**ML-Based Sentiment Detection in IMDB Movie Reviews**

Male vamshi (15834164, malev@coventry.ac.uk)

7120CEM - Natural Language Processing

Module Leader Name: Dr. Omid Chatrabgoun

**Abstract**

The research examines sentiment assessment of IMDB movie evaluations through machine learning approaches. The available dataset contains 50,000 reviews labeled according to positive or negative sentiments. A processing pipeline performs cleaning operations then separates tokens while eradicating normal stopwords before Bag-of-Words and TF-IDF extract features from text data. The performance evaluation of machine learning models showed Logistic Regression produced 88.7% accuracy as its best outcome while Naive Bayes, SVM and Random Forest followed. Word2Vec serves as part of the project methodology to generate word embeddings for semantic word associations while also creating visualizations of the model sentiment patterns.

**1.Introduction**

The automatic sentiment prediction technology from textual documents requires system development because it enables applications of opinion mining and customer feedback analysis alongside others.

The preprocessing methodology begins by eliminating unwanted textual elements starting from HTML tags and special characters along with common words known as stopwords. The text normalization produces tokenized parts from both stemming and lemmatization that preserve word original roots. Machine learning processes the prepared text after applying the To Boolean function where two feature extraction types TF-IDF and Bag-of-Words (BoW) are used.

Four standard machine learning methods consisting of Logistic Regression alongside Naive Bayes along with Support Vector Machine (SVM) and Random Forest perform training operations on the sanitized dataset. The evaluation framework utilizes two metrics: accuracy metrics together with classification reports and confusion matrices. The mentioned model provides the best performance by achieving an 88.7% accuracy measure.

The project shows efficient integration of traditional machine learning approaches with embedding procedures which demonstrates different inputs and extraction methods and selection methods influence prediction accuracy.

**2. Related Works**

The field of sentiment analysis for movie reviews demonstrated quick progress during recent years because of its combined usage of machine learning and natural language processing (NLP) together with deep learning techniques. The research by Hassan et al. (2024) demonstrated a method for sentiment polarity analysis of Roman Urdu movies reviews with limited support. The authors boosted sentiment prediction accuracy levels through implementing ensemble methods which utilized multiple classifiers. Ensemble techniques play a vital role to boost sentiment classifier performance when working with languages that have limited available resources according to the findings of this research. Sarhan et al. (2024) showed how sentiment analysis should be applied to movie recommendation systems by developing a system based on machine learning algorithms for recommending movies based on user review emotions. The user preference matching process of movies receives enhancement through sentiment analysis integration in the recommendation system developed by Deshmukh et al. (2024). The recommendations become more personalized through their sentiment analysis integration because end-users can identify movies which meet their preferences based on review sentiment analysis. According to Bhushan et al. (2024) their mood-based recommendation system applied sentiment analytics to determine matching movies that matched user emotional states.

**3. Methods**

Traditional sentiment analysis systems used the well-known machine learning models Naive Bayes with Logistic Regression and Support Vector Machines (SVM) for their proven success in text classification purposes. Standard text analysis these systems used elementary text processing methods that segmented information into basic units referred to as terms through tokenization along with stop word cleaning that removed redundant words including "and" and "the".Term Frequency-Inverse Document Frequency (TF-IDF) and Bag-of-Words (BoW) extracted numerical features from the text. Each document gets transformed into a word count vector under the BoW model while TF-IDF calculates term weights by measuring their uniqueness within the document against the whole dataset.

A graph with blue bars

AI-generated content may be incorrect.

Figure 1: Distribution of Sentiments in IMDB Reviews

The new system builds on standard text methods through advanced processing features which create better sentiment analysis in text documents. The system initiates text cleaning as its first step to guarantee data consistency and uniformity. The elimination procedure sanitizes the dataset by removing HTML marks and special characters together with numbers and redundant spaces since these elements create noise that obstructs text pattern revelation. After cleaning the text gets divided into separate words or tokens through tokenization processing. At the subsequent step the system eliminates stopwords since they do not supply significant sentiment information to allow the detection of meaningful terms. Through lemmatization and stemming techniques the system completes normalization of the text. Through these text preprocessing steps the content remains in its essential form which boosts the accuracy levels of sentiment evaluation.

