

**ST. TERESA'S COLLEGE (AUTONOMOUS),
ERNAKULAM
AFFILIATED TO MAHATMA GANDHI UNIVERSITY**



**SENTIMENTAL ANALYSIS OF MOVIE REVIEWS
PROJECT REPORT**

In partial fulfillment of the requirements for the award of the degree of

**BACHELOR OF SCIENCE IN
COMPUTER APPLICATIONS [TRIPLE MAIN]**

Submitted By

NEHA SURESH
III BSc Computer Applications [Triple Main]
Register No: SB21CA020

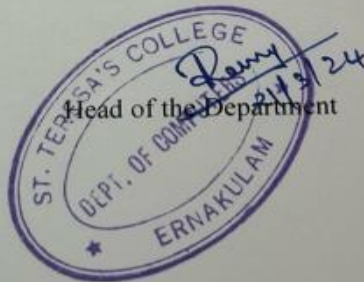
Under the Guidance of Asst.
Prof. Raji S Pillai

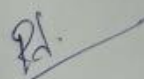
**DEPARTMENT OF COMPUTER APPLICATIONS
2021-2024**

CERTIFICATE




This is to certify that the project report entitled "Sentimental Analysis of MovieReviews" is a bona-fide record of the work done by NEHA SURESH (SB21CA020) during the year 2021 – 2024 and submitted in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Applications (Triple Main) under Mahatma Gandhi University, Kottayam.




Internal Examiner

Date: 21/03/24



External Examiner

DECLARATION

I, NEHA SURESH (Register no: SB21CA020), B.Sc. Computer Applications [Triple Main] final year student of St. Teresa's College (Autonomous), Ernakulam, hereby declare that the project submitted named "Sentimental Analysis of Movie Review" for the Bachelor's Degree in Computer Applications [Triple Main] is my original work. I further declare that the said work has not previously been submitted to any other university or academic body.

Date: 21/03/24

Place: Ernakulam



NEHA SURESH

ACKNOWLEDGMENT

First and foremost, I thank God Almighty for his blessings. I take this opportunity to express my gratitude to all those who helped me in completing this project successfully. I wish to express my sincere gratitude to the Provincial Superior and Manager Rev. Dr. Sr. Vinitha CSST, Director Rev. Sr. Emeline CSST and Principal Dr. Alphonsa Vijaya Joseph for providing all the facilities.

I express my sincere gratitude towards the Head of the department Ms. Remya CJ for the support. I deeply express my sincere thanks to my guide Asst.Prof. Raji S Pillai for her proper guidance and support throughout the project work. I thank all the other faculty of the department for their exemplary guidance and appreciation throughout the project.

My deepest appreciation goes to all my friends and relatives for their support. I convey my hearty thanks to my parents and my brother who encouraged me and provided all the opportunities for the completion of this project.

NEHA SURESH

ABSTRACT

Sentiment analysis, a component of opinion mining, is a novel concept in data mining that focuses on extracting emotions from online communication data, such as feedback in comments and reviews posted on the internet. The goal is to discern specific views or judgments on a particular topic. This process involves classifying text data into positive, negative, or neutral sentiments.

This paper presents a classification task of sentiment analysis using a movie database. The analysis involves categorizing movie reviews based on their sentiment polarity. The dataset likely contains a variety of opinions ranging from highly positive to highly negative, with some reviews being neutral. By applying sentiment analysis techniques, the system aims to automatically categorize these reviews into their respective sentiment polarities.

The classification task involves several steps, including data preprocessing, feature extraction, and model training. Various machine learning and deep learning algorithms can be employed for this task, such as Support Vector Machines (SVM), Naive Bayes, Logistic Regression, and Neural Networks. Additionally, lexicon-based approaches, such as the VADER sentiment analysis tool, can be used to analyze the sentiment of the text based on a predefined set of words and their associated sentiment scores.

