INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

Sentiment Analysis for Movie Database Management System Using Naive Bayes Algorithm

[1]Rhonnel S. Paculanan, [2]Jennifer T. Carpio, [3]Rosemarie M. Perreras, [4]Melchor G. Erise, [5]Shaine Maglapuz

[1,4]Arellano University

[2,3]San Beda University

Jose Rizal University

DOI : https://doi.org/10.51583/IJLTEMAS.2024.131224

Received: 20 December 2024; Accepted: 01 January 2025; Published: 14 January 2025

Abstract: According to F. Furtado and A. Singh (2021), movie recommendation systems aim to assist viewers by suggesting movies to watch, thereby avoiding the time-consuming and complex process of selecting from a vast library of movies, which can number in the thousands or even millions. This article seeks to reduce human effort by recommending movies based on user interests. To address these challenges, we propose a technique that combines content-based and collaborative approaches. This method produces more precise results than systems that rely solely on content-based methods. Content-based recommendation systems are limited to individual preferences and do not recommend items outside the user's immediate interests, restricting exploration. To overcome these limitations, we developed a system that addresses these concerns. The proposed movie recommendation system uses Cosine Similarity to suggest movies like the one chosen by the user. While existing algorithms generate recommendations, they often fail to answer the critical question, "Is this a film worth watching?" To enhance the overall experience, our system incorporates sentiment analysis of selected movie reviews using machine learning techniques. This study employs two supervised learning approaches, Naïve Bayes (NB) and Support Vector Machine (SVM), to improve accuracy and efficacy. A comparison of the two methods reveals that SVM achieved an accuracy score of 98.63%, while NB scored 97.33%, as reported by N. Pavitha et al. (2022). The project integrates a sentiment analysis module into the movie management system to enhance decision-making and user experience. The Naïve Bayes approach is used to classify and evaluate user reviews and feedback into positive, negative, and neutral categories. This analysis provides insights into audience preferences, overall satisfaction, and movie reception. The system utilizes this information to refine movie recommendations, analyze trends, and enhance content selection, all of which contribute to a more personalized user experience. The system is evaluated using the ISO 25010 framework, focusing on functionality, reliability, usability, effectiveness, and robustness. Findings show that respondents actively participated in the review process across multiple domains. Key results indicate high satisfaction levels in terms of ease of use, interface design, and system performance. Feedback also highlighted areas for improvement, such as optimizing loading times and enhancing specific features for a smoother user experience. Overall, the findings reflect positive engagement with the system, demonstrating its effectiveness in meeting user needs while identifying opportunities for future enhancements.

Keyword: Sentiment Analysis, Movie Database Management System, Naive Bayes Algorithm

I. Introduction

According to F. Furtado and A. Singh (2021), movie recommendation systems aim to assist moviegoers by suggesting movies to watch, rather than requiring them to navigate the time-consuming and complex process of selecting from a vast library of options, which can number in the thousands or even millions. This article seeks to reduce human effort by recommending movies based on user preferences. To address these challenges, we propose a methodology that combines both content-based and collaborative approaches, producing more explicit results than systems relying solely on content-based methods.

Content-based recommendation systems are limited to individual preferences and do not suggest items beyond a user's immediate interests, restricting exploration. To overcome this limitation, we focused on developing a system that addresses these concerns. According to N. Pavitha et al. (2022), this technique is particularly effective in managing large datasets and quickly identifying information that aligns with a user's preferences.

A movie recommendation system is presented that uses Cosine Similarity to suggest movies like the one selected by the user. While existing algorithms generate recommendations, they often fail to address the question, "Is this a film worth watching?" To improve the overall user experience, our system incorporates sentiment analysis of selected movie reviews using machine learning techniques.

This research employs two supervised learning methods: Naïve Bayes (NB) and Support Vector Machine (SVM), to enhance accuracy and effectiveness. A performance evaluation of these methods shows that SVM achieved an accuracy score of 98.63%, outperforming NB, which scored 97.33%. Therefore, SVM is better suited for sentiment analysis tasks.

Every year, new films are released with diverse storylines and genres that appeal to a wide range of audiences. Online movie and video streaming platforms can retain user engagement by recommending movies that align with viewer preferences. Developing more personalized recommendations remains a significant challenge for recommendation engines. This article discusses filtering

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

options, including content-based filtering, which suggests movies based on a user's viewing history and preferences, and collaborative filtering, which uses the opinions and actions of similar users to make recommendations.

Sakina Salmani et al. (2021) implemented and compared collaborative filtering approaches, including user-based, item-based, SVD, and SVD++ algorithms. They demonstrated that a hybrid recommendation engine combining content-based filtering and SVD filtering models achieves the best performance and generates superior movie suggestions, ensuring sustained viewer engagement.

Recommender systems are models of user preferences designed to suggest products or services to explore or purchase. The Intelligent Movie Database Recommendation System Integrates Human-Computer Interaction and Machine Learning techniques. The proposed method is an advanced information filtering system that captures a viewer's facial expressions and emotions to recommend films dynamically. It also suggests movies based on user demographics, such as age, gender, and genre preferences. The recommended movie list is further refined using previous user ratings and reviews.

Objectives: The project aims to integrate a sentiment analysis module into the movie management system to enhance decision-making and improve the user experience. In this study, the Naïve Bayes approach is used to accurately classify and evaluate user reviews and feedback as positive, negative, or neutral. This study contributes to practice by uncovering audience preferences, general audience satisfaction, and overall movie reception. This information enables the system to inspire movie recommendations, analyze trends, and enhance content selection, all of which contribute to a more personalized and enriched user experience. The system will be evaluated using the ISO 25010 framework, focusing on characteristics such as functionality, reliability, usability, effectiveness, and robustness.

Conceptual Framework: This framework examines the Movie Database Management System using sentiment analysis and the Naïve Bayes algorithm. The primary goal is to improve user feedback analysis and personalize the user experience by studying emotional responses and sentiments across various movie genres and plots. Understanding the principles underlying movie management systems requires leveraging sentiment analysis to interpret user input and emotional reactions. This approach aims to provide personalized movie recommendations based on user preferences and past behavior, such as ratings and viewing history. The Naïve Bayes algorithm is utilized to recommend highly rated movies, enabling informed decision-making and enhancing user satisfaction.

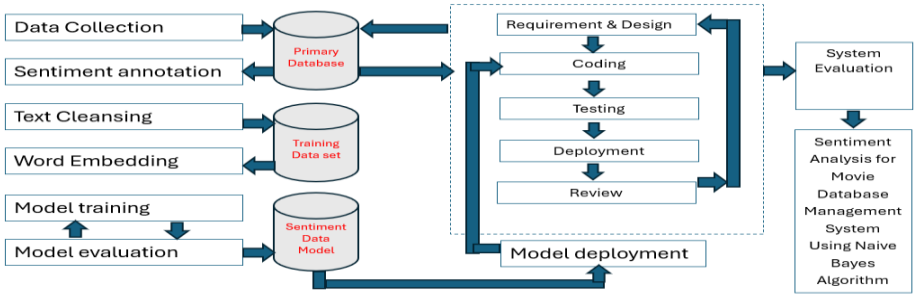

Figure 1. Conceptual Framework of Sentiment Analysis for Movie Database Management System Using Naive Bayes Algorithm

Figure 1 illustrates the parallel development process of the proposed system, which consists of two main components: sentiment model development using the Naïve Bayes method and system development using the Agile Scrum Model. The process begins with data collection, followed by sentiment annotation, where text data is labeled with tags reflecting the sentiment or emotional tone expressed in the content. This is followed by text cleansing and word embedding, which represents words as dense, continuous vectors in a multi-dimensional space. Next is model training, the process of teaching a machine learning model to identify patterns and relationships from labeled data, and finally, model evaluation, which assesses how well the trained model performs on unseen data. These steps collectively contribute to the development of the Sentiment Analysis for Movie Database Management System Using the Naïve Bayes Algorithm.

Simultaneously, system development will follow the Agile Scrum Model, involving requirements gathering, analysis and design, coding, testing, implementation, and review, all conducted in Python for the Movie Database Management System. To ensure the quality of the proposed system, it will be evaluated using a validated researcher-made questionnaire patterned after ISO/IEC 25010 standards.

According to Aravinth, Nivethitha, Saranya, and Nivethanandhini (2019), a vast amount of data in the form of tweets, blogs, status updates, and posts is present on social media and online platforms. Fortunately, sentiment analysis of this data allows us to understand the general sentiment of individuals or groups. Sentiment analysis refers to the examination of feelings and thoughts expressed in textual content.

Mondal (2023) states that the Movie Recommendation System with Sentiment Analysis (MRSA) predicts movie sentiment by utilizing IMDb ratings. The algorithm considers several variables, such as the frequency of votes on a movie and whether those

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

votes are positive or negative. Additionally, some researchers suggest that a person's voting history influences whether they like or dislike a particular movie.

II. Synthesis of Literature Review

The extensive data available on social media and online platforms makes sentiment analysis feasible, enabling the extraction of sentiments and perspectives from text. This study employs machine learning techniques to compare tweets about movies with reviews from the Times of India, providing valuable insights for market research and the film industry. A Movie Recommendation System with Sentiment Analysis can predict sentiment and recommend movies that align with users' emotional states by leveraging IMDb ratings and analyzing personal preferences and feelings.

A custom-developed website collects data from Kaggle and IMDb, offering users easy access to review classifications, recommendations, and movie information. The study emphasizes a personalized recommendation system built using the MovieLens Dataset and the Naïve Bayes algorithm. Opinion mining, including sentiment analysis, is used to train classifiers for precise sentiment classification and to categorize movie reviews on platforms like Twitter. Online reviews significantly influence movie suggestions, with machine learning techniques applied to recommend movies based on users' prior viewings. Additionally, sentiment analysis is utilized to provide personalized recommendations for Korean dramas based on user ratings and feedback.

Combining AI-based similarity algorithms with collaborative, content-based, and hybrid filtering methodologies enhances recommendation accuracy. An Adaptive Collaborative Intelligent Recommendation model integrates user behavior with adaptive filtering techniques to deliver precise recommendations. Despite advancements in the field, sentiment analysis remains challenging due to subjective interpretations, particularly when dealing with brief social media texts. However, innovations such as sentiment phrase pattern matching and advanced neural networks have improved upon earlier algorithms that relied on SVM and unigram features.

Automated systems like Opinion Finder and Pulse play a crucial role in sentiment analysis for applications such as political opinion mining, film reviews, and branding. By combining lexicon-based techniques with deep learning, real-time analysis across social networks has been enhanced, addressing challenges like data sparsity and the cold-start problem.

Research Design. The project employs a convergent mixed-methods design, incorporating both qualitative and quantitative data collection. Its primary goal is to develop a sentiment analysis framework for a movie management system using the Naïve Bayes algorithm. Alongside the qualitative data gathered through structured interviews and open-ended questions, the research collects quantitative data from film reviews via surveys, polls, and questionnaires.

The collected data is processed, tagged, and categorized into three sentiment types: Positive, Negative, and Neutral, before being input into the Naïve Bayes classifier. Quantitative data is statistically analyzed, evaluated, and presented in the form of tables, figures, and charts to identify trends, patterns, and correlations. Simultaneously, a thematic analysis will be conducted on the qualitative data to extract insights and identify recurring themes.

The structured approach aims to generate quantifiable and actionable information to enhance decision-making mechanisms within the movie industry. By analyzing the movie management system and improving recommendation accuracy, the project seeks to maximize client satisfaction and fulfillment.

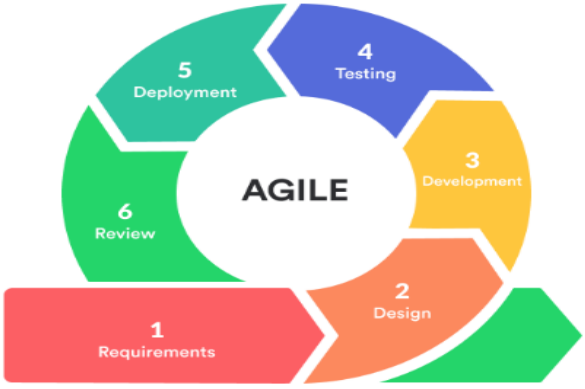

Figure 2. Agile Scrum Development Life Cycle

Figure 2 illustrates the Agile Scrum Development Life Cycle, which ensures efficient development aligned with user needs. This process involves defining project goals, implementing sentiment analysis to enhance movie recommendations, conducting feasibility studies, and gathering initial requirements. A team of developers, designers, and testers is formed to execute the project. The phases of this model include requirements and design, coding, testing, deployment, and review. By following this structured approach, the software developed is expected to meet user expectations and adhere to user standards.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

Website security:

1. aUser Authentication and Authorization:

1.1 Strong Password Policies: Enforce strong password creation rules (e.g., minimum length, complexity) to ensure that user accounts are protected from unauthorized access.

1.2 Multi-Factor Authentication (MFA): Implement MFA to add an extra layer of security, requiring users to provide two or more verification factors (e.g., a password and a one-time code sent to their phone).

2. Data Encryption

2.1 SSL/TLS Encryption: Use SSL/TLS protocols to encrypt data transmitted between the user's browser and the server, protecting sensitive information such as login credentials, movie reviews, and sentiment analysis results.

2.2 Data Encryption at Rest: Ensure that sensitive data, such as user reviews and feedback, is encrypted when stored in the database to prevent unauthorized access in case of a breach.

3. Database Security

3.1 SQL Injection Prevention: Use prepared statements or ORM (Object-Relational Mapping) libraries to prevent SQL injection attacks by ensuring that queries are executed with sanitized input.

3.2 Database Access Control: Restrict database access to only authorized users and applications, ensuring that sensitive data is not exposed unnecessarily.

4. Sentiment Analysis Data Security

4.1 Data Anonymization: Ensure that user data used for sentiment analysis (e.g., reviews, comments) is anonymized to protect user privacy. This could involve removing personally identifiable information (PII) from the data before processing.

4.2 Data Retention Policies: Implement clear data retention policies to limit how long user data is stored. Ensure that data is securely deleted after it is no longer needed.

5. Security for Sentiment Analysis Algorithms

5.1 Model Protection: Protect the sentiment analysis model from unauthorized access or tampering, ensuring that the algorithm used to process movie reviews remains secure.

5.2 Model Integrity Checks: Regularly verify the integrity of the Naive Bayes model and its associated data to ensure that it hasn’t been altered or corrupted by an external attacker.

6. Incident Response Plan

6.1 Breach Detection and Response: Establish a plan for detecting, responding to, and recovering from security breaches. This should include notifying affected users and relevant authorities in case of a data breach.

6.2 Disaster Recovery: Implement a disaster recovery plan to quickly restore service in the event of a major security incident or system failure.

Statistical treatment methods, such as percentage and weighted mean using a Likert scale, were applied to ensure accurate interpretation of the data. The study evaluated a total of 1,000 respondents, consisting of 880 users and 120 IT professionals. The data were analyzed and interpreted based on the ISO 25010 framework, focusing on the following characteristics: functionality, reliability, usability, effectiveness, and robustness.

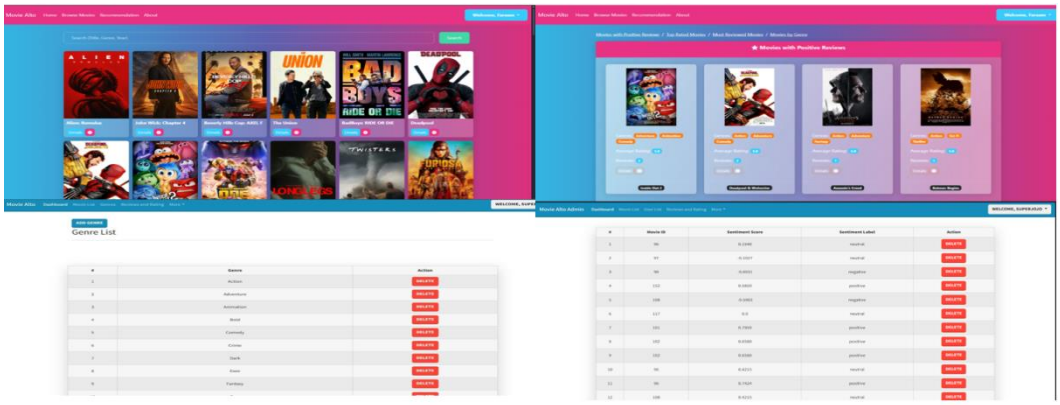

Figure 3. The Sentiment Analysis for Movie Database Management System Using Naive Bayes Algorithm Graphical User Interfaces

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

The figure 3 has 2 categories of user such as the User and the Administrator.

User Category: The figure on the top-left depicts the "Browse Movies" interface, which offers easy navigation and functionality. It allows users to browse movies in a grid structure and quickly access detailed information via a "Details" button. The search bar enables users to filter results by title, genre, or year, making it easier to find specific movies. A personalized greeting adds a nice touch, and the navigation menu simplifies access to other sections. This design further supports the project's goals by making movie discovery simple, intuitive, and personalized. The "User Recommendation Tab," shown on the top-right, enhances the user experience by providing curated movie recommendations based on ratings, reviews, and genres. It helps users discover films that match their preferences and provides comprehensive information to assist in making informed decisions. The "Details" button offers interactivity, which aligns with the project's goal of offering tailored movie recommendations and improving user satisfaction.

Administration Category: The bottom-left figure shows the "Genre List" in the admin panel of the movie management system, allowing administrators to manage movie genres by adding or deleting entries. This feature supports the system's objectives by organizing movies effectively, enabling precise sentiment analysis of user reviews, and improving the accuracy of recommendations. It also helps identify trends and tailor content strategies to align with user preferences. The bottom-right figure displays the "Admin Sentiment Score" tab, which shows sentiment analysis results for user reviews, including scores and labels (positive, negative, or neutral). This supports the system's objectives by helping administrators analyze feedback trends, refine movie recommendations, and make data-driven decisions.

Table 1: ISO 25010 characteristic Average Mean using 1000 respondents

Respondent(1000)	User (880)				Technical (120)			
ISO 25010	Male (490)		Female (390)		Male (73)		Female (47)	
	WM	VI	WM	VI	WM	VI	WM	VI
Functionality	3.55	SA	3.61	SA	3.01	SA	3.71	SA
Reliability	3.46	SA	3.54	SA	2.94	A	3.45	SA
Usability	3.49	SA	3.63	SA	2.91	A	3.25	SA
Effectiveness	3.45	SA	3.45	SA	2.99	A	3.3	SA
Robustness	3.47	SA	3.56	SA	3.65	SA	3	SA
Average Mean (4.00)	3.48	SA	3.56	SA	3.1	SA	3	SA

The table indicates the ISO 25010 evaluation, where 880 user respondents, both male and female, participated. Male respondents had an overall weighted average mean of 3.48, with a verbal interpretation of 'Strongly Agree,' while female respondents had an overall weighted average mean of 3.56, also with a verbal interpretation of 'Strongly Agree.' Additionally, 120 technical respondents were evaluated. Male respondents had an overall weighted average mean of 3.1, with a verbal interpretation of 'Strongly Agree,' while female respondents had an overall weighted average mean of 3.0, also with a verbal interpretation of 'Strongly Agree.' The respondents strongly agreed that the website adheres to the qualifications outlined in the ISO 25010."

III. Summary of findings

The findings show that respondents actively participated in the review process across multiple domains. Key insights indicate that most respondents provided feedback on the system's operation, usability, and user satisfaction. The data revealed high satisfaction levels regarding ease of use, interface design, and system performance. Additionally, feedback highlighted areas for improvement, such as optimizing loading times and enhancing certain features for a smoother user experience. Overall, the findings reflect positive engagement with the system, demonstrating its effectiveness in meeting user needs while identifying specific areas for future enhancement.

IV. Conclusion

This study effectively demonstrated that the system met its objectives, particularly in improving user experience and providing seamless functionality. The system proved to be user-friendly, intuitive, and efficient, with respondents praising the simplicity of navigation and the appealing design. Features such as responsiveness, real-time feedback, and interactivity were well-received, underscoring the system's usefulness. However, areas for improvement were identified, including faster loading times, enhanced security, and greater scalability to accommodate a larger user base. Addressing these issues will enable the system to evolve in response to changing user expectations and technological advancements. The findings of this study align with previous research on movie recommendation systems, which emphasizes that user satisfaction in such platforms is strongly linked to ease of use, accurate customization, and system responsiveness. The addition of a Naïve Bayes-based sentiment analysis module, which categorizes user comments as positive, negative, or neutral, helps to meet these standards by improving recommendation accuracy and trend analysis. This technique is consistent with trends in advanced recommendation systems, where sentiment data is used to enhance user engagement and personalize recommendations. Consequently, this study adheres to best practices in recommendation systems research, showing significant promise for developing a reliable, accessible platform that adapts to user needs and technological progress.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

Recommendations: The following suggestions are made to further improve the system:

• Strengthen security measures to protect user data and privacy.

• Improve recommendation accuracy by tailoring suggestions more closely to users' previous movie preferences.

• Enhance customer satisfaction by designing a professional, visually appealing user interface.

• Improve sentiment analysis to provide more accurate insights from user reviews.

• Create an automated movie dataset to serve as a reference for future research in this field.

References

1. Furtado, F., & Singh, A. (2020). Movie recommendation system using machine learning. DOAJ (DOAJ: Directory of Open Access Journals). https://doi.org/10.22105/riej.2020.226178.1128
2. Pavitha, N., Pungliya, V., Raut, A., Bhonsle, R., Purohit, A., Patel, A., & Shashidhar, R. (2022). Movie Recommendation and Sentiment Analysis Using Machine Learning. Global Transitions Proceedings, 3(1). https://doi.org/10.1016/j.gltp.2022.03.012
3. Sakina Salmani, Sarvesh Kulkarni (2021), Hybrid Movie Recommendation System Using Machine Learning: DOI:10.1109/ICCICT50803.2021.9510058: Conference: 2021 International Conference on Communication information and Computing Technology (ICCICT)
4. H. Swathi ; Mr. S. S. Aravinth ; V. Nivethitha ; T. Saranya ; R. Nivethanandhini (2019): Sentiment Analysis Of Movie Review Using Data Analytics Techniques: Iconic Research And Engineering Journals Publisher: IRE Journals: e-ISSN: 2456-8880
5. Zhang, W. (2020). Sentiment analysis and web development of movie reviews using Naive Bayes and LSTM. Carolina Digital Repository. https://doi.org/10.17615/cvr7-kt78
6. Hybrid movie recommendation System using Machine learning. (2021, June 25). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/9510058
7. How you like that?: Development of a Korean drama recommendation system through sentiment analysis. (2024, January 28). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/10457273/