

# 《沙丘2》影评的情感分析：基于Naïve Bayes算法的公众舆论翻译研究

马欣欣

上海海事大学外国语学院，上海

收稿日期：2025年6月18日；录用日期：2025年7月25日；发布日期：2025年8月7日

## 摘要

情感分析是自然语言处理领域最重要的分支之一。本研究基于Naïve Bayes算法，针对《沙丘2》影评进行情感分析，旨在揭示公众对影片的情感倾向，为电影的字幕翻译提供参考。通过对来自IMDb网站的1065条影评数据进行预处理、特征提取和情感分类，本研究实现95%的准确率、精确度和召回率。研究结果表明，Naïve Bayes算法在情感分类任务中表现出色，有效区分了正面、负面和中立的情感态度。这一分析不仅为电影研究领域提供了有价值的反馈，也为电影字幕制作和营销提供了实际的舆情参考。

## 关键词

情感分析，《沙丘2》，Naïve Bayes

# Sentiment Analysis of *Dune: Part Two* Movie Reviews: Study on Public Opinion Translation Based on Naïve Bayes Algorithm

Xinxin Ma

College of Foreign Languages, Shanghai Maritime University, Shanghai

Received: Jun. 18<sup>th</sup>, 2025; accepted: Jul. 25<sup>th</sup>, 2025; published: Aug. 7<sup>th</sup>, 2025

## Abstract

Sentiment analysis stands as one of the most crucial branches within the field of natural language processing. This study employs the Naïve Bayes algorithm to conduct sentiment analysis on reviews of *Dune: Part Two*, aiming to uncover public sentiment towards the film and provide a reference framework for its subtitle translation. By preprocessing, extracting features, and classifying the

sentiment of 1065 review samples sourced from IMDb, this research achieved high performance metrics of 95% accuracy, precision, and recall. The results demonstrate that the Naïve Bayes algorithm excels in sentiment classification tasks, effectively distinguishing among positive, negative, and neutral attitudes. This analysis not only offers valuable feedback for the film studies domain but also provides practical insights derived from public opinion for movie subtitle production and marketing strategies.

## Keywords

Sentiment Analysis, *Dune: Part Two*, Naïve Bayes

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

随着互联网和社交媒体的发展，公众舆论的形成和传播变得更加迅速和广泛。尤其是在电影行业，影评不仅影响观众的观影决策，也对电影的市场表现和文化影响产生深远的作用。影评不仅仅是电影质量的评价，它还反映了观众的情感倾向、社会文化背景以及对影片的情感共鸣。因此，分析影评中的情感态度，尤其是识别和提取出影评中的情感倾向，已成为电影字幕翻译领域的重要课题。

情感分析(Sentiment Analysis)作为自然语言处理(NLP)的一个重要研究方向，已广泛应用于各类文本数据的情感倾向识别。通过情感分析，研究者可以从海量的文本中提取出正面、负面或中立的情感信息，从而帮助决策者理解公众的情感反应和态度。在电影评论的背景下，情感分析能够揭示观众对电影的评价倾向，并为电影制片方、营销人员和研究者提供有价值的反馈。例如，通过情感分析，电影制片方可以了解观众对电影中的情节、角色和主题的反应，进而优化后续的制作或营销策略。

《沙丘 2》(*Dune: Part Two*)是丹尼斯·维伦纽瓦执导的史诗科幻电影，根据弗兰克·赫伯特的小说改编，续集继续讲述保罗·厄尔迪斯(Paul Atreides)的命运与反叛故事。作为一部备受期待的电影，《沙丘 2》自上映以来，吸引了大量影迷和评论者的关注。影评中，既有对电影视觉效果和叙事深度的赞扬，也有对影片节奏与情节设计的批评。这些情感反应不仅反映了观众的个人感受，也折射出公众对影片主题、情感表达和角色发展的集体态度。

本研究旨在通过情感分析，基于 Naïve Bayes 算法，探讨《沙丘 2》影评中的公众舆论。通过对影评数据的情感倾向分类(正面、负面和中性)，本文将揭示观众对影片的整体情感态度，并进一步分析不同情感类别背后的潜在因素。这一研究不仅为电影研究领域提供了新的方法论视角，也为电影产业提供了关于观众反馈和市场趋势的实用信息。

## 2. 文献综述

随着机器学习算法在情感分析中的应用日益广泛，其处理大规模数据集并有效分类情感的能力得到了充分的体现。传统模型如逻辑回归(Logistic Regression)和朴素贝叶斯(Naïve Bayes)因其简单性和计算效率，已被广泛用于文本分类任务。