Overall, this paper contributes to the field of sentiment analysis by applying it to a specific domain, the movie database, and demonstrating the effectiveness of various approaches in classifying sentiment in text data.

CONTENTS

1. Introduction

1.1. About Project

1.2. About Organization

1.3. Objective of Project

2. Methodology

2.1. Input Data

2.2. Pre processing

2.3. Method Used

2.3.1. VADER

3. Source Code

4. Results

5. Conclusion

6. Literature Review

7. References

1. INTRODUCTION

1. ABOUT PROJECT

Sentiment analysis is a rapidly growing research area in text mining, particularly with the increasing popularity of online reviews on various websites. It involves computationally studying people's judgments, attitudes, and emotions towards entities like individuals, events, or topics. Opinion mining, a related field, extracts and analyzes opinions about an entity, while sentiment analysis specifically focuses on identifying sentiment words in text document and analyzing them.

In the context of movie reviews, sentiment analysis plays a crucial role, as opinions expressed in these reviews can impact audiences, film directors, and production companies. Movie reviews found online are often informal and unstructured, making them a suitable subject for sentiment analysis.

This paper collects reviews from websites and utilizes a lexical approach to determine the overall polarity of movie reviews, categorizing them into positive, negative, or neutral sentiments. The results are then graphically displayed to show the distribution of positive and negative polarities.

1.1 ABOUT ORGANIZATION

Founded in Cochin back in the year 1992, LCC started its journey as a premier Computer Education Centre and later advanced into one of the most reputed brands for computer learning throughout India. After conducting extensive market research and analysis, LCC's expert academicians and certified faculties collaborate with our leading computer technology company allies, such as, to develop pertinent certification courses.

1.2 OBJECTIVE OF PROJECT

The objectives of movie review analysis using sentiment analysis tools are to determine the overall sentiment (positive, negative, or neutral) of reviews, understand audience sentiments towards specific movies, and compare sentiments across different movies or genres to identify audience preferences. This analysis helps in gauging audience satisfaction, identifying trends, and making informed decisions in the film industry.

The main objective of the organization is to continuously optimize their customers' business through their world-class solutions, services and products. They ensure the success of the company by constantly and consistently satisfying the customers, shareholders, and employees.

2. METHODOLOGY

This section gives the description of the steps followed for the movie datasets mining for sentiment analysis. The movie review datasets is used and label to the polarity is provided as follows: positive, negative, and neutral.

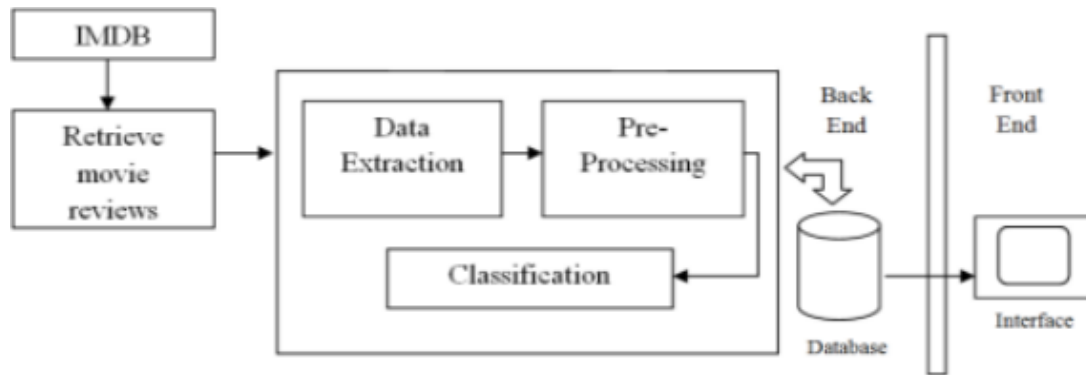


Fig 1. Architecture diagram for sentiment analysis

Movie reviews are used to determine the sentiment of each review.

