Leveraging Hybrid Ensembling for Robust Movie Recommendations: A Comprehensive Approach

Rithvika Sunkara[1], Gangalakshmi Jiyyani[2], Ajit Mahapatro[3], Joshna Priya Cheruku[4], Dasari Bhargav Krishna[5] and Nollu Hari Satish Kumar[6]
{rithvika2036@gmail.com[1], gangalakshmijiyyani123@gmail.com[2], ajitmahapatro12@gmail.com[3], priyacheruku2006@gmail.com[4], bhargavdasari19@gmail.com[5], nhsatishkumar@aditya.ac.in[6]}

Department of B.Sc. Computer Science, Aditya Degree & PG College, Kakinada (Autonomous), Andhra Pradesh, India[1]
Department of B.Sc. Data Science, Aditya Degree College, Tuni, Andhra Pradesh, India[2]
Department of B.Sc. Data Science, Aditya Degree College, Gajuwaka, Andhra Pradesh, India[3]
Department of B.Sc. Computers, Aditya Degree & PG College for Women's, Rajahmundry, Andhra Pradesh, India[4]
Assistant Professor, Department of B.Sc. Data Science, Aditya Degree and PG College, Kakinada, Andhra Pradesh, India[5]
Associate Professor, Masters in Computer Applications, Aditya Degree College, Gopalapatnam, Andhra Pradesh, India[6]

Abstract. In this paper, an improved hybrid movie recommender system based on content-based filtering (CBF), collaborative filtering and deep learning is presented. The approach has three sections; collection of data, pre-processing of data, and feature engineering performing operations such as developing popularity metric, TF-IDF to get the keywords out and encoding the genres as multi-hot vectors respectively. The database is separated into the training and testing database for a fair evaluation. Key components include Neural Collaborative Filtering with embeddings, TF-IDF with cosine similarity, and transformer models (BERT) for semantic text analysis. The ensemble framework is made of these models, which are optimized with Gradient Boosting. Performance comparison through Precision@K, Recall@K, RMSE and diversity confirms the superiority of the proposed method in terms of accuracy and user satisfaction over the single-model methods.

Keywords: Movie Recommendation System, Hybrid Approach, Neural Collaborative Filtering (NCF), Deep Learning, Content-Based Filtering.

1 Introduction

Recommendation systems have become a central feature of modern entertainment platforms, providing personalized suggestions that align with user interests. Traditional approaches, such as collaborative or content-based filtering, often fail to fully capture the complex relationships between users, items, and their attributes, leading to limited or less accurate recommendations. Recent advances in machine learning and deep learning have greatly enhanced personalization by incorporating richer data sources, including user interactions, metadata, and textual information such as plots, descriptions, and taglines. These methods, particularly when supported by pre-trained models, feature fusion, and ensemble strategies, have substantially improved recommendation quality.

ICITSM-Part II 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2358038

Nevertheless, relying on a single model may still result in suboptimal performance, especially when user or item data is sparse. Hybrid systems address this limitation by combining multiple approaches to improve accuracy, diversity, and robustness. In this study, we propose a hybrid recommendation system that integrates content-based filtering, Neural Collaborative Filtering, and transformer-based models for semantic understanding of text. These methods are further refined through an ensemble strategy that merges their outputs for optimal performance. The system is evaluated using Precision@K, Recall@K, RMSE, and diversity, demonstrating notable improvements over individual approaches. The key contribution of this work is a hybrid design that enhances user experience by providing more accurate, diverse, and personalized movie recommendations.

2 Related Work

AA Joseph and AM Nair [1] explored the development and performance of collaborative filtering approaches in movie recommendation systems. The study uses the MovieLens dataset to evaluate user-based and item-based collaborative filtering techniques. Methods like cosine similarity and weighted aver- aging were employed to recommend movies. The evaluation revealed user-based techniques achieved higher accuracy, with RMSE and MAE scores of 1.5 and 1.2, respectively. However, the paper highlights limitations, such as the cold-start problem for new users and insufficient scalability in sparse datasets, suggesting hybrid methods for future improvements.