目前，国内对情感分析的应用主要集中在商品评论、微博和新闻舆论三个领域。在商品评论方面，情感分析主要通过大数据挖掘和分析消费者对商品的评价，帮助消费者了解其他用户的态度倾向，从而优化购买决策；同时，生产商也可以利用这一分析了解消费者对产品的反馈，进而改进产品质量和服务

[1](张紫琼等, 2010: p. 84)。在舆情领域, 情感分析模型主要用于提取微博和新闻评论中的热点话题和公众观点, 进而识别公众评价的情感倾向, 进行褒贬情感的辨识。

孙长智[2] (2024)提出了一种深度主题自编码器模型(DTAM), 结合自编码器、生成对抗网络和 LDA 模型, 通过特征提取、生成和主题发现功能, 实现了对电影短评中主题词的精准分类。实验结果表明, DTAM 模型能有效反映评论中的主要观点和情感倾向, 为电影制作和市场营销提供了有价值的参考数据。

涂小琴[3] (2017)通过对豆瓣网站评分高(9.1 分)和评分低(5.2 分)的两部电影进行评论数据的收集, 利用 Python 网络爬虫获取并清理评论数据。接着, 基于改进的 TF-IDF 算法, 结合 PMI (Pointwise Mutual Information)算法对评论进行了分类。通过对评论中的分词进行 PMI 值计算, 最终选取出 PMI 最高的 15 个关键词, 并对这些关键词进行了详细分析与统计。分析结果揭示了评分差异背后的潜在情感倾向及观众的关注点。

张璐[4] (2019)通过运用 Python 技术对中国译介文学进行情感分析, 尤其是针对《三体》在亚马逊平台上的海外读者评论进行分析, 具有创新性。实验证明, 情感分析能够大规模读取并量化呈现海外读者的褒贬态度。研究发现, 整体而言, 海外读者对《三体》译本呈现正向情感态度, 尤其对小说的叙述角度评价较高, 情节和人物方面的评价较次。刘宇昆的译本接受度高于周华的译本。该研究再次验证了, 只有兼具世界性和民族性的作品才能在海外市场取得成功, 赢得读者的喜爱。

国外的情感分析研究不仅集中在电影、社交媒体和商品评论等典型领域, 还拓展到政治、公共卫生以及数据泄露等新兴领域, 显示了情感分析在不同应用场景中的广泛适用性。研究者们通过不同的数据源和分析方法, 探讨了如何从复杂的社交互动和消费行为中提取有意义的情感信息。

Makmun [5]等人(2023)针对《奥本海默》电影评论的情感分类研究中使用了朴素贝叶斯算法, 研究结果表明该算法能有效区分正面与负面情感, 达到了 96%的准确率, 平均精确率为 98%, 平均召回率为 90%, 平均 F1 分数为 93%。

Wongkar [6]等人(2019)则分析了有关 2019 年印度尼西亚总统候选人的推特情感数据, 使用朴素贝叶斯算法得出结果: 佐科维 - 马鲁夫·阿明(Jokowi-Ma'ruf Amin)获得 45.45%的正面情感和 54.55%的负面情感, 而普拉博沃 - 桑迪亚加(Prabowo-Sandiaga)则获得 44.32%的正面情感和 55.68%的负面情感。朴素贝叶斯模型在情感分类中的准确率达到 80.90%, 优于 SVM (63.99%)和 K-NN (73.34%)。

此外, Turmudi Zy [7]等在对推特数据泄露进行情感分析, 使用朴素贝叶斯算法的研究表明, 该模型表现出了卓越的准确性, 准确率为 98.33%, 精确率为 100%, 召回率为 97.13%。

Pristiyono [8]等人(2020)通过抓取推特数据, 以“新冠疫苗”(Vaccine COVID-19)为关键词, 使用朴素贝叶斯算法进行情感分析。研究中手动使用推特 API 获得访问令牌, 通过 Rapid Miner 工具提取所需数据, 并结合 Drone Emprit Academic Streaming Public Twitter 工具克服了手动抓取的限制, 获取了超过 6000 条推文。这些推文集中于 2021 年 1 月 15 日至 22 日的一周时间内。分析结果显示, 其中超过 3400 条为负面推文(占 56%), 约 2400 条为正面推文(占 39%), 其余 301 条(占 1%)为中性推文。

朴素贝叶斯是一种假设特征独立的概率模型, 在情感分析中得到了广泛应用。然而, 朴素贝叶斯在处理简短且结构简单的评论时表现良好, 但在面对更长、更复杂的文本反馈时, 由于其特征独立假设, 性能会受到一定限制。