2.1. INPUT DATA

The input data for sentiment analysis of movie reviews is sourced from movie review datasets. These datasets are used to select reviews for a specific movie, which are then displayed on a webpage. Reviews for newly released movies are continuously added to the datasets to ensure the analysis remains current.

2.2. PROCESSING

Text pre-processing involves two primary techniques:

Tokenization: This step involves breaking down the text into individual tokens or words. These tokens are then used for further analysis and processing.

Removal of Stop Words: Common stop words in English, such as "an," "a," "of," "the," "you," and "and" are removed from the text. These words do not carry significant meaning and their removal helps in improving the accuracy of sentiment analysis.

These pre-processing techniques are essential for preparing the text data for sentiment analysis. They help in reducing noise and irrelevant information, making the analysis more effective and accurate.

2.3. METHOD USED

After reviewing various papers, the VADER method emerged as a suitable approach for sentiment analysis of movie reviews. This method allows for the classification of sentiments into positive, negative, and neutral categories based on the content of the reviews. VADER, which stands for Valence Aware Dictionary and sentiment Reasoner, is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media texts. Its ability to handle nuances in informal language makes it particularly useful for analyzing the sentiments conveyed in movie reviews.

2.3.1 VADER

VADER (Valence Aware Dictionary and sentiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically designed to analyze sentiments in text, particularly in social media. It is widely used due to its ability to handle nuances in informal language, making it suitable for analyzing sentiments in movie reviews.

VADER works by assigning a polarity (positive, negative, or neutral) to each word in a text based on a pre-defined sentiment lexicon. The lexicon contains words scored based on their sentiment intensity, with scores ranging from -4 (extremely negative) to +4 (extremely positive). VADER also considers the context of words, such as capitalization and punctuation, to improve its accuracy.

In the context of sentiment analysis of movie reviews, VADER can be used to analyze the overall sentiment of a review by aggregating the sentiment scores of individual words. By summing up the scores and applying a threshold, the sentiment of the entire review can be classified as positive, negative, or neutral.

Additionally, VADER can be used to compare the sentiments of different movies or genres. By analyzing a large number of reviews, trends can be identified to determine which movies are generally better received by the audience.

Overall, VADER is a powerful tool for sentiment analysis of movie reviews due to its ability to handle informal language and its ease of use. It provides a quick and effective way to analyze the sentiments expressed in text, making it valuable for understanding audience reactions to movies.

3.SOURCE CODE

```
110
117 import nltk
118 import matplotlib.pyplot as plt
119 from nltk.sentiment.vader import SentimentIntensityAnalyzer
120
121 nltk.download('vader_lexicon')
122
123 def analyze_sentiment(review):
124     sia = SentimentIntensityAnalyzer()
125     sentiment_scores = sia.polarity_scores(review)
126
127     if sentiment_scores['compound'] >= 0.05:
128         return 'Positive'
129     elif sentiment_scores['compound'] <= -0.05:
130         return 'Negative'
131     else:
132         return 'Neutral'
133
134 import base64
135 from io import BytesIO
136 from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
137 import matplotlib
```