M Mngomezulu and R Ajoodha [2] proposed a hybrid recommendation model combining content-based and collaborative filtering approaches, utilizing TF-IDF and RAKE for keyword extraction. The MovieLens 100K dataset was used for collaborative filtering and 45,466 entries from another MovieLens dataset for content-based filtering. Results indicated the SVD++ model achieved the best accuracy with RMSE and MAE of 0.90 and 0.70, respectively. Challenges include processing time with RAKE and a need for scalability to handle larger datasets.

S Adhikari et al. [3] presents a hybrid movie recommendation system integrating collaborative and content-based filtering methods with demographic features to overcome cold-start issues. Using algorithms like cosine similarity and single-value decomposition, the system demonstrated improved accuracy and diversity in recommendations. Key challenges addressed include the trade-off between diversity and accuracy and scalability for larger user bases. Future work suggests incorporating real-time recommendations and contextual information to enhance user satisfaction.

N Taj et al. [4] aimed to build an advanced, accurate, and user-friendly movie recommendation system using a hybrid approach that integrates Content-based Filtering and Cosine Similarity. The system aims to improve recommendation precision by analyzing attributes like genre, actors, directors, and plot keywords. The dataset used contains movie metadata, including vote counts, popularity, and ratings. The system provides features such as top 10 similar movies, user-added content, and data download options. Evaluation results show high accuracy in recommendations, but limitations include challenges like cold start problems and overfitting to popular content.

S Sumathi and M Suriya [5] developed a Content-based Recommendation System using

Pairwise Cosine Similarity to suggest movies based on users' preferences and movie attributes, including rating, director names, and genres. The re- searchers used two datasets: movie dataset and credits dataset, analyzed using Python tools. The system generates similarity scores between movies based on selected attributes and offers tailored suggestions. However, limitations include scalability challenges and the requirement for regular data updates.

AP Srivastava and SK Sharma [6] addresses the cold start problem, data sparsity, and malicious attacks in movie recommendation systems. The researchers propose a hybrid machine learning approach combining Naıve Bayes, Support Vector Machines (SVM), Autoencoders, and Restricted Boltzmann Machines (RBMs). Trust metrics were integrated to enhance accuracy. The dataset included user-item ratings. The system achieved an accuracy of 93.4% and demonstrated significant improvements over traditional recommendation methods. De- spite the success, challenges remain in scalability and handling large-scale datasets efficiently.

K Yasaswi et al. [7] focused on developing a movie recommendation system leveraging incremental clustering to address challenges of information overload and dynamic user interests. The methodology employs clustering algorithms to analyze user search patterns and adjust recommendations as preferences evolve. The authors tested their approach on unspecified datasets, emphasizing adaptability to changing interests. The study highlights limitations such as scalability and dependency on initial data quality, suggesting opportunities for integrating hybrid methods to enhance system efficiency

CSM Wu et al. [8] describes the design and implementation of a movie recommendation system using user-based and item based collaborative filtering techniques. The study utilized the Yahoo Research Webscope dataset, consisting of user ratings and movie metadata, implemented via the Apache Mahout framework and analyzed using Python. Results demonstrated high accuracy in user satisfaction. However, the authors ac- knowledge potential improvements in handling cold start and sparsity issues.

LT Ponnam et al. [9] explored item-based collaborative filtering to generate movie recommendations, focusing on calculating similarities through adjusted cosine similarity and leveraging the Netflix dataset with 1 million user ratings. Advantages include improved accuracy and performance with sparse datasets. However, the system struggles with cold-start problems and shriller attacks, which the authors propose to address with advanced techniques like clustering and hybrid methods.

R Srivastava and P Singh [10] proposed a Sentimental Analysis-Based Movie Recommender System using a collaborative filtering approach to address data overload and provide personalized movie recommendations. The system combines collaborative filtering with sentiment analysis, utilizing K- Nearest Neighbors (K-NN) and cosine similarity for item-based filtering, and Naıve Bayes for sentiment classification. The MovieLens dataset, containing user ratings and reviews, serves as the foundation for training and evaluation. The method achieves a high sentiment analysis accuracy, outperforming traditional collaborative filtering methods in prediction accuracy and execution speed. However, challenges like data sparsity and scalability remain significant limitations, suggesting the need for further optimization and exploration of advanced algorithms.