通过分析这些影评, 不仅能发现最能引起观众共鸣的内容和需要改进的方面, 还能展示当前观众的趋势和偏好。这对电影行业来说至关重要, 有助于制定更具针对性的营销和制作策略。

与早期研究相比, 本研究在使用朴素贝叶斯算法进行《沙丘 2》影评情感分析时, 突出了几个不同点和创新之处。本研究还采用了 TF-IDF 特征选择方法, 并结合针对性的方式, 对与《沙丘 2》相关的最新

影评数据进行了处理。数据来源于 IMDb，并经过了数据预处理、数据划分和模型评估等多个步骤。尽管方法与《奥本海默》影评相似，本研究最终实现了 95% 的准确率，较一般电影影评分析有所提高，但仍低于《奥本海默》影评研究的准确率(低 1%)。尽管如此，该模型在情感分类(正面、负面和中立)方面仍表现良好，说明其在《沙丘 2》影评情感分析中的有效性。

### 3. 研究方法

在本研究方法中，使用的步骤或框架如下图 1：

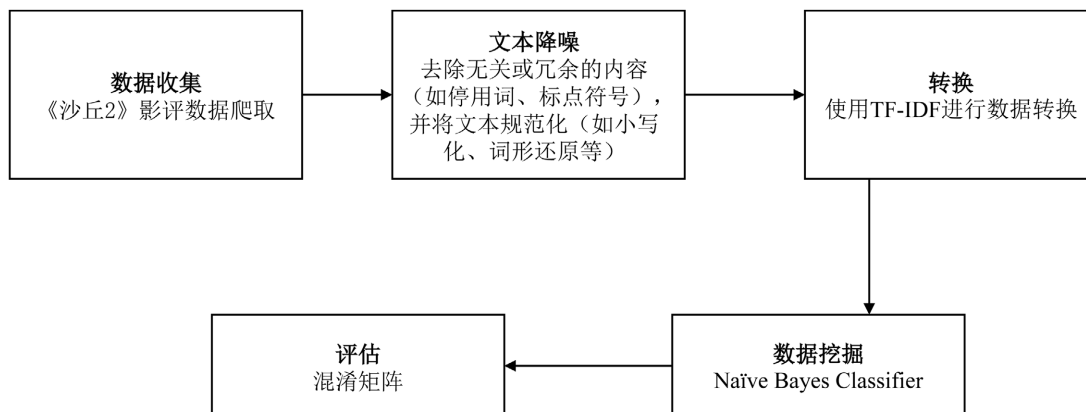


Figure 1. Research framework

图 1. 研究框架

本研究首先利用来自 Kaggle 网站的《沙丘 2》影评的定性数据。这些数据随后经过预处理，以清理文本、简化文本并计算每个单词在文档中的重要性。处理后的数据被用来训练 Naïve Bayes 算法，从而创建一个情感分类模型。通过测试，评估该模型在预测影评是否为正面或负面方面的准确性。评估指标包括准确率、精确度、召回率和 F1 值，用于衡量模型的表现。

**(1) 数据收集。**研究使用的数据集来自 IMDb 电影评论网站，共有 2586 条影评数据，经过筛选后保留了 1065 条数据，数据集来自可信网站 kaggle，见图 2。影评数据来源于多个平台，包括电影评论网站和社交媒体等。

**(2) 文本降噪。**数据清理是预处理的第一步，主要目标是去除文本中不必要的部分，减少干扰因素。

分词是将句子分解为单个单词作为文本分析中最小单位的过程，使得每个单词都可以被单独处理[9] (Darmawan 等, 2023)。通过将文本分割成单独的词元，可以进行各种高级分析，例如计算词频、执行词干提取或词形还原，以及构建语言模型。

在分词之后，下一阶段是过滤出对进一步分析重要的单词，以便只保留相关的单词[10] (Yuyun 等, 2021)。通常通过停用词过滤掉的单词是一些常见词，例如“or”(或者)、“at”(在)、“and”(和)，这些单词在日常语言中经常出现，但在分析上下文中并不具有深刻意义。除了停用词之外，特殊字符、数字以及过于常见或出现频率过低的单词也常常被过滤掉。通过过滤，研究者可以降低数据的维度，提高后续分析过程的计算效率。

词干提取是通过去除不必要的词缀将单词还原为其基本形式的过程[11] (Nurtikasari 等, 2022; Sundara & Ekaputri Arnas, 2020)。这是预处理中重要的一步。词干提取的主要目的是将具有相同词根但形态形式不同的单词合并，例如，“walking”和“walked”将被还原为其基本形式“walk”。因此，词干提取可以提高信息检索和文本分析的有效性。