For feature extraction, the system employs a combination of Bag-of-Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) techniques. Weighted terms in TF-IDF analysis become more significant because this method calculates word importance through comparison of document-based frequencies to overall corpus frequencies. Nevertheless BoW preserves raw word frequency patterns across all text. The current methodology reaches its maximum capability when extracting contextual word meanings since it does not work at full capacity.

The system educates multiple machine learning models that consist of Logistic Regression and Naive Bayes along with SVM and Random Forest which receive training input from extracted text features. The models get assessed by evaluating accuracy together with classification reports and confusion matrices to find out what model works best for precise sentiment identification. The new technique surpasses conventional systems because it provides better accuracy alongside reliability which leads to improved sentiment classification for IMDB movie reviews. The system delivers enhanced value to practical applications since it enables better comprehension of sentiment nuances which generates actionable insights.

**4. Experiments**

#### **a)Model Accuracy Comparison**

A graph of blue rectangular shapes

AI-generated content may be incorrect.

Figure 2: Model Accuracy Comparison

The above figure shows the accuracy of each model evaluated in the sentiment analysis system. The models were trained using TF-IDF features derived from the IMDB movie reviews dataset:

#### **b)Confusion Matrices**

The confusion matrices for each model provide further insight into how well the models classified the sentiment, showing the true positives, true negatives, false positives, and false negatives:

Figure 3: Confusion Matrix

A blue squares with white numbers

AI-generated content may be incorrect.

#### **c)Overall Performance Discussion**

The experiment results show that Logistic Regression delivered the best outcome with 88.7% accuracy rate whereas SVM achieved 87.9%. The policies implemented by both models proved successful in identifying positive versus negative sentiments while indicating minimal wrong categorizations in their corresponding confusion matrices. The models achieve higher accuracy rates because they efficiently find complex patterns in text data.

**5. Discussion**

Sentiment analysis conducted on IMDB movie reviews provides researchers with vital data about the performance capabilities of different machine learning models for sentiment classification. The Logistic Regression model proved to be the most effective binary classifier by reaching an accuracy rate of 88.7%. Executing sentiment classification in this domain demands the use of Logistic Regression because it manages high-dimensional data effectively with feature extraction methods such as TF-IDF.Additionally the confusion matrices show these data discrepancies. Logistic Regression obtained lower numbers of false positives and false negatives indicating its ability to achieve optimal precision-recall performance compared to competing models. Random Forest experienced more incorrect negative predictions indicating its difficulty to detect negative reviews compared to positive reviews because of its higher false negative count. Model selection should depend on the unique characteristics of testing datasets and practical goal requirements because these features affect operational performance outcomes. Better feature optimization could enhance the classification outcomes by reducing accuracy levels in Naive Bayes and Random Forest compared to other models.

**6. Conclusion**

A sentiment analysis system implemented traditional machine learning algorithms including Logistic Regression, Naive Bayes, SVM and Random Forest for classifying IMDB movie reviews. Logistic Regression delivered the best accuracy rate of 88.7 percent which proved its effectiveness for binary sentiment classification.The linear modeling approaches performed well according to confusion matrix analysis yet they needed adjustments to decrease their false negative and false positive error rates.

Traditional machine learning models of the system produced dependable results yet future improvements may emerge through implementing superior techniques. Several deep learning models including LSTM and BERT proved capable of detecting advanced patterns and context-based relationships in reviews which leads to improved accuracy alongside better performance generalization. The performance of the system could be improved through ensemble methods and optimal hyperparameter selection of present models. The proposed sentiment analysis system establishes strong fundament for sentiment classification yet it creates multiple opportunities for enhancement throughout future research.

