iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** XI **Month of publication:** November 2024

DOI: https://doi.org/10.22214/ijraset.2024.65043

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XI Nov 2024- Available at www.ijraset.com

Integrating Sentiment Analysis and Machine Learning for Improved Film Recommendation Systems

A.A.Patel[1], V.V.Ubhe[2], S.G.Vaidya[3], K.Deskmukh[4], V.Visal[5]
Pvpit College Bavdhan , Pune

Abstract: ***To do so, especially when it comes to digital entertainment platforms, which are evolving rapidly, giving users options for having what they want, and the importance of personalization has simply gone up only. This project focuses on movie recommendations using sentiment analysis, which increases the accuracy of recommendations by adding review sentiments based on movie ratings and reviews. This project presents a movie recommendation system that leverages collaborative filtering and sentiment analysis to deliver more accurate suggestions to movie enthusiasts, specifically targeting "Cinemaniacs" and "Film Junkies." The primary components of our approach include collaborative filtering, such as taking movie preferences from the user and performing natural language processing (NLP) techniques for sentiment analysis of user-generated reviews and ratings. By integrating user preferences, such as selected genres and viewing habits, with sentiment-driven insights from movie reviews, the system extends beyond traditional numeric ratings to capture nuanced satisfaction and dissatisfaction. The platform also includes genre-specific recommendations and predictive models that forecast upcoming movie releases based on individual tastes, trends, and pre-release sentiments. This innovative combination provides users with highly customized movie and web series suggestions, keeping them updated on new releases, while offering trailer predictions to preview potential interests. Through this multi-layered approach, the project aims to deepen viewer engagement by closely aligning content recommendations with emotional responses and genre preferences***

Keywords: Recommendation System, numerical rating, interest-based recommendation, genre preferences, trailer prediction, machine learning, sentiment analysis, natural language processing, text mining, cosine similarity.

I. INTRODUCTION

In the current digital age, streaming platforms offer a wide variety of movies and TV shows, making it difficult for users to find content that matches their interests. Most recommendation systems rely on ratings and viewing history to suggest content, but these often do not result in negative feedback from the users. Collaborative filtering helps identify patterns in user behavior and groups people with similar interests. However, this can overlook the experiences that users share in their reviews. To address this gap, the system uses natural language processing (NLP)-based sentiment analysis, which captures the emotions expressed by users (positive, negative, or neutral). Videos can be recommended not only to align with users' historical interests but also with their past thoughts. The combination of collaborative filtering and sentiment analysis is designed to create a more engaging and enjoyable user experience, helping users discover content that resonates with and is positively influenced by others. This project explored how thinking can increase the accuracy of recommendations and provide new insights into user recommendations.

II. LITERATURE SURVEY

1) *Implementation of Movie Recommendation System Using Hybrid Filtering Methods and Sentiment Analysis of Movie Reviews (2024)*

This study presents a method for incorporating movie review sentiment analysis into a hybrid recommendation system for streaming platforms. The study focuses on 4890 movies using a broad dataset containing detailed descriptions of the movies along with reviews. To employ demographic filtering, the popularity score of the movies was calculated, collaborative filtering was applied, and textual movie descriptions were vectorized using the countvectorizer method. The high-accuracy model "ControX/Sen1" was used to predict the sentiment of movie reviews. This hybrid recommendation system ranked the movies based on the user's preferences by employing cosine similarity and was further filtered with positive sentiment reviews.

©IJRASET: All Rights are Reserved

By including sentiment analysis, this research advances sophisticated movie recommendation systems by providing a comprehensive method for addressing user preferences and emotional resonance in film selection.

