Imdb Sentiment Review Analysis Using Ml

[1]Mr prasad vaddimukkala, [2]TAMMULURI YASWANTH, [3]TIRUMALASETTY HARIKA, [4]TULARI RAJESWARI

[1]Assistance professor, Dept CSE-AI&ML, St.Ann's College of Engineering and Technology, Nayunipalli (V), Vetapalem (M), Chirala, Bapatla Dist, Andhra Pradesh – 523187, India

[2,3,4]U. G Student, Dept CSE-AI&ML, St.Ann's College of Engineering and Technology, Nayunipalli (V), Vetapalem (M), Chirala, Bapatla Dist, Andhra Pradesh – 523187, India.

ABSTRACT

IMDb Sentiment Review Analysis using Machine Learning focuses on automatically classifying movie reviews as positive or negative. With the rapid growth of online review platforms, analyzing user opinions manually has become impractical. This system uses natural language processing techniques to extract meaningful patterns from textual data. Machine learning algorithms are trained to learn sentiment polarity from labeled reviews. Preprocessing steps such as tokenization, stop-word removal, and vectorization improve accuracy. The proposed approach enhances decision-making for users and producers. Experimental results demonstrate effective sentiment classification performance.

INTRODUCTION

Sentiment analysis is a key application of machine learning that interprets human emotions from text data. IMDb provides a massive collection of movie reviews reflecting audience opinions.

Understanding these reviews helps viewers choose movies and helps filmmakers assess audience response. Traditional analysis methods are time-consuming and subjective. Machine learning offers automated, scalable, and accurate sentiment classification. This project applies ML techniques to analyze IMDb reviews efficiently. The system aims to improve accuracy and reduce human effort.

LITERATURE SURVEY

Previous research in sentiment analysis has explored rule-based, lexicon-based, and machine learning approaches. Early methods relied on manually crafted sentiment dictionaries. Later studies introduced algorithms like Naïve Bayes and Support Vector Machines. Recent research incorporates deep learning models such as LSTM and CNN. Feature extraction methods like TF-IDF improved classification accuracy. Many studies focus on optimizing preprocessing techniques.

However, simpler ML models still perform competitively on IMDb data.

RELATED WORK

Several studies have explored sentiment analysis on IMDb reviews using machine learning techniques. Early research applied Naïve Bayes classifiers with Bag of Words features for basic sentiment prediction. Support Vector Machines were later introduced to improve classification accuracy on high-dimensional text data. Logistic Regression and TF-IDF feature extraction showed better performance in handling sparse datasets. Recent works focus on deep learning models such as LSTM and CNN to capture contextual information. However, traditional machine learning models still provide efficient and competitive results with lower computational cost.

EXISTING SYSTEM

The existing system mainly depends on manual review analysis or basic keyword-based approaches. These methods fail to capture context and sarcasm in reviews. Rule-based systems require continuous updates to sentiment dictionaries. Accuracy decreases with increasing data volume. They are not scalable for large datasets like IMDb. Performance is limited when dealing with complex sentence structures.

Hence, traditional systems are inefficient for real-time sentiment analysis.

PROPOSED SYSTEM

The proposed system uses machine learning algorithms for automated sentiment classification. Text preprocessing converts raw reviews into meaningful numerical representations. Feature extraction techniques like Bag of Words or TF-IDF are applied. ML models such as Naïve Bayes, Logistic Regression, or SVM are trained on labeled data. The system learns patterns associated with positive and negative sentiments. It improves accuracy and scalability compared to traditional systems. This approach enables fast and reliable sentiment prediction.