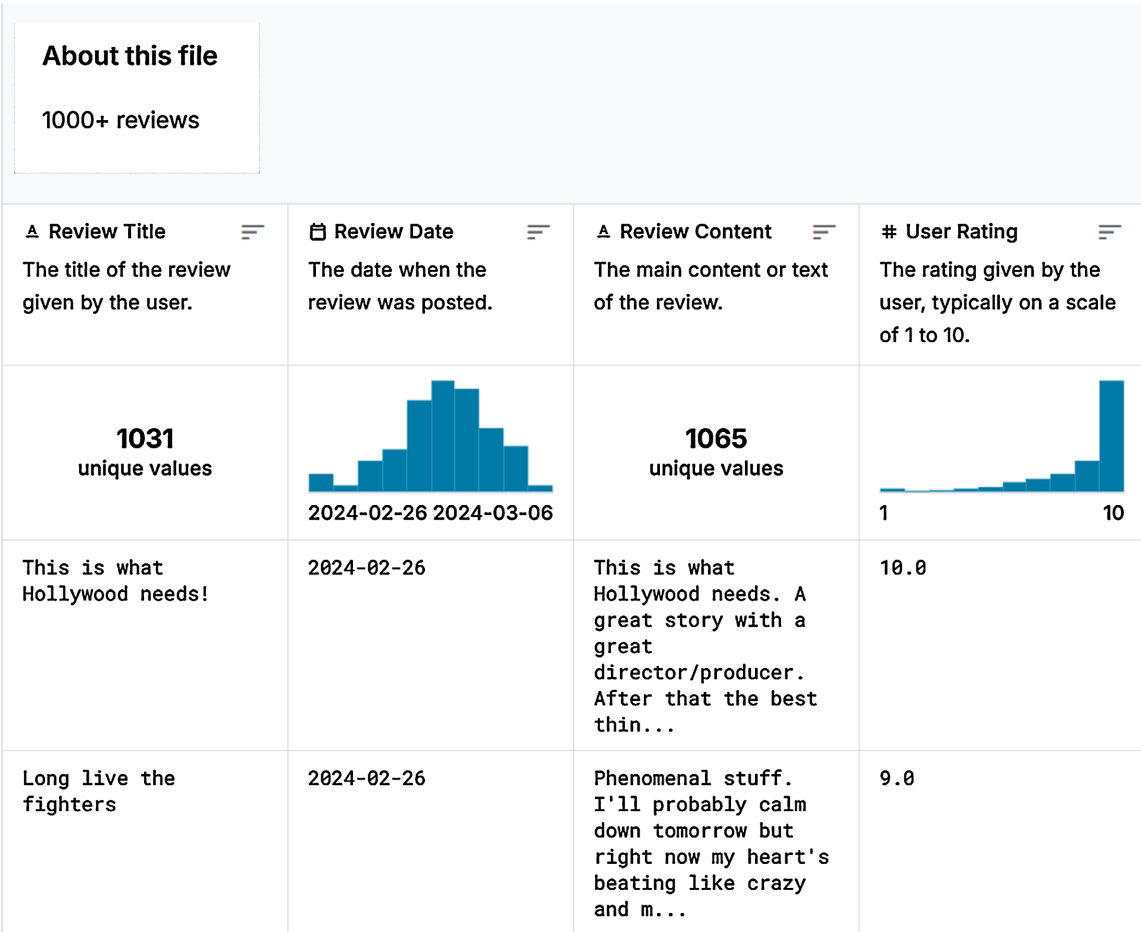


Figure 2. Data source  
图 2. 数据来源

大小写转换是通过将所有字符转换为小写形式来调整单词形式的过程[12] (Esa Tiffani, 2020)。该过程旨在标准化单词形式，从而简化分析过程。通过将所有单词统一为小写字母，可以减少相同单词形式的变体，使文本处理算法能够更高效地工作。例如，“Tom”和“tom”在大小写转换后会被视为相同的单词。

停用词去除是文本预处理中的一个过程，涉及使用停用词表从分词后的文本中移除不太重要的单词。停用词是一些常见单词，虽然它们在文本中频繁出现，但被认为语义价值较低(Kelvin 等，2022)。例如“this” (这个)、“or” (或者)、“are” (是)等。

**(3) 数据转换。**应用 TF-IDF 算法对清洗后的数据进行关键词的提取和频率计算，从而识别出评论中的高频关键词。该算法基于两个主要组成部分：词频(TF)和逆文档频率(IDF)，同时将数值评分转化为二元情感标签(正面或负面)。

**词频(TF)：**词频是指某个词在单个文档中出现的次数。它用来衡量一个词在该文档中的重要性，公式为：

$$TF(t) = \frac{\text{词 } t \text{ 在文档中出现的次数}}{\text{文档中的总词数}}$$

**逆文档频率(IDF)：**逆文档频率是通过取文档总数与含有特定词汇的文档数量之比的对数来计算得到的值：

$$IDF = \log \left( \frac{\text{语料库的文档总数}}{\text{包含该词的文档数} + 1} \right)$$

**TF-IDF:** TF 和 IDF 结合起来，得到每个词的 TF-IDF 值。TF-IDF 的最终公式为：

$$TF\text{-}IDF(t) = TF(t) \times IDF(t)$$

(4) **《沙丘 2》影评情感分析。**得到了文档的 TF-IDF 矩阵，接下来就可以将这些数据传递给机器学习模型 Naive Bayes 进行训练和预测。模型会基于这些数值化的特征，学习文本的模式和规律，进而进行分类(如情感分析、主题分类等)。

(5) **实验结果及分析。**使用测试集数据对模型进行评估。通过预测测试集中的情感标签，并将模型预测的结果与实际标签进行对比，来评估模型的表现。常用的评估指标有：准确率(Accuracy)、精确度(Precision)、召回率(Recall)、F1-Score。

4. 实验结果及分析

研究使用的数据包含了 2586 条影评数据，这些数据原本来自 IMDb 电影评论网站，然后通过筛选或去除冗余数据，将数据量减少到 1065 条，见表 1。虽然数据集的核心来源是 IMDb 电影评论网站，但这些影评数据也可以来源于多个平台，包括电影评论网站 Rotten Tomatoes、Metacritic、社交媒体如 Twitter、Facebook、Instagram 等社交平台。

Table 1. Movie review screening  
表 1. 电影评论筛选

|      | Review Title                                        | Review Date |
|------|-----------------------------------------------------|-------------|
| 0    | This is what Hollywood needs!\n                     | 2024-02-26  |
| 1    | Long live the fighters\n                            | 2024-02-26  |
| 2    | Ladies and gentleman... the PEAK of filmmaking...\n | 2024-02-28  |
| 3    | WOW! I need Dune Messiah now.\n                     | 2024-02-26  |
| 4    | The Sci-Fi/Fantasy Epic of our Generation\n         | 2024-02-26  |
| ...  | ...                                                 | ...         |
| 1061 | A visual masterpiece\n                              | 2024-03-02  |
| 1062 | Ok. But where is the music?!\n                      | 2024-03-04  |
| 1063 | Meh\n                                               | 2024-03-01  |
| 1064 | Dune: Part Two\n                                    | 2024-03-05  |
| 1065 | Matrix Revolutions meets Lawrence Arabia\n          | 2024-03-04  |

接下来对数据进行了标签处理，将评分列转换为包含情感标签的列。评分较高(如 8~10 分)表示积极情感，评分较低(如 1~4 分)表示消极情感，而中间分数(如 5~7 分)表示中性情感。

以下是已经标记的数据结果，见表 2。

预处理是将初始文本或原始文本作为输入，应用一些基本的操作，以改变或去除在后续处理中无用的文本元素的阶段[13] (Najjichah *et al.*, 2019)。

这是使用 Python 编程语言进行数据预处理的结果，见表 3。

使用词云进行情感可视化，可以有效地展示文本数据中最重要的词汇及其出现频率，帮助我们直观地分析情感分析结果和文本内容。根据 TF-IDF 算法对清洗后的数据进行关键词的提取和频率计算，

**Table 2.** Data results  
**表 2.** 数据结果

|      | Review Content                                     | User Rating | Sentiment |
|------|----------------------------------------------------|-------------|-----------|
| 0    | This is what Hollywood needs. A great story wi...  | 10.0        | positive  |
| 1    | Phenomenal stuff. I'll probably calm down tomo...  | 9.0         | positive  |
| 2    | This is the kind of movie that is impossible t...  | 10.0        | positive  |
| 3    | If you liked or loved the first one, the same...   | 10.0        | positive  |
| 4    | Had the pleasure to watch this film in an earl...  | 10.0        | positive  |
| ...  | ...                                                | ...         | ...       |
| 1061 | Dune Part 2 is dope to the most extent. Visuall... | 8.0         | positive  |
| 1062 | First I want to address the majority of "Ones"...  | 7.0         | neutral   |
| 1063 | It wasn't a bad movie... but I would rather ha...  | 6.0         | neutral   |
| 1064 | Dune: Part Two is a cinematic triumph, a colos...  | 7.0         | neutral   |
| 1065 | I know. Poor Stilgar. Without him, we're lost...   | 7.0         | neutral   |