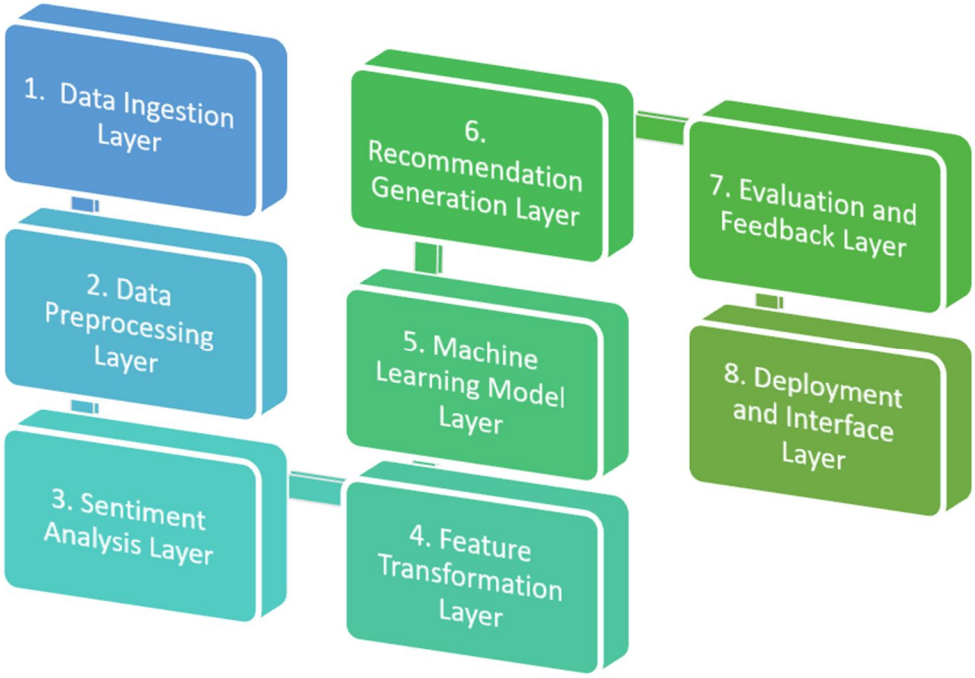
2) Machine Learning-Based Sentiment Analysis of Movie Review (2023)

This research focuses on sentiment analysis, which is a rapidly growing field in machine learning that can provide valuable insights into audience opinions and preferences in various domains, such as movie reviews. Sentiment analysis can help marketing efforts and improve the quality of products or services.

3) New Machine Learning Model to Movie Recommender and Sentiment Analysis (2023)

This study proposes a new supervised machine learning model for recommending movies, addressing the issue of an overload of massive data in the movie industry. The proposed model combines cosine similarity, sentiment analysis, and naïve Bayes or support vector machine classifiers to provide more accurate and efficient movie recommendations for users. This work is essential for improving the user experience in movie recommendation systems and reducing the time and effort required for users to find movies that match their interests.

III. SYSTEM ARCHITECTURE



1) Data Ingestion Layer

The system begins by aggregating movie review data and metadata from multiple sources, thereby providing a comprehensive dataset for training and recommendation.

Review Data Collection: Sources include IMDb, Rotten Tomatoes, and social media platforms, focusing on English-language reviews for consistency. Each review included sentiment-rich text and metadata.

Metadata Gathering: Collects relevant information such as movie genre, cast, director, and release date, enhancing the recommendation context.

Data Aggregator Module: Centralizes data across sources, standardizing it into a single structured repository, and ensuring consistent processing.

2) Data Preprocessing Layer

Data preprocessing transforms raw text into a structured input, preparing it for sentiment analysis and classification.

Text cleaning: Extraneous characters, including punctuation, HTML tags, and URLs, are removed to improve text quality.
Tokenization and Lemmatization: Tokenizes text and applies lemmatization, converting words to their base form (e.g., "watched" becomes "watch") to reduce dimensionality.
Stop word removal: Common words that do not carry sentiment are filtered out, focusing on the analysis of sentiment-rich words.
Feature Engineering: Extracts basic linguistic features, such as review length, word frequency, and n-grams, contributing to sentiment strength and clarity.

3) Sentiment Analysis Layer

Sentiment analysis quantifies user opinions and converts qualitative texts into quantitative sentiment scores.
Sentiment Scoring Engine: Uses Valence Aware Dictionary and sEntiment Reasoner (VADER) to score each review between -1 (negative) and +1 (positive). This tool was optimized for short-form text analysis, balancing efficiency, and accuracy.
Aspect-Based Sentiment Analysis (ABSA) (optional): Identifies sentiment associated with specific movie aspects (e.g., plot, visuals, acting), offering an aspect-specific sentiment score for each aspect.
Sentiment classification: Each review is categorized into positive, neutral, or negative classes based on empirically set thresholds.
Positive: >0.5
Neutral: between -0.5 and 0.5
Negative: < -0.5
Sentiment aggregation: The aggregated sentiment score for each movie is calculated by averaging the individual review scores, providing a consensus sentiment.

4) Feature Transformation Layer

This layer structures sentiment data and additional review features into vectors for machine learning.
Vectorization: Converts cleaned text data into numerical vectors using TF-IDF, capturing the relevance of specific words and phrases.
Feature Vector Construction: Combines aggregated sentiment scores, aspect-specific scores, and other review metadata into a feature vector for each movie.
Normalization and Scaling: Ensures numerical consistency by normalizing data, which is critical for algorithms sensitive to feature scaling.