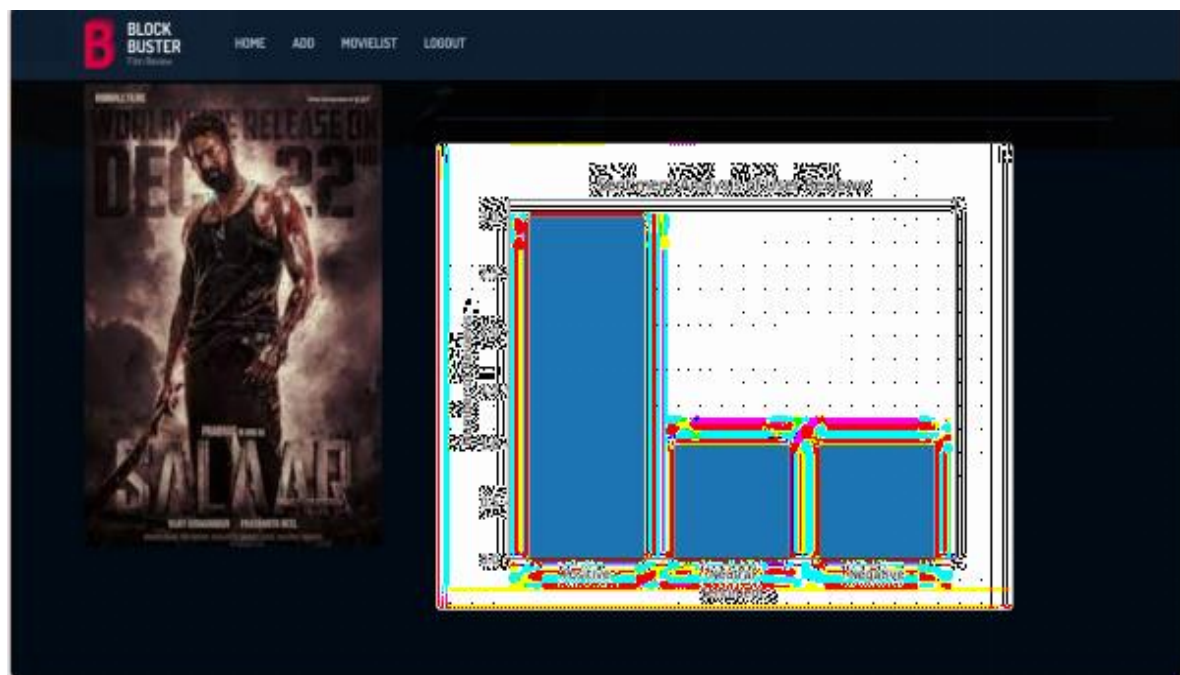
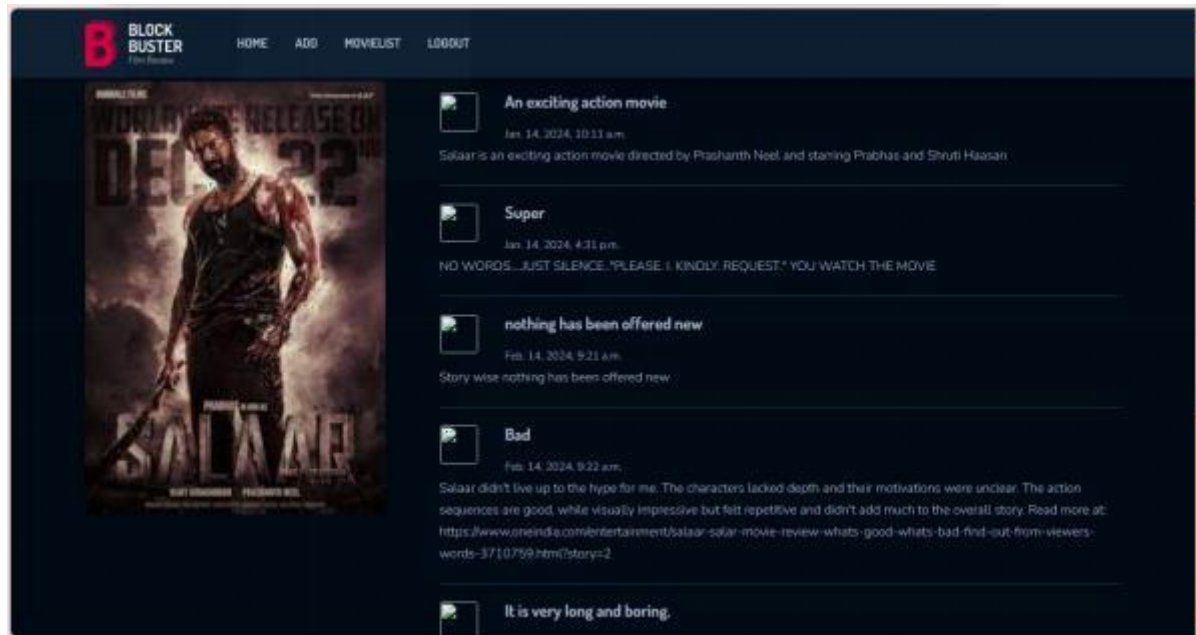
```
137 matplotlib.use('Agg')
138
139 def generate_and_get_plot(movie_id):
140     reviews = add_review.objects.filter(movie_id=movie_id)
141     sentiments = [analyze_sentiment(review.review) for review in reviews]
142
143     sentiment_counts = {
144         'Positive': sentiments.count('Positive'),
145         'Neutral': sentiments.count('Neutral'),
146         'Negative': sentiments.count('Negative')
147     }
148
149     labels = sentiment_counts.keys()
150     values = sentiment_counts.values()
151
152     fig, ax = plt.subplots()
153     ax.bar(labels, values)
154     ax.set_xlabel('Sentiment')
155     ax.set_ylabel('Number of Reviews')
156     ax.set_title('Sentiment Analysis of User Reviews')
```

```
157
158     plot_data = BytesIO()
159     fig.savefig(plot_data, format='png')
160     plt.close(fig)
161
162     plot_data.seek(0)
163     plot_base64 = base64.b64encode(plot_data.read()).decode()
164
165     return plot_base64
166
167
168
169 v def detail(request):
170     id = request.GET.get('id')
171     n = Movie.objects.get(id=id)
172     reviews=add_review.objects.filter(movie=n)
173 v     for review in reviews:
174         review.sentiment = analyze_sentiment(review.review)
175
176     plot_base64 = generate_and_get_plot(id)
```

```
177
178     return render(request, 'User/movieSingle.html', {"n":n, "reviews":reviews, 'plot_base64':plot_base64})
179
180 v def addreview(request):
181     uid=request.session['uid']
182     mid=request.GET.get("mid")
183     user=Registration.objects.get(user=uid)
184     movie=Movie.objects.get(id=mid)
185 v     if request.POST:
186         title=request.POST['title']
187         review=request.POST['review']
188         re=add_review.objects.create(title=title, review=review, user=user, movie=movie)
189         re.save()
190         return redirect("/view_movie")
191     return render(request, 'User/add_review.html')
192
193 # def analyze_and_plot_sentiment_user(request):
194 #     id = request.GET.get('id')
195 #     reviews = Review.objects.filter(movie_id=id)
196
```

```
196
197 #     for review in reviews:
198 #         review.sentiment = analyze_sentiment(review.review)
199
200 #     plot_base64 = generate_and_get_plot(id)
201
202
203
204 #     context = {
205 #         'plot_base64': plot_base64,
206 #
207 #     }
208 #     return redirect(request, 'User/sentiment_analysis_user.html', context)
```

4. RESULTS



5. CONCLUSION

In conclusion, this paper has explored the use of the VADER (Valence Aware Dictionary and sentiment Reasoner) method for sentiment analysis of movie reviews. Through a comprehensive review of literature, it was established that VADER is a well- suited approach for analyzing sentiments in movie reviews due to its ability to handle nuances in informal language, which is prevalent in such reviews.

The VADER method utilizes a sentiment lexicon to assign polarity (positive, negative, or neutral) scores to words in a text. By aggregating these scores, the overall sentiment of a movie review can be determined. This approach allows for the efficient analysis of large volumes of text, making it ideal for processing the numerous reviews that are typically available for popular movies.

Furthermore, VADER can be used to compare sentiments across different movies or genres, providing valuable insights into audience preferences and trends. Its ease of use and effectiveness make it a valuable tool for researchers and industry professionals interested in understanding audience reactions to movies.

Overall, this paper has demonstrated the effectiveness of the VADER method for sentiment analysis of movie reviews. Future research could explore the application of VADER in other domains or the development of more sophisticated sentiment analysis tools to further enhance our understanding of audience sentiments.

6. LITERATURE REVIEW

SL NO.	DETAILS ABOUT THE RESEARCH PAPER	METHODOLOY USED	MERITS	DEMERITS
1.	<p>Paper name: Twitter Sentiment Analysis of COVID-19 In India: VADER Perspective</p> <p>Authors: K Bhadra, A Dash, S Darshana, M Pandey, SS Rautaray, RK Barik</p> <p>Published year: 2023</p> <p>Journal:2023 International Conference on Communication, Circuits, and Systems (IC3S)</p>	<p>COV19Tweets dataset from IEEE</p> <p>VADER sentiment analysis tool</p>	<p>VADER sentiment analysis tool performs better than TextBlob in categorizing COVID-19-related tweets.</p> <p>The research provides a user-friendly web application to visualize and compare sentiment scores of COVID-19 events in India.</p>	<p>No information provided about the demerits of the study.</p>
2.	<p>Paper name: Sentiment Analysis on IMDB Movie Reviews using Machine Learning and Deep Learning Algorithms</p> <p>Authors: <u>K Amulya</u>, <u>S. Swathi</u>, <u>Pille Kamakshi</u>, <u>Y. Bhavani</u></p> <p>Published Year: 2022</p> <p>Journal:2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)</p>	<p>Machine Learning (ML) algorithms</p> <p>Deep Learning (DL) algorithms</p>	<p>Sentiment analysis helps improve business based on customer reviews. Deep learning algorithms provide accurate and efficient results for sentiment analysis.</p>	<p>The paper does not mention any specific demerits of using machine learning and deep learning algorithms for sentiment analysis.</p>
3.	<p>Paper name: Sentiment analysis on IMDB using lexicon and neural networks</p> <p>Authors: Zeeshan</p>	<p>Supervised method for sentiment classification</p> <p>Lexicon-based approach using</p>	<p>Sentiment analysis can provide valuable insights for decision-making processes.</p> <p>Neural networks</p>	<p>Complexities in human language can lead to misinterpretation of sentiment.</p> <p>Lexicon-based</p>

	<p>Shaukat¹, Abdul Ahad Zulfiqar¹, Abdul Ahad Zulfiqar², Chuangbai Xiao¹, Muhammad Azeem¹, Muhammad Azeem³, Tariq Mahmood¹</p> <p>Year:2020</p> <p>Journal:SN Applied Sciences, 2020 - Springer</p>	WordNet dictionaries	trained on movie reviews achieved a final accuracy of 91%.	approach may not capture all nuances of sentiment.
4.	<p>Paper name: Analysis of sentiment based movie reviews using machine learning</p> <p>Authors:<u>Sachin Chirgaiya</u>¹, <u>Deepak Sukheja</u>¹, <u>Niranjan Shrivastava</u>¹, <u>Romil Rawat</u></p> <p>Year:2021</p> <p>Journal: Journal of Intelligent & Fuzzy Systems, 2021</p>	Classifier model trained using feature extraction and ranking Natural language processing (NLP) for sentiment analysis	Classifier model trained with feature extraction and ranking achieves 97.68% accuracy. NLP used to implement proposed model for sentiment classification of movie reviews.	Lack of grammatical structures in movie comments NLP used to implement proposed model and compare with existing models
5.	<p>Paper name: An Approach of Sentiment Analysis for Movie Reviews</p> <p>Authors: M.B,C.S</p> <p>Year: 2022</p> <p>Journal: 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT)</p>	Bag-of-words model Defining adjectives and adverbs, managing negations, bounding word frequencies, and employing Word Net synonyms knowledge	Unique machine learning strategy for identifying phrases in movie reviews. Various text aspects extracted for determining polarity of movie reviews.	
6.	<p>Paper name: Sentiment Analysis On Movie Reviews Using Recurrent Neural Network</p> <p>Authors: Sumesh Kumar Nair, Ravindra Soni</p>	Sentiment analysis using Recurrent Neural Network (RNN) Text	LSTM Recurrent Neural Networks provide high accuracy and polarity in sentiment analysis. Deep learning	Sentiment analysis faces criticism due to its unrealistic nature. Document level sentiment analysis may not be suitable