**Table 3.** Results of data preprocessing using Python  
**表 3.** Python 编程语言进行数据预处理的结果

|   | Original Review                                   |
|---|---------------------------------------------------|
| 0 | This is what Hollywood needs. A great story wi... |
| 1 | Phenomenal stuff. I'll probably calm down tomo... |
| 2 | This is the kind of movie that is impossible t... |
| 3 | If you liked or loved the first one, the same...  |
| 4 | Had the pleasure to watch this film in an earl... |
|   | Processed Review                                  |
| 0 | hollywood need great stori great directorprodu... |
| 1 | phenomen stuff ill probabl calm tomorrow right... |
| 2 | kind movi imposs justic talk kind experi never... |
| 3 | like love first one appli one person love one...  |
| 4 | pleasur watch film earli screen complet blown...  |

从而识别出评论中的高频关键词。

此外，为可视化处理后的电影评论数据，本研究还利用 WordCloud 库生成词云图字号越大代表出现的次数越多，如图 3 所示，以直观展示关键词的分布情况。

从这幅《沙丘 2》的词云图可以看出，观众的评论集中于电影的核心元素，如“movie”“film”“story”和“character”等，这些词在图中占据显著位置，反映了观众对电影整体质量、情节发展以及人物塑造的高度关注。同时，“Dune”和“Part”也表明电影的品牌效应和续集特性得到了讨论。诸如“great”“good”“visual”和“scene”这样的积极词汇频繁出现，表明许多评论对电影的视觉效果、场景设计和整体观感持正面态度。此外，主角“Paul”和导演“Denis Villeneuve”也被频繁提及，显示出观众对角色表演和导演功力的认可。总体来看，这幅词云揭示了观众对《沙丘 2》在故事、视觉和人物表现上的高度评价，整体情感倾向积极。



**Figure 3.** Keywords word cloud  
**图 3.** 关键词词云

TF-IDF (词频 - 逆文档频率)是一种在自然语言处理中使用的词权重计算方法,旨在评估某个词在特定文本中的重要性,并结合整个文档语料库进行考量。TF-IDF 权重常用于信息检索领域。该方法会为在特定文档中频繁出现的词赋予更高的权重,但如果该词在多个文档中同时出现,则会降低其权重。

数据划分是指将数据集中的数据分为两个不同的子集，即训练数据和测试数据，用于训练和测试机器学习模型。其目的是评估模型在未见过的数据上的性能表现有多强[14] (Putra 等, 2024)。机器学习模型通过使用数据集的一部分进行训练，以学习数据中的模式和关联，从而能够对新的、未见过的数据进行预测。在训练过程中，模型依赖标注数据来发现规律，而更多高质量的训练数据可以显著提升模型的性能。同时，测试数据是一个独立的数据子集，用于评估模型的表现。在训练阶段，测试数据不会被使用。测试时，模型会对测试数据生成预测标签，这些预测结果与真实标签进行比较，从而计算评估指标，包括 F1-score、准确率、精确率和召回率。

通过准确率、分类报告和混淆矩阵等指标对预测结果进行评估。这些评估有助于理解情感分类过程的有效性以及模型的性能。分类结果如下图 4 所示：

Accuracy: 0.9460966542750929

## Classification Report:

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| negative     | 0.99      | 0.94   | 0.97     | 174     |
| neutral      | 0.87      | 0.99   | 0.93     | 177     |
| positive     | 0.99      | 0.90   | 0.94     | 187     |
| accuracy     |           |        | 0.95     | 538     |
| macro avg    | 0.95      | 0.95   | 0.95     | 538     |
| weighted avg | 0.95      | 0.95   | 0.95     | 538     |

**Confusion Matrix:**

```
[[164    9    1]
 [  0 176    1]
 [  1  17 169]]
```

**Figure 4.** Evaluation classification for prediction results  
**图 4.** 预测结果评估分类

根据分类测试结果，训练后的模型能够以 95% 的准确率、精确率和召回率对测试数据进行分类。如

此高的准确率表明, Naïve Bayes 模型能够有效识别《沙丘: 第二部》影评中的正面、负面和中性情感。上述结果的可视化如下图 5 所示:

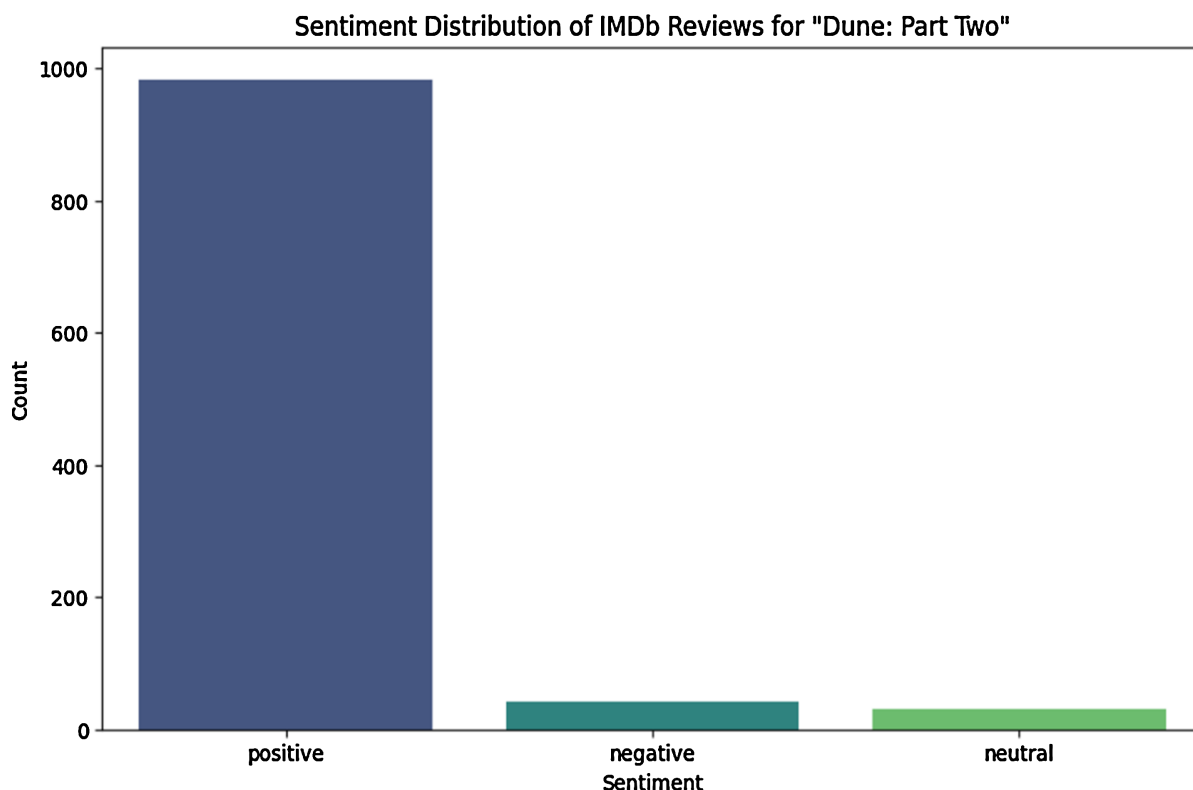


Figure 5. Sentiment distribution of reviews for *Dune: Part Two*

图 5. 《沙丘: 第二部》影评情感分布图

## 5. 讨论

本研究通过 Naïve Bayes 算法对《沙丘 2》影评进行情感分析, 结果表明, 该方法在情感分类任务中表现出色。然而, 朴素贝叶斯假设特征独立, 这可能对处理复杂文本造成一定限制。

此外, 数据来源单一(仅限 IMDb)可能对其他影评平台的适用性产生影响。通过分析影评中的关键词和情感倾向, 发现观众对视觉效果和叙事深度普遍给予正面评价, 而对情节节奏和部分角色塑造存在一定争议。这为电影制作方优化续集提供了有价值的反馈。与类似研究相比(如《奥本海默》影评情感分析), 本研究采用的 TF-IDF 特征提取方法具有更高的灵活性, 但分类精度略有差距。

未来研究建议尝试更复杂的模型(如 BERT 或 LSTM), 以更好地捕捉文本中的深层语义信息。此外, 结合多平台、多语言的数据集, 探讨不同文化背景对电影评价的影响, 为跨文化电影研究提供更多视角。

## 6. 总结

本研究利用朴素贝叶斯算法对《沙丘 2》的影评进行了情感分析。通过预处理和 TF-IDF 特征选择, 模型实现了 95% 的高准确率, 能够有效区分正面、负面和中性情感。研究结果表明, 朴素贝叶斯作为一种简单高效的文本情感分类方法, 结合合适的特征提取技术, 在情感分析任务中表现优异。

分析结果显示, 观众的评论主要集中在电影的视觉效果、叙事深度和人物塑造上, 整体反馈以正面情感为主。研究揭示了这些核心元素对观众满意度的显著影响, 为电影制作方在优化电影内容和制定营

销策略方面提供了宝贵的见解。

尽管模型表现良好,但其依赖特征独立假设,且数据来源单一(IMDb 影评)的局限性,可能对结果的通用性带来一定影响。未来研究可以尝试采用更先进的模型(如 BERT 或 LSTM),结合更大规模且多样化的数据集,尤其是跨文化的评论数据,进一步探讨不同文化背景对电影评价的影响。本研究为电影影评情感分析提供了重要的参考,为理解观众偏好和提升电影行业竞争力奠定了基础。

## 参考文献

- [1] 张紫琼, 叶强, 李一军. 互联网商品评论情感分析研究综述[J]. 管理科学学报, 2010, 13(6): 84-96.
- [2] 孙长智, 易铭瑒, 蒋京京, 等. 基于深度主题自编码器模型的电影短评数据情感分析研究[J]. 情报科学, 2024, 42(11): 69-75.
- [3] 涂小琴. 基于 Python 爬虫的电影评论情感倾向性分析[J]. 现代计算机(专业版), 2017(35): 52-55.
- [4] 张璐. 从 Python 情感分析看海外读者对中国译介文学的接受和评价: 以《三体》英译本为例[J]. 外语研究, 2019, 36(4): 80-86.
- [5] Noviansyah, B., Effendi, M.M. and Achmad, Y. (2024) Sentiment Analysis of Oppenheimer Movie Reviews: Naïve Bayes Algorithm for Public Opinion. *Journal of Computer Networks, Architecture and High Performance Computing*, **6**, 1406-1414. <https://doi.org/10.47709/cnahpc.v6i3.4393>
- [6] Wongkar, M. and Angdresey, A. (2019) Sentiment Analysis Using Naive Bayes Algorithm of the Data Crawler: Twitter. 2019 *Fourth International Conference on Informatics and Computing (ICIC)*, Semarang, 16-17 October 2019, 1-5. <https://doi.org/10.1109/icic47613.2019.8985884>
- [7] Turmudi Zy, A., Nugroho, A., Rivaldi, A. and Afriantoro, I. (2022) Analisis Sentimen Terhadap Pembobolan Data Pada Twitter Dengan Algoritma Naive Bayes. *Jurnal Teknologi Informatika Dan Komputer*, **8**, 202-213. <https://doi.org/10.37012/jtik.v8i2.1240>
- [8] Pristiyo, Ritonga, M., Ihsan, M.A.A., Anjar, A. and Rambe, F.H. (2021) Sentiment Analysis of COVID-19 Vaccine in Indonesia Using Naïve Bayes Algorithm. *IOP Conference Series: Materials Science and Engineering*, **1088**, Article 012045. <https://doi.org/10.1088/1757-899x/1088/1/012045>
- [9] Darmawan, S., Agusvina, N., Lusa, S. and Sensuse, D.I. (2023) Knowledge Management Factors and Its Impact on Organizational Performance: A Systematic Literature Review. *JOIV: International Journal on Informatics Visualization*, **7**, 161-167. <https://doi.org/10.30630/joiv.7.1.1644>
- [10] Yuyun, Hidayah, N. and Sahibu, S. (2021) Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, **5**, 820-826. <https://doi.org/10.29207/resti.v5i4.3146>
- [11] Nurtikasari, Y., Alam, S. and Hermanto, T.I. (2022) Analisis Sentimen Opini Masyarakat Terhadap Film Pada Platform Twitter Menggunakan Algoritma Naive Bayes. *INSOLOGI: Jurnal Sains Dan Teknologi*, **1**, 411-423. <https://doi.org/10.55123/insologi.v1i4.770>
- [12] Esa Tiffani, I. (2020) Optimization of Naive Bayes Classifier by Implemented Unigram, Bigram, Trigram for Sentiment Analysis of Hotel Review. *Journal of Soft Computing Exploration*, **1**, 1-7.
- [13] Najjichah, H., Syukur, A. and Subagyo, H. (2019) Pengaruh Text Preprocessing Dan Kombinasinya Pada Peringkat Dokumen Otomatis Teks Berbahasa Indonesia. *Jurnal Teknologi Informasi*, **15**, 1-11.
- [14] Putra, F., Tahiyat, H.F., Ihsan, R.M., Rahmaddeni, R. and Efrizoni, L. (2024) Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, **4**, 273-281. <https://doi.org/10.57152/malcom.v4i1.1085>