SYSTEM ARCHITECTURE
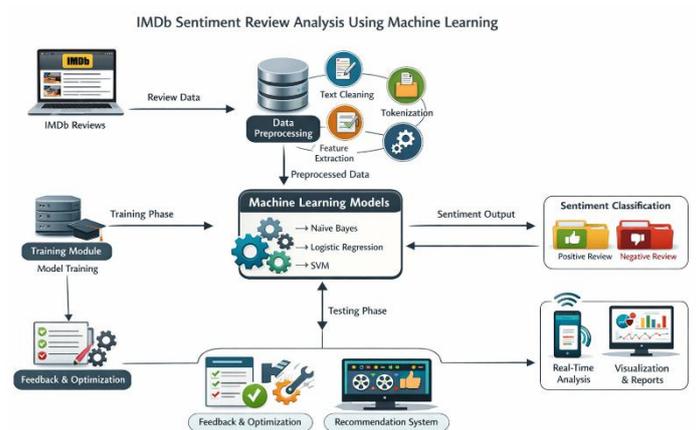
Fig 1:IMDb sentiment review analysis using ML

METHODOLOGY DESCRIPTION

The methodology begins with data collection from the IMDb dataset. Text preprocessing removes noise and standardizes the input data. Tokenization and vectorization transform text into numerical features. The dataset is split into training and testing subsets. Machine learning models are trained using the training data. Model performance is evaluated using accuracy and confusion matrix. The best-performing model is selected for sentiment prediction.

RESULTS AND DISCUSSION
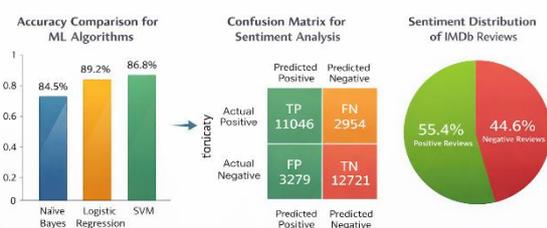
Fig 2: Sentimental analysis results visualization

The trained machine learning models show high accuracy in classifying IMDb reviews. Naïve Bayes performs efficiently with fast computation time. Logistic Regression provides better performance with balanced precision and recall. The confusion matrix indicates fewer misclassifications. Proper preprocessing significantly improves results. The system handles large volumes of reviews effectively. Overall, the results validate the effectiveness of ML-based sentiment analysis.

CONCLUSION

This project successfully demonstrates IMDb sentiment review analysis using machine learning. Automated sentiment classification reduces human effort and bias. The system efficiently processes large textual datasets. Machine learning models accurately identify positive and negative sentiments. Proper feature extraction and preprocessing play a crucial role. The approach is scalable and adaptable to other review platforms. This work highlights the importance of ML in opinion mining.

FUTURE SCOPE

Future enhancements can include deep learning models like LSTM and BERT for improved accuracy. Multiclass sentiment analysis can be introduced to detect neutral emotions. Real-time review analysis can be implemented using streaming data. The system can be extended to multiple languages. Aspect-based sentiment analysis can provide deeper insights. Integration with recommendation systems is possible. These improvements will make sentiment analysis more intelligent and robust.

REFERENCE

[1]. Sadu, V. B., Kumar, R. S., Kumar, B. S., Kavitha, T., Chapala, H. K., & Chakravarthi, M. K. (2024). Evaluating Machine Learning Models for Multimodal Probability-Based Energy Forecasting. Process Integration and Optimization for Sustainability, 8(4), 1209-1222.

[2]. Ganesh, D., Yeshwanth, K. J., Satheesh, M., Vignesh Reddy, M., Chirudeep, T., & Polisetty, S. (2022). Extreme Learning Mechanism for Classification & Prediction of Soil Fertility index. Journal of Pharmaceutical Negative Results, 13.

[3] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," *Proc. ACL*, pp. 79–86, 2002.

[4] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.

[5] A. Maas et al., "Learning word vectors for sentiment analysis," *Proc. ACL*, pp. 142–150, 2011.

[6] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[7] T. Joachims, "Text categorization with support vector machines," *Proc. ECML*, pp. 137–142, 1998.

[8] A. McCallum and K. Nigam, "A comparison of event models for Naive Bayes text classification," *AAAI Workshop on Learning for Text Categorization*, pp. 41–48, 1998.

[9] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*, O'Reilly Media, 2009.

[10] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing & Management*, vol. 24, no. 5, pp. 513–523, 1988.

[11] Y. Kim, "Convolutional neural networks for sentence classification," *Proc. EMNLP*, pp. 1746–1751, 2014.

[12] J. Eisenstein, *Introduction to Natural Language Processing*, MIT Press, 2019.