取：使用 TF-IDF 向量化方法将评论文本转换为特征向量，并设置最大特征数为 2000，去除低频词（ $\text{min_df}=2$ ）和高频词（ $\text{max_df}=0.95$ ），之后对特征向量进行归一化处理

（2）模型训练与预测：使用规则判定的结果作为训练标签（积极、中性、消极），采用随机森林分类器进行训练，并输出每条评论属于三种情感的概率分布情况

（3）预测结果处理：获取模型预测的情感类别，同时保留预测的概率值，用于后续歌曲情感分布的计算

2.2.2 评论与歌曲情感映射规则

为了将单条评论情感标签（积极/中性/消极）整合为歌曲整体的情感特征，设计了歌曲主导情感的判定规则，旨在依据评论词汇的情感分布来精准刻画歌曲的情感倾向，具体流程如下：

（1）情感分布统计：使用随机森林模型对每条评论进行情感预测，获取每首歌曲所有评论的情感概率分布，计算积极（ P_{pos} ）、中性（ P_{neu} ）和消极（ P_{neg} ）情感的平均概率；

（2）阈值判定规则：若积极情感占比 $P_{\text{pos}} \geq 0.45$ ，判定为积极情感；若消极情感占比 $P_{\text{neg}} \geq 0.20$ 或超过积极情感比例，判定为消极情感；若中性情感占比 $P_{\text{neu}} \geq 0.50$ 或积极与消极情感差异小于 15 个百分点，判定为中性情感；其他情况下，取概率最高的情感类别作为主导情感。

该规则充分考量歌曲评论情感分布特点，借助机器学习模型的概率预测以及合理的阈值设置^[14]，将单条评论的情感预测整合为歌曲的整体情感特征。系统考虑了各类情感的占比情况，又引入情感差异的阈值判断，让情感分类结果更加客观准确。这种方法为后续基于情感的音乐推荐提供了可靠的数据支撑，能够满足用户在不同情感状态下的音乐需求。

2.3 相关机器学习算法

本系统采用三种经典机器学习算法构建歌曲情感分类模型，并比较各算法性能指标来确定最佳方案。这些算法在情感分析领域有着广泛的应用，接下来会详细阐述各算法的基本原理与实现特点。

（1）朴素贝叶斯

基本原理：

朴素贝叶斯基于贝叶斯定理构建概率分布模型^[15]，其核心公式为：

$$P(y|x) = \frac{P(y) \prod_{i=1}^n P(x_i|y)}{P(x)} \quad (2-2)$$

其中 y 表示情感类别， $x=(x_1,x_2,x_3)$ 为歌曲情感分布特征向量（积极/中性/消极评论的比例）。算法通过最大后验概率（MAP）决策规则进行分类：

$$\hat{y} = \arg \max_y P(y) \prod_{i=1}^3 P(x_i|y) \quad (2-3)$$

实现特点：

在特征输入环节，歌曲情感分布向量涵盖了积极、中性以及消极评论的比例，能直观地呈现出歌曲的情感构成，为模型训练提供了清晰的数据基础。在应用场景上，该算法适用于特征相对独立的场景，具备计算轻量高效的优势，可快速给出分类结果，满足实时性要求较高的应用需求。在优化配置方面，采用 MultinomialNB 来实现，适合处理离散化的情感分布数据，在文本分类等任务中表现出色，能够有效的提升分类效率与准确性。不过朴素贝叶斯也有局限性，它无法捕捉情感分布间的复杂非线性关系，当特征之间存在强相关性或数据中存在复杂的模式时，模型的分类性能可能会受到一定限制。

（2）支持向量机

基本原理：

支持向量机通过寻找最优超平面将不同类别的数据分开，目标是最大化类别间的间隔^[16]。对于线性不可分的情况，SVM 引入核函数（kernel function）将低维特征映射到高维空间，使数据线性可分。优化目标函数为：

$$\begin{aligned} \min_{w,b} & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t.} & y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \end{aligned} \quad (2-4)$$

其中 C 为惩罚参数，为了控制模型的复杂度与训练误差的平衡。

实现特点：

在应用配置环节，SVM 采用线性核函数来处理高维特征空间，同时设置了 $C = 1.0$ 及 `class_weight = 'balanced'` 这两个参数，依靠这种方式应对复杂的情感分类任务，同时平衡不同情感类别，进而提升对少数类情感的识别能力。它的优势在于在高维空间里表现得比较稳定，可对特征间的弱相关性进行有效处理。即便所涉及的特征数量很多，它也可以对不同情感类别的特征空间加以划分，保证分类结果有稳定性与可信度。在决策边界方面，SVM 构建最优超平面划分不同情感类别的特征空间，支持多类情感分类任务，决策边界清晰，为分类任务提供了不错的可解释性。

（3） 随机森林

基本原理：

随机森林是一种集成学习方法，借助构建多棵决策树并把多数表决的结果当作最终预测^[17]。其关键步骤包括：

自助采样：从原始训练集中有放回地随机抽取样本，以此来构建多个子训练集合，这样做可增加样本的多样性，减少出现过拟合的风险。

随机特征选择：在训练每棵决策树时，从所有特征中随机挑选出子集用于节点分裂，这回使得每棵树只能基于部分特征进行决策，进一步降低特征间的相关性，提高模型的泛化能力。

决策树构建：基于信息增益或基尼不纯度等指标构建决策树，这两个指标可以有效地衡量特征对于分类任务的贡献程度，帮助模型选择最优的特征进行分裂，构建出高效的决策树。

集成预测：将所有决策树的预测结果借助投票机制集成成最终结果，这种集成方式可综合多棵决策树的智慧，减少单棵决策树可能存在的偏差，提高分类结果的稳定性和准确性。

实现特点：

随机森林算法的实现有多方面的创新与优势。借助集成学习方法，使用 100 棵决策树进行协同预测，有效提升模型的稳定性和准确性。每棵决策树凭借随机特征选择和样本抽样，可处理情感特征间的非线性关系和交互效应，相比于线性模型更能捕捉复杂情感数据中的隐藏模式。依靠集成多棵决策树的预测结果，减少了单棵决策树带来的随机性和不确定性，从而提高分类的准确性和泛化能力。

2.4 情感匹配规则

本系统结合用户需求，采用基于用户情绪需求与歌曲情感匹配的推荐规则。推荐系统根据用户的情绪强度（积极/消极/中性）和需求类型（鼓励/舒缓/共鸣）为用户推荐最匹配的歌曲。系统把用户的情绪强度 E_u 和需求类型 N_u 作为输入，结合歌曲的情感特征 E_i ，通过以下策略计算匹配分数：

$$Score(u, i) = w_1 * BaseScore(E_u, E_i) + w_2 * DiversityFactor \quad (2-5)$$

其中：

$Score(u, i)$ 为用户 u 对歌曲 i 的最终匹配得分； $DiversityFactor$ 为引入的有着随机性的多样化因子； $BaseScore(E_u, E_i)$ 为基础匹配分数，根据不同需求类型采用不同的计算策略：

- (1) 用户需求为鼓励时，选择类别为积极的歌曲： $BaseScore =$ 积极情绪百分比
- (2) 用户需求为舒缓时，选择类别为中性的歌曲： $BaseScore =$ 中性情绪百分比
- (3) 用户需求为共鸣时，系统首先根据用户情绪强度选择匹配的情感类别：

即当用户情绪强度 $\geq 60\%$ 时，仅匹配主情感类别为积极的歌曲；用户情绪强度 $\leq 30\%$ 时，仅匹配主情感类别为消极的歌曲；用户情绪强度在 $30\%-60\%$ 之间时，仅匹配主情感类别为中性的歌曲。对于情感类别匹配的歌曲，当情绪强度差异 $\leq 20\%$ 时： $BaseScore = (1 - |用户情绪强度 - 歌曲情感强度|) * 0.5$

经过实验，设置了最优权重系数 $w_1=0.95, w_2=0.05$ ，并通过最后的匹配分数确定推荐歌曲的匹配度。在保证情感匹配准确性的同时增加了推荐结果的多样性。这样系统就能够综合考虑情感类别、用户需求和推荐多样性三个维度，为用户生成更加符合情绪需求的音乐推荐列表。

2.5 本章小结

本章详细介绍了歌曲情感分析与推荐系统的相关技术基础。阐述了自然语言处理与情感分析的基本概念与理论基础，涵盖文本预处理、特征提取等关键技术；分析了歌词与评论文本的特殊性，为系统设计提供理论支撑。深入探讨了三种机器学习算法在情感分类中的应用原理与实现方法。最后，介绍情感匹配规则，即综合歌曲情感类别、用户情绪状态和用户需求三个维度将歌曲合理的推荐给用户。这些核心技术的深入研究为系统设计与实现奠定了坚实的理论基础。

第 3 章 系统分析

3.1 需求分析

3.1.1 用户需求分析

在数字化音乐消费场景中，用户对音乐推荐的需求呈现出显著的情感化与个性化特征。传统音乐推荐系统只是依赖用户历史行为数据或音乐基础特征，没办法精准匹配用户实时情感需求，使得推荐结果与用户心理预期存在着差异。依靠对网易云音乐用户评论以及国内外音乐系统的现状分析，总结得出以下三点用户需求：

情感分析需求：用户期望在特定情绪状态下，可获取与自身情绪相契合或具有情绪调节作用的音乐推荐。比如说，当用户处于低落情绪状态，可选择治愈的歌曲来获得鼓舞，也可挑选舒缓的歌曲以平复心情。这就要对系统中的歌曲进行情感分析，给出歌曲的情感类别。本系统要分析网易云音乐评论中的情感特征，构建包含积极、消极和中性三种情感类别的分类模型，以准确识别歌曲评论中的情感倾向，把评论中的情感映射到歌曲中，依据用户评论来定义歌曲类别，找出歌曲所涵盖的情感特性。

个性化推荐需求：用户在音乐推荐系统里希望满足情感的匹配，获得更为个性的推荐结果。本系统应采用三层推荐机制，借助用户情感强度、情绪状态以及多样性因子综合推荐歌曲，在考虑歌曲类别、用户情感以及用户需求的同时，增加推荐的随机性，确保推荐结果的个性化及多样性，更加精准的满足用户需求。

交互便捷性需求：在追求高效与即时反馈的数字时代，用户对音乐推荐系统的交互体验有了更高要求。一方面，系统要提供简洁直观的交互界面，支持用户使用滑块及按钮等便捷方式快速表达自身需求；另一方面，系统还要包含清晰的操作说明与可视化结果展示，降低用户使用门槛，提升整体交互体验的流畅性与友好性。

3.1.2 功能需求分析

本系统的功能主要围绕情感分析、个性化推荐以及交互便捷性三个方面展开。系统需对从网易云音乐平台获取的热门歌曲以及对应的用户评论数据进行预处理，实现中文分词、去除停用词、特殊字符清洗等操作，并利用 TF-IDF 方法对文本进行向量化处理，以便于机器学习模型的训练和评估。

在情感分析方面，系统要集成多种主流机器学习算法，包括朴素贝叶斯、支持向量机和随机森林，采用最佳的算法对用户评论进行积极、中性、消极三种情感分类，并计算情感概率分布和情感强度，提升情感识别的准确性。依据每首歌曲不同情感的评论所

占比例，运用自定义的阈值规则，剖析出歌曲的情感类别。基于情感分析所得到的结果，在知晓歌曲类别的情形下，系统可依照用户当下的情绪状态以及需求，设定多样性因子，智能化地推荐与用户情感需求相契合的歌曲，达成个性化推荐功能，推荐功能要考虑情感匹配，而且要结合用户当下的情绪需求，及时地推荐契合用户情绪状态的歌曲，保证推荐内容有相关性与多样性。系统还需要提供便利的交互功能，设置丰富多样的可视化展示界面，比如生成情感词云图、模型混淆矩阵以及性能对比图等，帮助用户直观地理解分析结果与推荐依据，系统界面应当简洁且友好，支持用户方便地输入情感需求、浏览推荐结果。

经过上述功能的有机结合，系统能够为用户提供高效、智能、个性化的音乐情感推荐服务，满足用户在数字化音乐消费场景下的多样化需求。

3.2 可行性分析

3.2.1 经济可行性分析

本系统在开发以及运行过程中主要依赖于 Python 语言及其成熟的开源生态体系，包括 pandas、scikit-learn、jieba、matplotlib 等主流数据处理与机器学习库，这些工具都是免费开源软件，不需要额外的软件采购费用，系统的开发与部署可在普通的个人计算机或实验室服务器上完成，对硬件配置要求不高，不需要高昂的硬件投入。数据采集主要基于公开的网易云音乐平台，避免了数据购买等额外开销。后期维护和升级主要依赖开源社区的持续支持，维护成本低廉。系统的开发、测试与维护成本较低，经济投入小，有良好的经济可行性。

3.2.2 操作可行性分析

本系统采用模块且人性化的设计理念，打造出低门槛、高交互性的操作体系，用户不需要具备专业的编程或数据分析背景即可顺利使用。系统界面友好，支持用户通过滑块、按钮等直观方式输入情感需求，大幅降低了操作门槛。数据预处理、情感分析、模型训练以及个性化音乐推荐等核心流程都达成了自动化。用户只要按照提示完成简单的操作，就能获取个性化音乐推荐结果。系统还整合了丰富的可视化功能，像情感词云图、模型混淆矩阵等，帮助用户直观理解分析和推荐结果，提升系统的可用性和用户体验。本系统有较高的操作可行性，可满足不同层次用户实际使用需求，拥有良好推广及应用前景。

3.2.3 技术可行性分析

本系统在技术实现层面有很高的可行性，系统选用 Python 作为开发语言，拥有丰富且成熟的数据处理、自然语言处理以及机器学习库，可有效支持数据采集、文本预处理、特征提取、模型训练与评估以及结果可视化等各个环节。情感分析与推荐部分采用朴素贝叶斯、支持向量机、随机森林等主流算法，这些算法在学术界以及工业界都有广泛应用，相关实现和调优方法成熟可靠。系统的数据预处理融合中文分词、停用词过滤、正则表达式清洗等多种技术手段，提升了情感分析的准确性。特征工程采用 TF-IDF 等成熟方法，将文本数据转变为适合机器学习模型处理的数值特征。模型训练与评估支持多种算法的对比与选择，还可以自动输出多项评估指标，方便全面分析模型性能。可视化模块利用 matplotlib 和 wordcloud 等工具，生成直观的图表和词云，提升结果的可解释性以及用户体验。系统整体架构清晰，模块间耦合度低，便于后续功能扩展和维护。依靠成熟的技术框架和丰富的开源资源，本系统在技术实现上完全可行，可高效稳定地完成音乐情感分析与个性化推荐的各项任务。

3.3 本章小结

本章围绕音乐情感分析与推荐系统的实际应用需求，系统分析了用户在情感共鸣、个性化推荐以及交互便捷性等多个方面的核心诉求，并据此提出相应的功能需求。系统要达成对评论数据的高效预处理以及多模型情感分类，还需要结合用户当下的情绪状态，智能生成个性化的音乐推荐结果，并借助可视化手段提升结果的直观程度和交互体验，为后续系统设计与实现提供了明确的目标和方向。

第 4 章 系统设计与实现

本章详细阐述基于情感分析的音乐推荐系统的总体架构设计、核心功能模块实现以及关键技术难点突破，展示系统从设计理念到具体实现的完整过程。

4.1 系统总体架构

4.1.1 设计思路

本系统采用了模块化的设计思想，把音乐情感分析与推荐任务分为数据采集、情感分析、歌曲推荐和用户交互四大核心模块。各模块既相互独立又紧密协作，整体架构如图 4-1 所示：

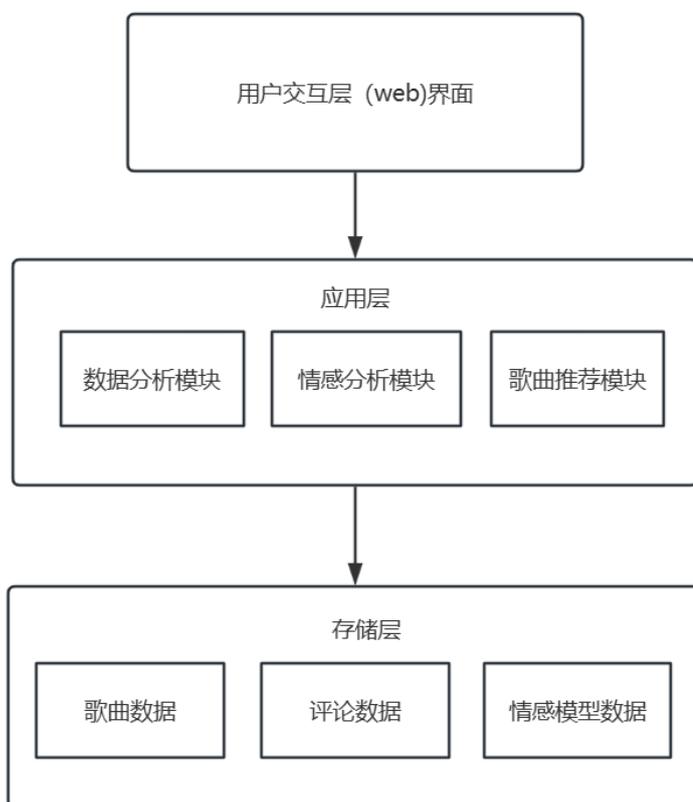


图 4-1 系统架构图

该架构遵循着以下设计原则：

- (1) 模块化设计：功能划分清晰，便于独立开发、测试与维护。
- (2) 松耦合高内聚：模块间借助明确定义的接口交互，降低依赖。
- (3) 数据流驱动：各组件根据数据处理流程协作，形成完整链路。
- (4) 可扩展性：预留扩展接口，利于后续功能拓展与算法优化。

4.1.2 技术栈选择

各层级技术栈如下：

（1）用户交互层：

前端框架：Flask-Bootstrap

页面渲染：Jinja2 模板引擎

交互组件：JavaScript、jQuery、AJAX

（2）应用层：

Web 框架：Flask

情感分析：Scikit-learn、NLTK、Jieba

数据处理：Pandas、NumPy

可视化：Matplotlib、Seaborn

（3）数据存储层：

文件存储：CSV 格式数据文件

内存缓存：应用内缓存机制

技术选型全面考虑系统需求、开发效率与性能等多方面因素，其中轻量级的 Flask 框架可对响应速度起到保障作用，而丰富的机器学习库则可为情感分析以及推荐算法提供支持。

4.2 核心功能模块实现

4.2.1 数据采集模块

本研究打造了一套数据采集模块，从网易云音乐平台获取热门歌曲以及评论数据，为后续的情感分析和歌曲推荐提供可靠的数据基础。该模块采用多线程爬虫架构来保证数据采集的高效且精准。具体来说，先借助 API 获取包含歌曲 ID、歌名、歌手等关键信息的热门歌曲榜单数据，再针对每首歌曲获取其中的各 200 条用户评论，并把用户 ID、评论内容等相关数据记录下来，最后过滤无效数据并统一数据格式，把处理后的数据存储成 CSV 文件。在数据采集过程中运用自适应爬取频率和请求等待原则来避免反爬虫限制，提高了数据采集的效率与稳定性。完成了对 100 首热门歌曲的爬取任务，每首歌曲采集 200 条评论数据，形成一个含有 20000 条评论的数据集，如图 4-2 所示。



图 4-2 部分评论数据

4.2.2 情感分析模块

该模块为系统的核心模块，主要功能是对采集到的评论数据进行深度的情感倾向分析，并建立起歌曲情感分类模型。在情感分析过程中，对评论文本进行了一系列预处理操作，包括分词、停用词过滤和特征提取等关键步骤。预处理结果如图 4-3 所示。



图 4-3 部分预处理数据

运用自定义的词典规则以及 TF-IDF 特征提取技术把处理后的评论数据划分为积极、中性和消极三种不同的情感倾向类别。在完成分类之后统计分析每首歌曲的情感比例分布情况，生成一种可以量化反映歌曲情感特征的情感分布向量。本研究结合音乐领域情感词典并采用依据词典的情感分析方法，使得情感分析的准确性以及专业性得以提高。依据评论情感分布特征，训练并评估朴素贝叶斯、支持向量机和随机森林三种主流的机器学习模型。经过模型训练与评估流程，随机森林在混淆矩阵上所展现的精确率、准确率、召回率等最高，被确立为系统的最佳分类器。

4.2.3 歌曲推荐模块

该模块作为系统核心，分析用户当前情绪并构建歌曲情感推荐模型。具体来说，系统首先对歌曲库中的每首歌曲进行情感预测。预测过程中，随机森林模型基于歌曲评论的情感分布特征即积极、中性和消极评论百分比进行分类，结合阈值设置将歌曲划分为积极、中性和消极三种情感类别。随后按上文设置的匹配规则为用户推荐歌曲，依据匹配规则计算每首歌曲与用户情感和需求的匹配得分，并按匹配得分对歌曲进行降序排序，从得分最高的前 20 首歌曲中随机选择 5 首作为最终推荐列表。这种基于实时情感需求的推荐策略突破了传统协同过滤依赖历史行为数据和基于内容推荐依赖音频特征的局限，直接响应用户的情感状态变化，提供更加个性化音乐推荐服务。

4.2.4 用户交互模块

该模块为用户提供清晰的交互体验，让用户可以便捷地输入情绪需求并获取个性化推荐。运用响应式布局设计了包括首页、歌曲详情页、歌曲分类页、歌曲推荐页以及数据分析与可视化页在内的多个用户界面。其中歌曲推荐页面以简洁友好的设计风格呈现，用户可凭借滑动条表达当前情绪强度，还可以点击按钮来表达情绪需求，这样系统就可以根据用户当前的情绪状态针对性地推荐歌曲。后端则基于 Flask 框架构建，严格遵循 RESTful 原则设计主要接口^[18]，保障了系统的易用性和可维护性。这一模块达成了用户情绪和需求的准确输入和个性化推荐的直观展示，并借助数据分析与可视化功能，强化了用户对推荐结果的理解和信任。相关用户界面如图 4-4 至图 4-8 所示。



图 4-4 首页界面

首页界面主要介绍系统的功能和使用方法，并列举了五首系统所涵盖的推荐歌曲，底部设置的两个按钮分别为“寻找适合您情绪的音乐”和“查看热门歌曲库”，点击按钮可以分别到达歌曲推荐页和歌曲详情页，方便用户进行进一步操作。

热门歌曲库

搜索歌曲、歌手...

网易云音乐热门歌曲 Top 100

排名	歌曲	歌手	专辑
1	唯一	G.E.M.邓紫棋	T.I.M.E.
2	爱错	王力宏	恋爱占星音乐全精选
3	春雪	裘德	春雪
4	于是	郑润泽	于是
5	跳楼机	LBI利比	跳楼机
6	罗生门 (Follow)	梨冻紫	罗生门 (Follow)
7	你	郑润泽	你
8	如果爱忘了 (live)	汪苏泷	声生不息·家年华 第12期
9	我爱你但是我要回家	ET	我爱你但是我要回家
10	有些	颜人中	失眠症候群
11	特别的人	方大同	危险世界
12	还是会想你	林达浪	还是会想你
13	把回忆拼好给你	王贰浪	把回忆拼好给你

图 4-5 歌曲详情界面

该界面按照排行榜一一显示了网易云音乐的前 100 首热门歌曲以及它们的歌手信息、专辑信息。并设置搜索功能，用户可以在右上角的搜索框中搜索想听的歌曲。

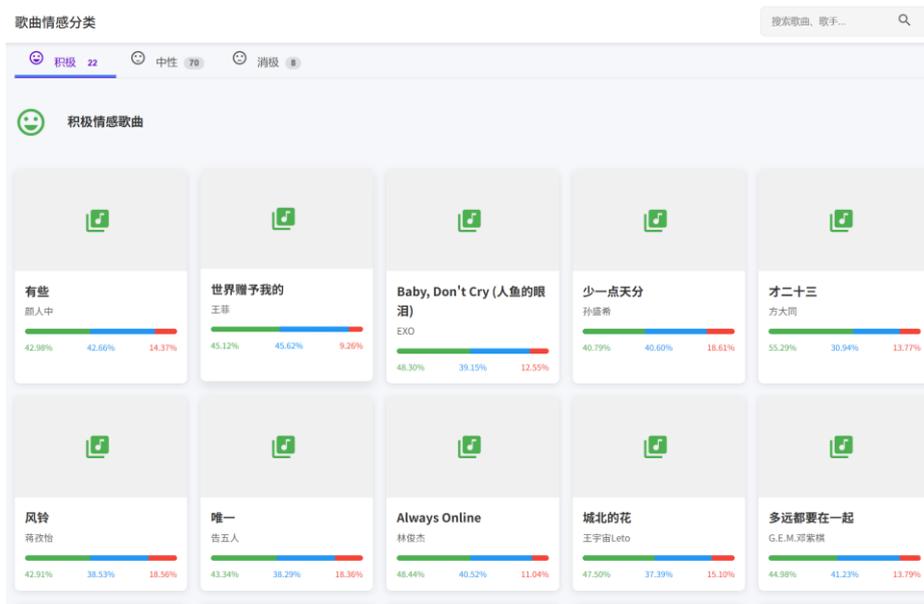


图 4-6 歌曲分类界面

该页面采用经模型评估确定的最佳机器学习算法随机森林对评论数据进行情感预测，后结合自定义的评论与歌曲映射规则将歌曲的类别划分为积极、中性、消极三类，为后续的歌曲推荐功能奠定了基础。



图 4-7 歌曲推荐界面

该页面展示了用户情绪与歌曲匹配的主要窗口，用户可以根据自己当前的情绪状态，利用滑块输入自己的情感强度，并借助下面三个选择按钮来表明自己的情感需求，最后点击获取推荐歌曲按钮来匹配适合自己情绪需求的歌曲。

20 首歌曲中随机选择指定数量的 5 首歌曲作为最终推荐结果，强化了推荐的多样性。这种多层次的策略设计保证了推荐结果与用户情感需求的匹配度，又避免了推荐结果的单一化。

4.4 本章小结

本章详细介绍基于情感分析的音乐推荐系统的设计与实现过程。系统采用模块化架构把复杂任务分解为四大核心模块，各模块协同工作形成完整的处理流程。在情感分析上结合单条评论情感判定规则和机器学习技术构建高性能的歌曲情感分类模型；在推荐引擎方面创新性地设计情感匹配算法，可依据用户当前情绪状态和需求智能推荐匹配的歌曲；而对于用户交互则提供了直观友好的 Web 界面实现情绪输入和个性化推荐展示。

第 5 章 测试与评估

5.1 评价指标

为了全面评估系统性能，本项目采用了以下评价指标：

5.1.1 情感分析算法评价指标

- (1) 准确率：正确分类样本数占总样本数的比例，直观反映模型整体分类效果
- (2) 精确率：预测为正例的样本中真正例的比例，评估模型对正例预测的准确性
- (3) 召回率：真正例中被正确预测为正例的比例，评估模型捕获所有正例的能力
- (4) F1 分数：精确率和召回率的调和平均值，兼顾精确率与召回率的综合指标
- (5) 混淆矩阵：直观展示分类结果的正确与错误情况与各类别之间的误分情况

5.1.2 推荐系统评价指标

- (1) 基础匹配分数：依据用户情绪需求来计算匹配程度，分数范围在 0 至 1 之间。
- (2) 多样性因子：取值范围在 0.95 至 1.0 的随机值，用于增加推荐结果的多样性。
- (3) 最终推荐分数：综合基础匹配分数和多样性因子所得到的评分，用于歌曲排序和筛选。

5.2 情感分析模型评估

5.2.1 模型训练与测试数据

本研究把网络音乐平台所收集的歌曲评论数据作为情感分析的基础数据集。该数据集在完成预处理后，包含 13708 条评论，这些评论涉及 100 首热门歌曲，平均每首歌曲大约有 137 条用户评论。为保证模型评估的准确性与结果的可靠性，运用分层抽样的方式把数据集按照 8:2 的比例划分为训练集与测试集，同时设置固定的随机种子（`random_state=42`）来保证实验的可重复性。分层抽样策略确保训练集和测试集中的情感类别分布保持一致，避免了数据划分过程中可能出现的类别不平衡问题。情感类别分布如表 5-1 所示：

表 5-1 评论情感类别占比

情感类别	评论数量	百分比
积极	5414	39.50%
中性	6342	46.26%
消极	1952	14.24%
总计	13708	100%

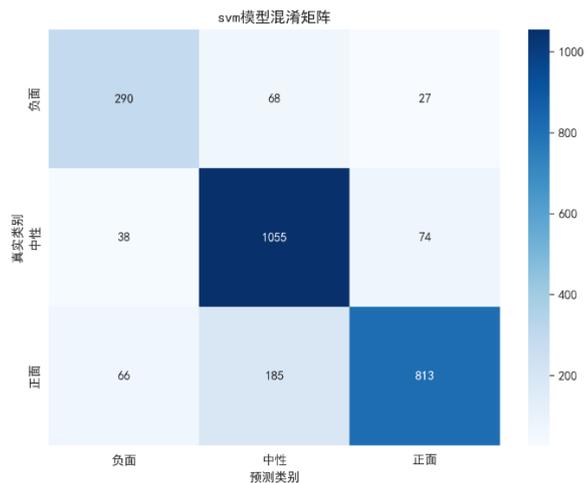


图 5-5 支持向量机

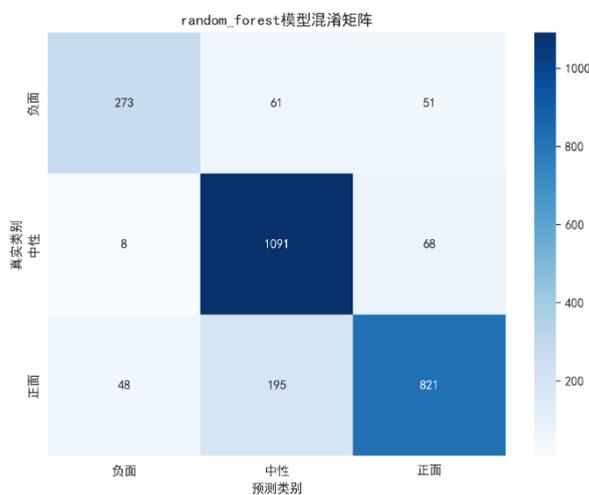


图 5-6 随机森林

为了更好的评价模型分类性能，选取三种模型整体分类的准确率、精确率、召回率、F1 分数作为评价指标，结果如表 5-2 所示。

表 5-2 模型性能评估指标

指标	准确率	精确率	召回率	F1 分数
朴素贝叶斯	0.7324	0.7598	0.6419	0.6620
支持向量机	0.8249	0.8107	0.8071	0.8064
随机森林	0.8352	0.8377	0.8052	0.8173

通过对三种机器学习模型的混淆矩阵分析，我们可以深入了解各模型在音乐评论情

感分类任务上的表现特点。

（1）模型性能对比分析

随机森林模型表现出色，总体准确率达到 0.8352，是三个模型中最高的；精确率为 0.8377，显示出较高的预测准确性；召回率为 0.8052，表明有较好的检出能力；F1 分数为 0.8173，说明模型在精确率和召回率之间取得了很好的平衡。

朴素贝叶斯模型表现不均衡，总体准确率较低为 0.7324；精确率为 0.7598，说明预测的准确性有待提高；在消极情感识别的召回率极低，说明在处理消极情感这类少数类别时能力较弱。

支持向量机模型表现中等，总体准确率达到 0.8249，仅略低于随机森林；精确率为 0.8107，显示出稳定的预测能力。整体性能表现良好。

（2）关键问题分析

类别不平衡问题：

数据集中消极情感评论仅占 14.24%，存在明显的类别不平衡问题^[19]。支持向量机模型通过设置 `class_weight='balanced'` 参数较好地处理了这个问题，在保持 0.8249 的总体准确率的同时，对少数类也有不错的识别能力；朴素贝叶斯在处理消极情感这类少数类别时表现最差，召回率仅为 0.6419，说明对少数类的检出能力较弱；随机森林虽然总体准确率最高，但在类别不平衡问题的处理上仍有提升空间，可以考虑通过调整类别权重或采样策略来进一步改善对少数类的识别能力。

情感边界识别问题：

中性情感与其他类别之间存在边界模糊问题。随机森林在中性与积极情感的边界区分上表现较好，各类别的精确率和召回率都较为平衡。而朴素贝叶斯的边界区分能力最弱，特别是在消极——中性边界。这种情感边界模糊性反映了自然语言表达的复杂性，尤其在音乐评论中，用户常使用模糊情感表达。未来可引入模糊分类法，以提高边界识别精度。

算法特性与任务适配性：

各算法在情感分析任务中展现出不同的优势。支持向量机在处理高维特征空间和查找最优分类边界方面表现稳定。随机森林的决策树集成特性使其能够有效捕捉评论中词汇与情感的复杂非线性关系。而朴素贝叶斯尽管性能较弱，但训练速度快，适合快速迭代和原型开发。

综合分析，随机森林模型表现出色，主要原因有：随机森林借助集成 100 个决策树

来进行投票预测，每个决策树使用不同的特征子集展开训练，在提升预测稳定性和准确性的同时还提高了模型的泛化能力，可有效降低过拟合风险。在特征处理方面，该模型可以有效处理 TF-IDF 生成的高维稀疏特征，会自动进行特征选择，以此识别出对情感分类最为关键的词语特征，并且对特征的非线性关系有较好的建模能力。不过所有模型在处理消极情感识别时仍有改进空间，这可能需要通过数据增强或模型优化来提升。

5.3 系统总体评估

5.3.1 推荐系统性能

推荐系统所给出的评估结果显示，基于情感匹配的推荐方法能有效满足用户不同情绪状态下的音乐需求。接下来从情感匹配机制、推荐多样性两方面展开分析：

（1）情感匹配机制

系统依据用户不同的情感需求达成了三种差异化的推荐策略，其中鼓励策略会严格挑选积极情感歌曲，保证推荐歌曲的主情感类别是“积极”，将歌曲的积极情感占比当作基础匹配分数，借助阈值控制来保证推荐歌曲的积极情感强度；舒缓策略会优先推荐主情感类别为“中性”的歌曲，按照歌曲的中性情感占比来计算匹配度，以此帮助用户实现情绪平和的状态；共鸣策略会依据用户情绪强度实现三档匹配机制，在高强度时匹配积极情感歌曲，低强度时匹配消极情感歌曲，中等强度时匹配中性情感歌曲，并且控制情绪强度差异在 20% 以内，保证情感共鸣的精准性。

（2）推荐多样性与相关性平衡

在保证情感匹配的前提下，系统设计了多层次评分机制以平衡推荐的相关性与多样性，运用权重为 0.95 的基础匹配分数来保证推荐结果契合用户情感需求，同时引入权重为 0.05、取值范围在 0.95 至 1.0 之间的随机因子来提升推荐的多样性，这两个因子借助综合评分公式 2-5 加以整合，基于此采用“Top-20 随机抽取”的分层选择策略，先挑选得分最高的 20 首歌曲，接着随机抽取 5 首作为最终推荐结果。这种基于可量化指标的多层次推荐机制，保证了推荐结果与用户情感需求的基本匹配，还凭借随机因子和分层选择策略有效防止了推荐结果的单一化，为用户提供了相关又多样的音乐选择。

5.3.2 系统优势与不足

本系统在情感分析任务中展现出优势，尤其在积极和消极情感识别方面有着不错的表现。引入依据情绪状态的推荐策略，为用户提供了丰富的互动体验。不过系统也存有一些局限：当下歌曲库仅仅包含 100 首歌曲，评论数据集规模为 13708 条，可能导致长

期使用时推荐结果出现重复情况；在情感分类方面，中性情感与其他类别存在边界模糊的问题，随机森林模型对应中性情感的召回率远高于精确率，模型倾向于将样本分类为中性，而积极情感的精确率高但召回率较低，意味着有大量积极样本被错误分类为中性。未来可借助引入深度学习模型等手段进行优化^[21]。系统还没有充分考虑用户的历史偏好，歌词内容在情感分析中的潜在价值也没有得到充分利用^[22]，个性化推荐机制仍有待完善。

5.4 本章小结

本章对音乐情感分析以及推荐系统的实验结果展开详细的剖析。依据多项评价指标的综合评估，证实随机森林模型在歌曲情感分类任务上有优越性，相较于其他对比算法而言更为出色。基于该情感分析模型构建的推荐系统展现出良好的情感匹配能力。但实验结果也揭示系统在中性情感识别、歌曲库规模和个性化推荐等方面有改进空间。未来的研究方向将集中在扩充音乐库、优化中性情感识别算法以及引入基于用户历史偏好的个性化推荐机制上，提升系统的性能和用户体验。

第 6 章 结论

6.1 总结

在当今数字化时代的浪潮下，互联网音乐产业蓬勃发展，音乐已成为人们日常生活中不可或缺的精神食粮。随着网络技术的飞速进步，国内外各大音乐平台都在不断的创新优化，用户对音乐的需求也日益多样化和个性化。在此背景下，音乐推荐系统作为一种智能化的音乐推荐技术应运而生并受到广泛关注。它既能帮助用户节省寻找心仪音乐的时间，又能精准推荐符合用户口味的歌曲，提升用户的音乐体验。但传统的音乐推荐方式，像基于内容推荐或简单协同过滤，往往难以精准捕捉用户复杂多变的情感需求，推荐结果的个性化有待进一步提高。所以开发高效、精准的音乐推荐系统具有极为重要的现实意义。本研究从传统推荐方式的不足出发，致力于将用户的实时情绪与推荐系统建立联系。通过分析用户评论文本的情感，提出融合用户情绪状态和需求的混合推荐方法，成功实现了基于情感分析的个性化音乐推荐，为用户带来更贴合其情绪需求的音乐享受。本文的主要工作包括：

(1) 首先介绍音乐情感分析与推荐系统的研究背景和意义，并对国内外的发展现状与趋势进行概述，明确研究方向与研究内容。

(2) 针对音乐评论文本的情感分析技术，本文达成了完整的情感分析流程，先是构建评论文本的获取机制，借助网络爬虫技术获取音乐平台上的用户评论，随后对中文文本开展预处理，包括分词、去停用词以及特征提取。接着比较朴素贝叶斯、支持向量机和随机森林三种情感分类模型的优缺点，最终选定随机森林模型作为系统的核心分类器，依据情感分析结果，为每首歌曲构建情感特征，包含积极、中性和消极情感的分布比例，

(3) 基于情感特征设计并实现情感导向的推荐算法，系统依照用户当前的情绪状态，结合不同情感需求，实现三种差异化的推荐策略，使用户在消极状态下能获得鼓励，又能找寻情感共鸣，智能调节推荐侧重点，在维持情绪和调节情绪之间实现平衡。

(4) 开发完整的音乐情感分析与推荐系统 Web 应用。该系统涵盖数据采集、情感分析、可视化展示和个性化推荐等功能模块，还提供直观的用户界面，支持热门歌曲浏览、歌曲情感分析、情感词云展示以及基于情感的个性化推荐，为用户提供全新的音乐探索体验。

6.2 展望

虽然本研究已经收获了一定的成果，但仍存在着进一步优化和拓展的空间。未来研

究主要从以下几个方向改进：

（1）情感分析技术深化：引入像 BERT 这类深度学习模型的预训练语言模型，提升情感分析的准确性和细粒度。同时探寻更为丰富的情感分类体系，从简单的三分类扩展到多维情感空间，更精细地描绘音乐情感特征。

（2）多模态特征融合：把评论文本分析与音频特征、歌词内容分析结合到一起，构建多模态的音乐情感表示。依靠融合不同来源的情感信息，提高情感特征的全面性与准确性，为推荐系统提供更为丰富的特征输入。

（3）动态用户情感建模：开发出可动态跟踪用户情感状态的机制，借助分析用户历史行为、当前选择以及实时反馈，构建时序化的用户情感模型，达成更精准的情感预测以及推荐调整。

（4）推荐策略优化：改进混合推荐算法，研究情感因素与传统特征的最优权重分配方式。设计更为复杂的推荐策略，如情感序列推荐与情境感知推荐，提高系统对用户情感需求的适应性。