	Year: 2018	classification using LSTM (Long short-term memory) model	methods, such as LSTM, can learn directly from low-level features.	for evaluating multiple entities.
7.	Paper name: Sentiment Analysis of Movie Reviews using POS tags and Term Frequencies Authors: Oaindrila Das, Chandra Rakesh, Balabantaray Year: 2014 Journal: International journal of computer applications, 2014	NLP and machine learning techniques SVM lite with a single kernel function	Sentiment analysis helps in judging and predicting people's views. Sentiment analysis can be used in various applications such as marketing and recommendation systems.	-
8.	Paper name: A Literature Review on Sentiment Analysis Authors: Venkata Sai, Abhishikth Tholana Year: 2017	Microsoft Azure analyser tool Sentiment polarity categorization	Sentiment analysis can provide benefits and quality information from social media data. Sentiment analysis can be used for product reviews and opinions.	The paper mentions the curse of dimensionality in feature vector formation. The paper discusses the limitations and challenges faced in sentiment polarity categorization.
9.	Paper name: Sentiment Analysis for IMDb Movie Review Using Support Vector Machine (SVM) Method Authors: Fidya Farasalsabila, Verra Lestari, D Diffraan, Nur Cahyo, Tutik Lestari, Rusdi Fahmi, Al Islami, M Akbar Maulana Year: 2023 Journal: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi, 2023	Support Vector Machines (SVM) classifiers Bags of Word (BoW) and Term Frequency–inverse Document Frequency (TF-IDF)	SVM achieved high accuracy for sentiment analysis on IMDb movie reviews. The use of TF-IDF resulted in the highest accuracy test.	

7. REFERENCES

Bhadra, Kunal, et al. "Twitter Sentiment Analysis of COVID-19 In India: VADER Perspective." 2023 International Conference on Communication, Circuits, and Systems (IC3S). IEEE, 2023.

Amulya, K., et al. "Sentiment analysis on IMDB movie reviews using machine learning and deep learning algorithms." 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE, 2022.

Ali, Nehal Mohamed, Marwa Mostafa Abd El Hamid, and Aliaa Youssif. "Sentiment analysis for movies reviews dataset using deep learning models." International Journal of Data Mining & Knowledge Management Process (IJDMP) Vol 9 (2019).

Shaukat, Zeeshan, et al. "Sentiment analysis on IMDB using lexicon and neural networks." SN Applied Sciences 2 (2020): 1-10.

Chirgaiya, Sachin, et al. "Analysis of sentiment based movie reviews using machine learning techniques." Journal of Intelligent & Fuzzy Systems 41.5 (2021): 5449-5456.

Manikandan, B., and S. Chakaravarthi. "An Approach of Sentiment Analysis for Movie Reviews." 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT). IEEE, 2022.

NAIR, SUMESH KUAMR, and RAVINDRA SONI. "Sentiment Analysis on Movie Reviews using Recurrent Neural Network." (2018).

Das, Oaindrila, and Rakesh Chandra Balabantaray. "Sentiment analysis of movie reviews using POS tags and term frequencies." International journal of computer applications 96.25 (2014): 36-41.

Cahyo, D. Diffran Nur, et al. "Sentiment Analysis for IMDb Movie Review Using Support Vector Machine (SVM) Method." Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi 8.2 (2023): 90-95.