5) Machine Learning Model Layer

This core layer applies the random forest classifier, which is a powerful ensemble method, to identify complex patterns in the sentiment data.
Random Forest Classifier: Trained on feature vectors, this model predicts the recommendation relevance of movies based on user sentiment. Random forest handles high-dimensional and nonlinear data, making it ideal for nuanced sentiment features.
Hyperparameter Tuning: Hyperparameters such as the number of estimators, maximum tree depth, and minimum samples per split were optimized through grid search and cross-validation, ensuring high accuracy without overfitting.
Feature Importance Analysis: Identifies the most predictive features (e.g., sentiment scores and review length) in movie ranking, providing insights into which sentiment factors most influence recommendations.

6) Recommendation Generation Layer

The recommendation layer applies ranking algorithms to prioritize movies based on sentiment scores, and filtering results to deliver only the most relevant recommendations.
Movie Ranking Engine: Ranks movies by their sentiment-based aggregate score, assigning a higher weight to those with a positive sentiment consensus.
Threshold-Based Filtering: Filters movies below a certain sentiment threshold, ensuring that recommendations are both relevant and positive.
Personalization Module (optional): Incorporates additional user preferences (e.g., favorite genres) to refine recommendations, making them highly tailored.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

7) Evaluation and Feedback Layer

This layer measures the recommendation quality of the model and continuously improves it based on real-time feedback.

Quantitative Metrics Evaluation: Calculates metrics such as precision, recall, F1 score, and AUC to assess the classification accuracy.

User Feedback Integration (optional): Collects user feedback on recommended movies to enhance personalization and fine-tune the model in future iterations.

Model Retraining Trigger: Automatically initiates retraining based on feedback and new data, keeping the system aligned with recent trends and preferences.

8) Deployment and Interface Layer

The deployment layer makes recommendations accessible to end users through an API and provides an intuitive user interface.

API Interface: RESTful API exposes recommendations to front-end applications, enabling integration with the web or mobile platforms.

User Interface: Offers an interface where users can view personalized recommendations and provide feedback.

Logging and Monitoring Module: Tracks system performance, user interactions, and error rates, ensuring system reliability and offering data for ongoing refinement