参考文献

- [1] 余梦琴.个性化音乐推荐系统的设计与实现[D].华中科技大学,2022.DOI:10.27157/d.cnki.ghzku.2022.005003.
- [2] 王方圆,张国华.融合协同过滤的 XGBoost 在音乐推送上的应用研究[J].科技创新与应用,2024,14(11):49-52.DOI:10.19981/j.CN23-1581/G3.2024.11.012.
- [3] 李雯萱.中西方不同情绪音乐及色彩匹配感知研究[D].上海音乐学院,2024.DOI:10.27319/d.cnki.gsy.2024.000112.
- [4] 龚志,邵曦.基于多模态的音乐推荐系统[J].南京信息工程大学学报(自然科学版),2019,11(01):68-76.DOI:10.13878/j.cnki.jnuist.2019.01.013.
- [5] R A ,R.S S ,Deepika G , et al.AI based Music Recommendation system using Deep Learning Algorithms[J].IOP Conference Series: Earth and Environmental Science,2021,785(1):
- [6] 杨建,刘磊,康欣欣.基于用户行为和音频特征的音乐推荐系统设计与实现[J].《无线互联科技》,2023,20(09):55-57.
- [7] 李津.基于知识图谱的个性化音乐推荐系统设计与实现[J].《科学技术创新》,2024,(02):127-130.
- [8] 佟雪娜.生成式人工智能对音乐传播的变革探讨[J].福建论坛(人文社会科学版),2024,(02):44-53.
- [9] Awotunde, Joseph Bamidele, et al. "Personalized Music Recommendation System Based on Machine Learning and Collaborative Filtering." IEEE Xplore, 2023.
- [10] Keigo S ,Ren T ,Takahiro O , et al.[Paper] Deep Reinforcement Learning-based Music Recommendation with Knowledge Graph Using Acoustic Features[J].ITE Transactions on Media Technology and Applications,2022,10(1):8-17.
- [11] J. Sha, N. Basara, J. Freedman and H. Xu, "FLOR: A Federated Learning-based Music Recommendation Engine," 2022 International Conference on Computer Communications and Networks (ICCCN), Honolulu, HI, USA, 2022, pp. 1-2, doi: 10.1109/ICCCN54977.2022.9868921.
- [12] 赵京胜,宋梦雪,高祥,等.自然语言处理中的文本表示研究[J].软件学报,2022,33(01):102-128.DOI:10.13328/j.cnki.jos.006304.
- [13] 王钦炀,施水才,王洪俊.文本情感分析综述[J].软件导刊,2025,24(01):193-202.
- [14] 张忠林,吴挡平.基于概率阈值 Bagging 算法的不平衡数据分类方法[J].计算机工程与科学,2019,41(06):1086-1094.
- [15] 宋雨泽,温学兵.贝叶斯分类的研究及应用[J].理论数学,2023,13(10):3023-3029.
- [16] 岳朝霞,刘甲.SVM 分类理论与算法的若干新进展[J].应用数学进展,2021,10(12):4535-4541.
- [17] 向进勇,王振华,邓芸芸.基于随机森林算法的机器学习分类研究综述[J].人工智能与机器人研究,2024,13(1):143-152.
- [18] 于守健,魏聪,黄鹏.REST 风格的 Web 服务在移动终端的应用[J].计算机科学与应用,2014,4(8):144-151.

- [19] 姚世祎, 杨盛腾, 李裕梅. 基于混淆矩阵的机器学习分类评价指标研究及 Python 实践[J]. 数据挖掘, 2022, 12(4): 351-367.
- [20] 王晓. 加密流量数据集类别不平衡的研究[J]. 理论数学, 2024, 14(1): 23-33.
- [21] 尹岚清. 基于深度学习的多模态音乐情感识别的研究[D]. 东华大学, 2023. DOI:10.27012/d.cnki.gdhuu.2023.001350.
- [22] T N J , Alexander C , Jowan W , et al. Lyrical Trends: An Analysis of Music's (Mis) Use of Vertigo.[J]. Otology & neurotology : official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology, 2023, 44(6):

致 谢

行文至此，落笔为终。四年本科生涯即将落下帷幕，始于 21 年秋，终于 25 年夏，今年，我二十二岁。回首往昔，热泪盈眶。我品尝过成功的琼浆，也体会过失败的苦楚，我携稚嫩而来，满载成熟而归。四年的种种也许难做到一生铭记，却会一生感恩。

一朝沐杏雨，一生念师恩。能够顺利完成这篇论文，首先要感谢我的毕业论文指导老师肖雨腾老师，很幸运能在茫茫人海中成为您的学生，从论文选题到多次论文的修改中，每一步都离不开您的耐心指导与帮助，感谢您细心的帮我指出问题，给予修改意见，让我能够明确方向，不断完善论文。也感谢四年学习生涯中遇到的所有老师，感谢你们在学习和生活中给予的所有帮助。同时，也祝愿各位老师工作顺利，平安喜乐。

父母之爱子，则为之计深远。想起儿时父亲的谆谆教导，望向我时眼神里是盖不住的期许，您教会我读书识字，督促我好好学习，少时任性，竟从未写起过您，后来提笔，已是遗憾。那年您对我说的话仍历历在目，终有一天，我会完成您的夙愿，向您证明我可以！希望时空是个圆圈，若干年后，我们在世界尽头相聚。母亲，这么多年您一个人抚养我们姐弟俩长大，我人生中走过的每一段路，看过的每一道风景，都离不开您的托举，您的肩膀明明瘦小，却努力撑起了我的一片天！千言万语汇成一句谢谢，谢谢您最最最亲爱的母亲，谢谢您不舍昼夜的陪伴，谢谢您无微不至的关怀，谢谢您细致入微的照料，谢谢您不是超人，却为我变成了万能。

山河不重足，重在遇知己，愿岁并谢，与友长兮。感谢这些年来始终陪伴着我的好朋友们，因为有了你们，每一次相聚和重逢都有了意义。感谢一直以来的点点滴滴，孤独时的陪伴，难过时的安慰，每个自卑敏感的瞬间，都是友谊在拯救我。希望在奔赴未来的路上，我们都有着生生不息的热爱，如风自由，如星灿烂。

或许人生本身就是一场盛大的告别，我们在告别中挥手，再见又“再见”。于是，我告别校园，告别同窗，告别恩师，告别挚友，告别沿途中经过的所有风景，每一次晚风，每一场日落，每一阵欢笑，每一幕泪滴……因为我始终相信，告别是为了下一次的相遇，岁月悠长，我们终将再见！

最后，无论是否词不达意，仍想感谢那个走得很慢却一直在往前走的自己。十几载求学路，我看天地、见众生，也找自己。感谢无数个奋笔疾书的日夜，无数个含泪坚持的瞬间，知不足而奋进，望远山而前行。

我与我周旋久，宁做我！人生如旷野，阅己、越己、悦己。

附 录