**7.References**

1. Hassan, M. E., Maab, I., Hussain, M., Habib, U., & Matsuo, Y. (2024). Polarity Classification of Low Resource Roman Urdu and Movie Reviews Sentiments using Machine Learning-based Ensemble Approaches. IEEE Open Journal of the Computer Society. <https://ieeexplore.ieee.org/abstract/document/10707202/>
2. Sarhan, A. M., Ayman, H., Wagdi, M., Ali, B., Adel, A., & Osama, R. (2024). Integrating machine learning and sentiment analysis in movie recommendation systems. Journal of Electrical Systems and Information Technology, 11(1), 53. <https://link.springer.com/article/10.1186/s43067-024-00177-7>
3. Deshmukh, N., Yadav, L., & Deharkar, A. (2024). Movie Recommendation System with Sentiment Analysis. Int Res J Modernization Eng Technol and Sci, 6(5). <https://www.researchgate.net/profile/Lowlesh-Yadav-2/publication/382530440_MOVIE_RECOMMENDATION_SYSTEM_WITH_SENTIMENT_ANALYSIS/links/66a218efc6e41359a83c05fe/MOVIE-RECOMMENDATION-SYSTEM-WITH-SENTIMENT-ANALYSIS.pdf>
4. Bhushan, S., Singh, A. K., Kumar, S., & Bhuyan, B. P. (2024). Mood-Based Movie Recommendation System Using Sentiment Analysis. In International Conference on Machine Learning, Advances in Computing, Renewable Energy and Communication (pp. 235-246). Springer, Singapore. <https://link.springer.com/chapter/10.1007/978-981-97-5231-7_20>
5. Bhaskar, G. (2024). MOVIE RECOMMENDATION SYSTEM USING SENTIMENT ANALYSIS FROM MICRO BLOGGING DATA. International Journal of Information Technology and Computer Engineering, 12(2), 567-580. <https://ijitce.org/index.php/ijitce/article/download/544/505>
6. Nor El-Deen, D., El-Sayed, R. S., Hussein, A., & Zaki, M. (2024). Sentiment analysis for movie recommendations: Harnessing opinion mining systems to analyze user reviews. The Egyptian Statistical Journal, 68(1), 1-14. <https://esju.journals.ekb.eg/article_358152.html>
7. Sharma, U., Saraogi, Y., Bhasme, S., Bharsakhale, V., Gupta, S., Khandare, H., & Deshmukh, S. S. (2024). Leveraging Machine Learning for Personalized Movie Recommendations with Sentiment Analysis. Journal Impact Factor, 3(2). <https://ijirid.in/3-2-24Apr/3-2-3-Uma%20Sharma-Yuvraj%20Saraogi-Shreyah%20Bhasme-Vedant%20Bharsakhale-Simran%20Gupta-Harshad%20Khandare-Prof%20S%20S%20Deshmukh.pdf>
8. Nkhata, G., Anjum, U., & Zhan, J. (2025). Sentiment analysis of movie reviews using bert. arXiv preprint arXiv:2502.18841. <https://arxiv.org/abs/2502.18841>
9. Bi, R. (2024). Long-term and short-term memory network based movie comment sentiment analysis. Applied and Computational Engineering, 36, 150-155. <https://www.ewadirect.com/proceedings/ace/article/view/10082>
10. Paneru, B., Thapa, B., & Paneru, B. (2025). Sentiment Analysis of Movie Reviews: A Flask Application Using CNN with RoBERTa Embeddings. Systems and Soft Computing, 200192. <https://www.sciencedirect.com/science/article/pii/S2772941925000109>

**8. APPENDIX**

1. Dataset link: [IMDb Movie Reviews Dataset | Papers With Code](https://paperswithcode.com/dataset/imdb-movie-reviews)

2.link to complete code: