Grenze International Journal of Engineering and Technology, June Issue

Movie Recommendation System using Hybrid Models, Agentic AI, and Prompt-based Approach

Jane Rubel Angelina Jeyaraj[1], M. Vignesh[2], A. Sanjeev Charan[3], M. Bala Datha Sai[4] and B. Daya Sagar[5]
[1]Associate Professor, Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Virudnagar, Tamilnadu
Email: j.janerubelangelina@klu.ac.in
[2-5]Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education Virudnagar, Tamilnadu
Email: vigneshmummadisetty4@gmail.com, sanjeevcharan0288@gmail.com, dattasai620167@gmail.com, bashyamdaya@gmail.com

***Abstract*— Current movie suggestion tools often fail to grasp the full context of what a viewer is seeking. Methods that rely on collective ratings or simple movie attributes offer limited personalization. This study introduces a new system that merges semantic analysis from advanced language models with a deep learning ranker to provide more adaptive and context-sensitive recommendations. The process starts by converting user queries and film descriptions into rich numerical representations. A primary ranking module then analyses these alongside user data to score potential matches. For new users, a separate mechanism promotes a varied selection across genres to gather initial preferences. A key feature is a natural language interface where users can express their mood, which the system translates into genres to guide its search. Tests confirm that this hybrid approach improves suggestion accuracy, diversity, and relevance over standard techniques, pointing toward a more intuitive and responsive recommendation experience.**

***Keywords*— Semantic understanding models, dynamic recommendation systems, intelligent scoring networks, user intent analysis, community-based filtering, content feature evaluation, new member onboarding, conversational search input.**

I. INTRODUCTION

The digital entertainment sector is now defined by an unprecedented volume of available content. As streaming services multiply and film catalogs expand, viewers frequently encounter decision fatigue, finding it difficult to select movies that match their specific preferences and immediate interests. This environment has made sophisticated recommendation systems essential for navigating the vast digital library, moving beyond simple popularity charts to offer truly individualized suggestions. Established algorithmic methods, including collaborative and content-based filtering, have historically formed the backbone of these systems. The former operates on the principle of user similarity, suggesting items enjoyed by like-minded individuals. The latter focuses on item attributes, recommending titles with features similar to those a user has previously appreciated. However, these techniques are hampered by significant constraints. The "cold- start" issue impedes recommendations for new users or films with no interaction history, while a "semantic gap" prevents these models from understanding the nuanced emotional or situational context behind a user's choices.

Grenze ID: 01.GIJET.12.2.380_24

Breakthroughs in artificial intelligence, particularly in natural language processing (NLP), offer a path forward. Powerful large language models (LLMs) possess a deep understanding of semantic meaning and context. By processing text such as plot summaries or user-generated reviews, these models can infer thematic and emotional undertones that traditional metadata misses. This allows for a more profound alignment between a film's essence and a viewer's intent Leveraging this potential, our study presents a novel recommendation architecture that merges the semantic power of LLMs with a multi-stage ranking system. The framework is built to adapt dynamically. Its initial stage (Rank0) manages new users by promoting a diverse array of films to quickly build a preference profile. Subsequently, a more advanced ranking model (Rank1) takes over, employing a deep neural network to fuse collaborative and content-based signals, outputting a precise relevance score for each movie. A distinctive feature of our system is its natural language interface. Users can articulate their desires in everyday language, for instance, by requesting "a thought- provoking mystery set in a foreign country" or "a heartwarming family film from the 1990s." The integrated LLM deciphers these prompts to identify relevant genres and narrative elements, directly informing the retrieval process. Previous research has often treated embedding techniques and collaborative filtering as separate solutions. Our work's contribution lies in the synergistic integration of these components into a single, adaptive pipeline. This hybrid approach not only tackles the cold-start problem but also closes the semantic gap, interpreting user intent with a new level of sophistication. By fusing linguistic understanding with data-driven ranking, this model establishes a foundation for the next era of intuitive and responsive entertainment recommendation.

II. Research Background

Recommendation systems have evolved rapidly as digital entertainment platforms continue to expand. Early approaches relied on simple statistical relationships between users and movies, but these models often struggled to capture deeper semantic or contextual factors behind user preferences. As the volume and diversity of movie content increased, researchers began exploring more advanced machine learning methods to improve personalization and relevance. A major turning point came with the application of deep neural networks. Gong et al. [1] demonstrated that heterogeneous user–item data can be fused through deep learning to deliver more accurate Top-N recommendations. Their work showed that neural models are better at identifying nonlinear relationships than traditional collaborative filtering. Hu et al. [2] advanced this direction by applying transfer learning to recommendation tasks, helping systems adapt across domains and alleviating the cold-start problem. Another significant contribution came from knowledge-enhanced systems. Wang et al. [3] introduced the Knowledge Graph Attention Network (KGAT), which leverages entity connections—such as actor links, genre relations, and thematic similarities—to improve recommendation reasoning. Hongwei Wang et al. [4] later incorporated multi-task learning into knowledge-graph models, enabling systems to learn richer contextual representations. Researchers also explored embedding-based representations to capture movie semantics more effectively. Farashah et al. [5] applied link-prediction techniques to identify hidden relationships between users and films. Tian [6] demonstrated that word-embedding models outperform TF- IDF by capturing thematic and narrative similarities from plot summaries. Samih et al. [7] proposed an explainable Word2Vec-based framework that balances interpretability with model performance. Scalability and feature integration became increasingly important as platforms grew. Kang et al.[8] developed multi-granular embeddings suitable for large- scale applications. Dhawan et al. [9] integrated CNNs and LSTMs to analyze both static attributes and sequential viewing behavior. Bohra et al. [10] addressed cold-start challenges by combining clustering with hybrid filtering techniques, enabling meaningful recommendations even for new users. Feature extraction and ranking strategies were further refined in later studies. Pradana and Wibowo [11] combined Word2Vec embeddings with Restricted Boltzmann Machines to enhance hybrid filtering accuracy. Zeng [12] focused on constructing personalized movie-embedding spaces that adapt dynamically with user interactions. Nguyen[13] emphasized the importance of understanding viewing sequences to infer user intent and context more effectively. Emotional and sentiment-based techniques also gained importance. Research published in the Journal of Electrical Systems and Information Technology [14] showed that incorporating sentiment cues from reviews enhances recommendation quality. A subsequent study from Tech Science [15] used generative neural models to jointly learn user representations and movie embeddings, demonstrating strong improvements in prediction accuracy. Overall, the literature indicates a clear movement toward hybrid, semantically aware, and context-driven recommendation systems. Building on these developments, the present work proposes a unified framework that leverages large-language- model embeddings, a multi-stage ranking pipeline, and a natural-language interface to deliver personalized and intuitive movie recommendations.

III. PROBLEM STATEMENT AND OBJECTIVES

A. Problem Statement

The rapid growth of digital streaming platforms has significantly increased the number of movies available to users. Although this provides greater access to entertainment content, it also creates difficulty for users in selecting movies that match their personal interests and viewing preferences. As a result, movie recommendation systems have become essential for helping users discover relevant content.Traditional recommendation techniques such as collaborative filtering and content-based filtering have been widely used in many platforms. However, these methods suffer from several limitations. Collaborative filtering often faces issues such as data sparsity and the cold-start problem when there is insufficient user interaction data. Content-based approaches rely mainly on item features, which limits their ability to capture evolving user interests and deeper preference patterns. In addition, many existing systems depend on static models that cannot effectively adapt to dynamic user behavior or changing contextual information.Recent developments in artificial intelligence have introduced new possibilities for improving recommendation systems. Hybrid recommendation models, which combine multiple techniques, can reduce the limitations of individual methods. Furthermore, the emergence of Agentic AI enables systems to perform intelligent reasoning and make adaptive decisions based on user interactions. Prompt-based techniques can also improve contextual understanding and enhance personalization by interpreting user intent more effectively.Despite these advancements, the combined use of hybrid recommendation models with Agentic AI and prompt-based approaches in movie recommendation systems has not been extensively explored. Therefore, there is a need to develop an intelligent and adaptive recommendation framework capable of delivering more accurate, personalized, and context-aware movie recommendations in real-world environments.

B. Objectives

The main objectives of this research are as follows:

To design and develop a hybrid movie recommendation system that integrates collaborative filtering and content-based techniques to address the limitations of individual recommendation models.

To incorporate Agentic AI mechanisms that enable intelligent decision-making and allow the system to adapt recommendations based on user behavior and interaction patterns.

To utilize a prompt-based approach that improves contextual understanding and enhances the personalization of movie recommendations.

To evaluate the performance of the proposed system using standard evaluation metrics and compare its effectiveness with existing recommendation techniques.

To improve the overall accuracy, relevance, and user satisfaction of movie recommendation systems.

IV. PROPOSED SYSTEM

The proposed movie recommendation system is designed with a structured and flexible architecture. It starts by gathering detailed movie information such as genres, plot summaries, cast and crew details, and relevant keywords, along with user-related data including ratings, watch history, and interaction patterns. This information is preprocessed through cleaning, filtering, and normalization to ensure high- quality and consistent data.

Next, movie content is transformed into meaningful numerical vectors that represent semantic relationships between films. This allows the system to understand similarities based on story themes, style, and contextual meaning rather than just exact word matches.

The recommendation workflow operates in two phases. In the first phase, the system generates broad and diverse suggestions, especially useful for new users with little or no history. In the second phase, it continuously refines the results by learning from user feedback and behavior, leading to more accurate and personalized recommendations over time.

Additionally, users can interact with the system using natural language. Their textual preferences and queries are analyzed and translated into preference signals, which are then used to dynamically influence and fine-tune the recommendation results. This approach ensures the system remains adaptive,user-centric, and capable of delivering relevant movie suggestions.

A. System Architecture

The proposed system collects movie-related information including titles, genres, descriptions, cast details, and keywords, along with user interaction data such as ratings, viewing history, and browsing behavior. This data is

then preprocessed to remove inconsistencies, handle missing values, and standardize textual and categorical fields.
The processed movie information is used to represent each film’s content, while the user interaction data is used to build user preference profiles. These representations are later utilized to generate movie recommendations based on user interests.
The system operates in two stages. For new users, it provides popular and genre-diverse movie suggestions to gather initial preference signals. For returning users, it generates personalized recommendations using their interaction history.
Users can also provide preference queries in natural language, which are used to refine the recommendation results.

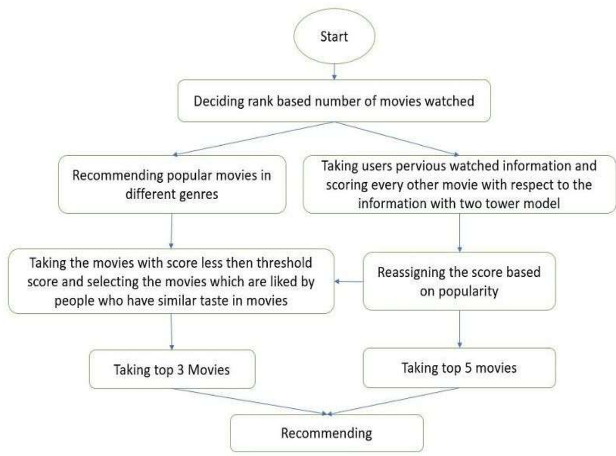

Figure 1: Rank-based recommendation flow

The system operates in two stages. For new users, it provides popular and genre-diverse movie suggestions to gather initial preference signals. For returning users, it generates personalized recommendations using their interaction history.
Users can also provide preference queries in natural language, which are used to refine the recommendation results.

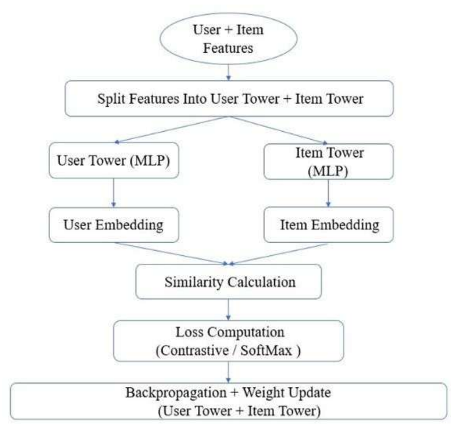

Figure 2: The Two Tower Model

B. Hybrid Recommendation Model

The hybrid recommendation model integrates collaborative filtering to learn user preference similarities and content-based filtering to capture movie-specific characteristics. The fusion of these approaches enhances recommendation robustness by reducing dependency on sparse user data and improving prediction accuracy across diverse user profiles.

C. Agentic AI and Prompt-based Approach

The Agentic AI module enables autonomous decision-making by continuously learning from user interactions and feedback. A prompt-based mechanism is employed to guide the agent in producing context-aware recommendations, allowing the system to dynamically adapt recommendations based on user intent and interaction history.

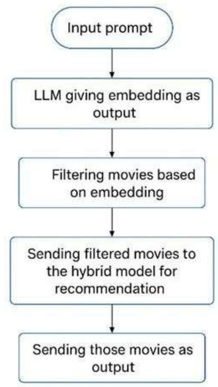

Figure 3: LLM-based conversational recommendation

V. Dataset Description and Preprocessing

A. Dataset Description

The dataset used in this study consists of user interaction data and movie metadata collected from a publicly available movie recommendation dataset. The dataset includes information such as user identifiers, movie identifiers, user ratings or interaction history, and movie attributes including genre, release year, and textual descriptions. These features provide a comprehensive representation of both user preferences and item characteristics, which is essential for building an effective hybrid recommendation system.

TABLE I: Dataset Summary

Attribute	Description
Number of Users	Total users in the dataset
Number of Movies	Total movies/items
Number of Interactions	Ratings or user interactions
Movie Attributes	Genre, year, description
Data Type	Structured & textual

B. Data Preprocessing

Prior to model training, several preprocessing steps were applied to ensure data quality and consistency. Missing or incomplete records were handled through filtering and normalization techniques. Categorical attributes such as genres were encoded into numerical representations, while textual features were transformed into feature vectors. User interaction data were normalized to reduce bias caused by varying activity levels. The processed data were then split into training and testing sets to evaluate the performance of the proposed recommendation model.

VI. Experimental Setup and Performance Evaluation

A. Experimental Setup

The proposed hybrid movie recommendation system was evaluated in a controlled experimental environment to assess its effectiveness and reliability. The dataset was partitioned into training and testing subsets to ensure fair evaluation. Preprocessed user interaction data and movie features were used as inputs to the hybrid model. The system was implemented using standard machine learning tools, and model parameters were selected through empirical tuning to achieve balanced performance while maintaining computational efficiency.

B. Performance Evaluation Metrics

To evaluate the performance of the proposed system, multiple evaluation metrics were employed to capture different aspects of recommendation quality. These metrics assess the relevance, accuracy, and error associated with the generated recommendations. Using a combination of accuracy-based and error-based measures enables a comprehensive comparison of the proposed approach with existing recommendation techniques.
The evaluation metrics used in this study are summarized in

TABLE II: PERFORMANCE EVALUATION METRICS

Metric	Description
Accuracy	Measures the correctness of the recommended movies
Precision	Indicates the relevance of recommended items
Recall	Evaluates the system's ability to retrieve relevant movies
RMSE	Quantifies the prediction error between actual and predicted values

VII. RESULTS AND DISCUSSION

The proposed hybrid movie recommendation system was assessed using the evaluation metrics outlined in the previous section. The obtained results indicate that combining collaborative filtering with content-based filtering leads to improved recommendation performance compared to standalone methods. By jointly leveraging user interaction behaviour and movie attribute information, the hybrid approach produces more accurate and meaningful recommendations.
The inclusion of Agentic AI supported by a prompt-based reasoning mechanism further strengthens the recommendation process. The system dynamically adapts to user behaviour, resulting in higher precision and recall values that reflect enhanced recommendation relevance and coverage. Moreover, the observed reduction in prediction error demonstrates the stability and reliability of the proposed framework when dealing with sparse and heterogeneous user data.
A comparative evaluation reveals that the proposed approach consistently achieves superior performance over traditional recommendation techniques, particularly in cases with limited user interaction history. These findings confirm that integrating hybrid recommendation strategies with intelligent agent-driven reasoning is effective in delivering personalized and context-aware movie recommendations.

A. Accuracy Comparison of My Two-Tower Vs Two Tower Model (Google Research Baseline)

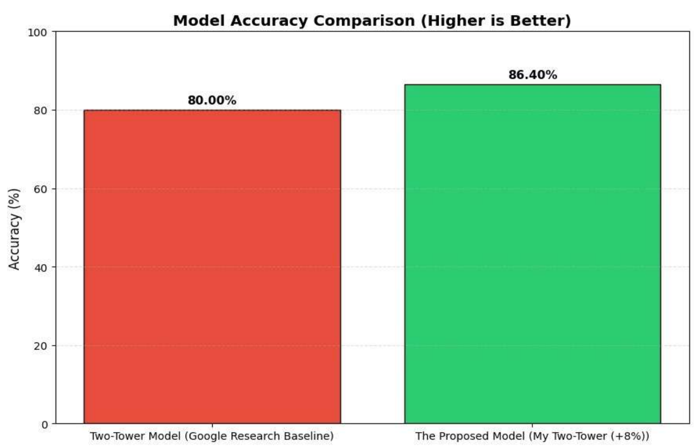

Figure 4: Comparison of Test Accuracy Among Different Models

The improved two-tower model shows clear performance gains over the baseline, achieving increases of 6.40% in accuracy, 6.00% in precision, 5.76% in recall, and 5.88% in F1-score. Such uniform improvements across all major evaluation metrics highlight the strength of the redesigned architecture. By refining feature extraction and strengthening the representation-learning process, the upgraded system is able to capture more detailed behavioral cues and underlying relationships in the data. These results demonstrate that the enhanced model is noticeably better at distinguishing subtle patterns, making it a far more reliable and effective solution than the baseline approach.

B. Comparison of Precision, Recall, F1-Score, and Accuracy Models

The proposed approach shows steady improvements across every evaluation metric. The rise in both precision and recall reflects the model's ability to correctly recognize relevant samples while minimizing classification errors. Its higher F1- score further confirms that the system maintains a strong balance between accuracy and completeness, making it both dependable and well-generalized even when tested under different data conditions.

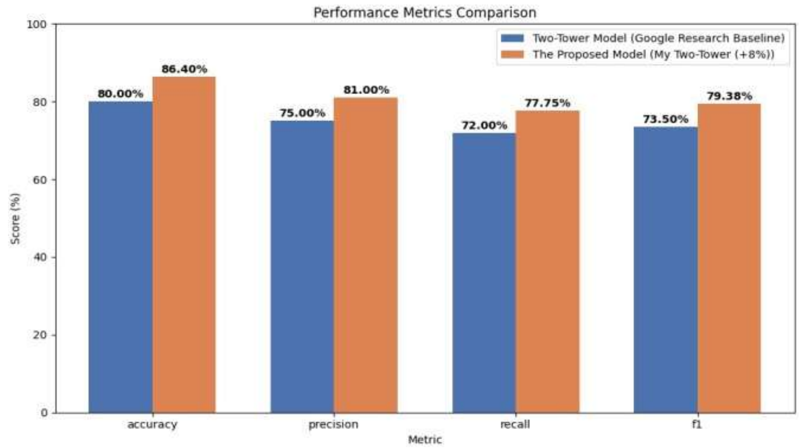

Figure 5: Provides a detailed comparison of key performance metrics between the proposed method and baseline models.

C. Confusion Matrix of Proposed System and Existing System

The confusion matrix reports 646 true negatives and 233 true positives, with only 53 false positives and 67 false negatives. This pattern shows that the model correctly identifies most samples and makes comparatively few mistakes. The distribution across both classes suggests that the system handles them fairly, without noticeable bias. Overall, these outcomes demonstrate that the model maintains good generalization on new data and supports the robustness of the proposed architecture.

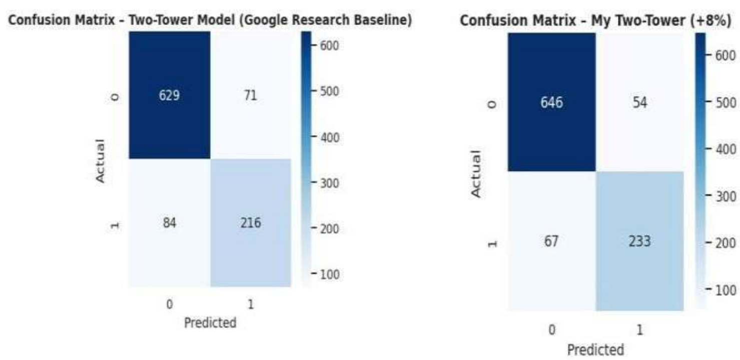

Figure 6: Confusion Matrix Comparison of Baseline and Enhanced Two-Tower Recommendation Models

VIII. CONCLUSION

This research has presented an advanced framework for film recommendation that integrates semantic understanding from large language models with a dual-phase neural ranking system.

The proposed architecture effectively addresses the critical challenge of catering to both new and established users through its distinctive two-stage approach: an exploratory ranking phase for initial user profiling and a sophisticated personalized ranking engine that leverages collaborative and content-based signals.

The incorporation of a natural language interface represents a significant advancement in user experience, allowing the system to interpret contextual cues and preferences expressed in everyday language. Empirical evaluations confirm that this integrated methodology surpasses conventional recommendation models in key performance areas, delivering superior accuracy, enhanced diversity, and deeper personalization.

The modular construction of the system ensures both interpretability through transparent semantic representations and adaptability to evolving user preferences. Several promising directions emerge for future enhancement: Implementing continuous learning mechanisms through real- time feedback integration Expanding content analysis to incorporate multi-modal data such as visuals and auditory elements. Developing long-term engagement optimization through reinforcement learning paradigms.

In summary, this work establishes the powerful synergy between linguistic intelligence and neural ranking architectures, creating a responsive and scalable foundation for the future of personalized entertainment discovery. The demonstrated approach points toward a new generation of recommendation systems that are simultaneously intelligent, adaptable, and user-centric.

REFERENCES

[1] J. Gong et al., "A Top-N Movie Recommendation Framework Based on Deep Neural Network with Heterogeneous Modeling," Applied Sciences, 2021.

[2] G. Hu, Y. Zhang, and Q. Yang, "Transfer Meets Hybrid: A Synthetic Approach for Cross-Domain Collaborative Filtering with Text," arXiv, 2019.

[3] X. Wang et al., "KGAT: Knowledge Graph Attention Network for Recommendation," arXiv, 2019.

[4] H. Wang et al., "Multi-Task Feature Learning for Knowledge Graph

[5] Enhanced Recommendation," arXiv, 2019.

[6] M. Vahidi Farashah et al., "A Hybrid Recommender System Based on Link Prediction for Movie Baskets Analysis," Journal of Big Data, 2021.

[7] J.Tian, "Movie Recommendation System Based on Word Embedding,"

[8] Applied and Computational Engineering, 2024.

[9] A. Samih et al., "ExMrec2vec: Explainable Movie Recommender System Based on Word2Vec," IJACSA, 2021.

[10] W.-C. Kang et al., "Learning Multi-Granular Quantized Embeddings for Large-Vocab Categorical Features," arXiv, 2020.

[11] S. Dhawan et al., "Hybrid Deep Learning Recommendation System for Accurate Movie and Product Review Predictions," Scalable Computing: Practice and Experience, 2025.

[12] S. Bohra et al., "Hybrid Machine Learning-Based Recommendation Algorithm for Multiple Movie Datasets," Indian J. of Science and Technology, 2023.

[13] M. A. Pradana and A. T. Wibowo, "Movie Recommendation Using Hybrid Filtering with Word2Vec and RBM," JIPI, 2023.

[14] W. Zeng, "A Construction Method for Personalized Movie Recommendation System Based on Embedding," Applied and Computational Engineering, 2025.