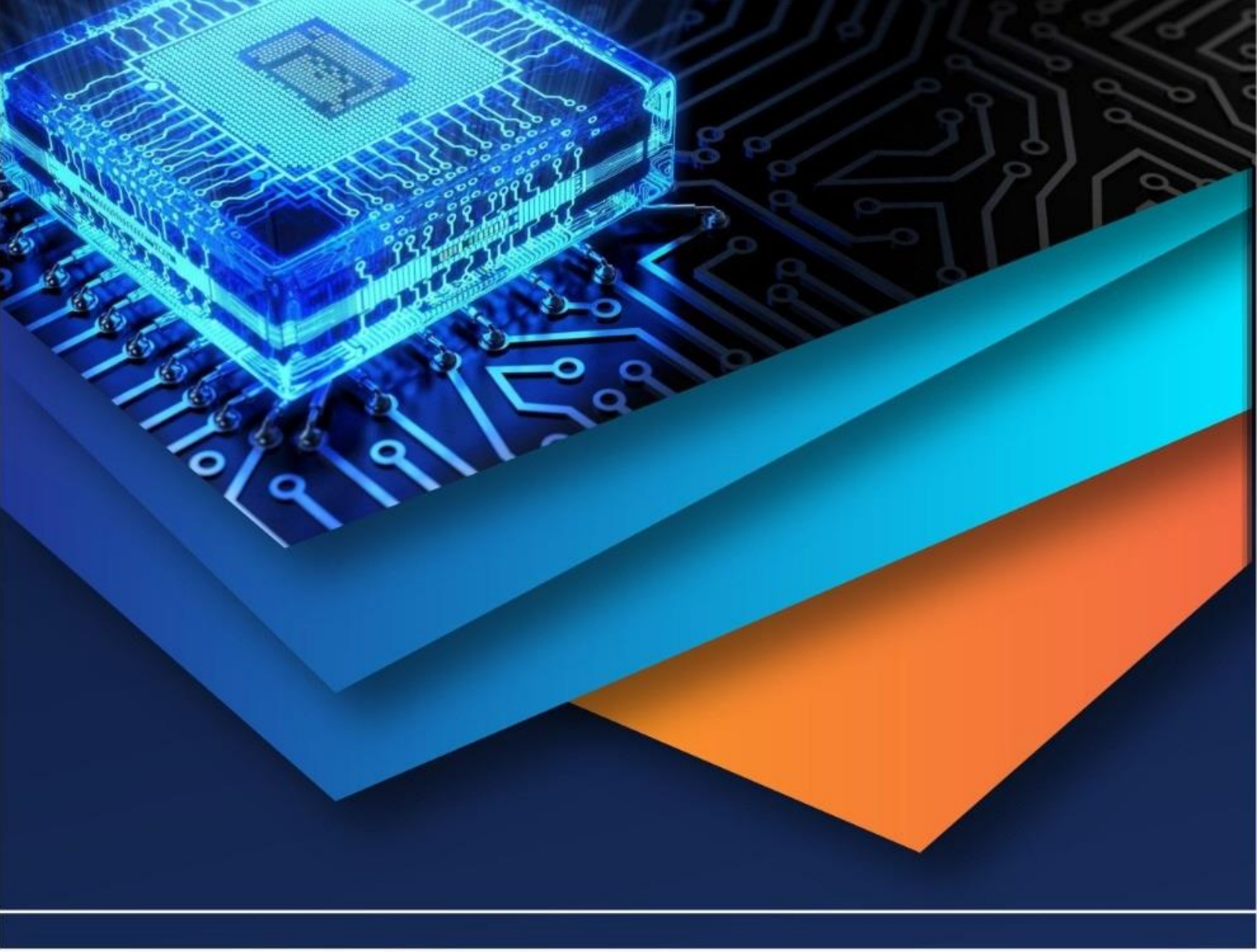
IV. CONCLUSION

In this study, we developed a movie recommendation system that leverages sentiment analysis and random forest classification to enhance the recommendation accuracy and user satisfaction. By integrating natural language processing with a robust machine learning classifier, the system effectively captures nuanced user preferences embedded in textual reviews and transforms subjective sentiment into actionable insights. The use of sentiment scoring and, optionally, aspect-based sentiment analysis ensures that recommendations are not only based on aggregate ratings, but also on deeper emotional feedback related to specific movie attributes, such as acting, plot, and visuals.

The random forest model's capability to handle high-dimensional data and identify complex patterns has proven to be well suited for this recommendation task. Feature engineering and threshold-based filtering further optimize the recommendation pipeline, prioritize movies with high positive sentiments, and reduce the influence of neutral or mixed reviews. The system's architecture is designed to be modular and scalable, allowing for easy integration with new data sources, updates to NLP methods, and real-time user feedback, thereby supporting ongoing improvements and adaptability to evolving trends

REFERENCES

[1] B. Pang, L. Lee et al., "Opinion mining and sentiment analysis," Foundations and Trends® in information retrieval, vol. 2, no. 1–2, pp. 1–135, 2008.
[2] Naskar, I. & Joseph, N. P. (2024, February 16 2024). Implementation of Movie Recommendation System Using Hybrid Filtering Methods and Sentiment Analysis of Movie Reviews. https://doi.org/10.1109/icwite59797.2024.10502695
[3] Marappan, R., & R, S. (2023, November 23). New Machine Learning Model to Movie Recommender and Sentiment Analysis. https://doi.org/10.1109/icaiccit60255.2023.10465888
[4] Soubraylu, S., and Rajalakshmi, R., "Hybrid convolutional bidirectional recurrent neural network-based sentiment analysis on movie reviews", Computational Intelligence, 37(2), 735-757, 2021.
[5] M Viswa, Murali Vishnu, and TG Nancy Victor, "A Collaborative Filtering based Recommender System for Suggesting New Trends in Any Domain of Research", 2019, 978-1-5386-9533-3/19/$31.00 ©2019 IEEE
[6] Van Meteren, Robin, and Maarten Van Someren. "Using content-based filtering for recommendation." In Proceedings of the Machine Learning in the New Information Age: MLnet/ECML2000 Workshop, Vol. 30, pp. 47-56. 2000.
[7] ControX/Sen1. "ControX/Sen1: Sentiment Analysis Model based on BERT." Huggface Model Hub 2022. [Online]. Available: https://huggingface.co/ControX/Sen1
[8] B. Pang, L. Lee et al., "Opinion mining and sentiment analysis,"Foundations and Trends® in information retrieval, vol. 2, no. 1–2, pp. 1–135, 2008.® in information retrieval, vol. 2, no. 1–2, pp. 1–135, 2008.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)