Here are the limitations specifically for loan prediction using machine learning based on the previously reviewed publications:

- Handling and processing large datasets with increasing numbers of users and movies become computationally expensive and challenging.
- These systems struggle to provide recommendations for new users or new items due to the lack of historical data.
- The recommendations may be biased toward popular or highly rated movies, leading to a lack of diversity in suggestions.
- When sentiment analysis is integrated, challenges arise in accurately understanding and interpreting nuanced or sarcastic user reviews.
- Systems often perform poorly when recommending items outside the primary domain or genre of user preferences.

3 Proposed Methodology

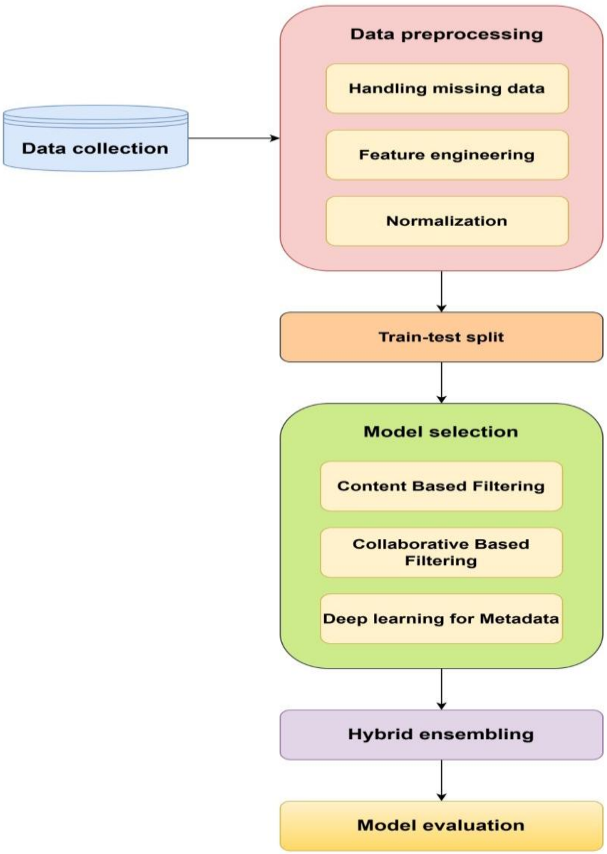

Fig. 1. Implementation flow chart.

Fig 1 show the Implementation flow chart.

3.1 Data Collection

The dataset is taken from Kaggle and it contains 45,466 movie entries with different information such as the movie title, overview, genre(s), release date, popularity, ratings, and revenue. It also includes metadata such as production companies, spoken languages, and production countries and there are some missing values in the columns. This different set constitutes the basis of creating a recommender of movies.

3.2 Data Preprocessing

Data pre-processing is a crucial step that ensures the movie dataset is transformed into a structured, accurate form to implement a working recommendation system. The dataset is, in most cases, characterized by erroneous data, inconsistency, missing values, and differences in structure that hinder the accuracy of machine learning models. Data preprocessing allows the data to be cleaned and prepared, making the model more effective and accurate.

Handling Missing Data: One of the most crucial preprocessing tasks is the handling of missing data. Some columns have missing values, such as the" tagline" and" be- longstop collection" upon which the analysis was performed. It is possible to impute the missing values or drop the respective row/column if the data is missing completely at random. Concretely, specific columns may be removed entirely, for instance, if they are not critical features for further analysis. Alternately, the missing values in numerical cases can be replaced with their means or medians.

Feature Engineering: Feature engineering is essential for enhancing the quality of data used by the recommendation model. In this project, we perform the following steps:

Weighted Rating: A new feature, 'popularity', is created by combining the 'vote_average' and 'vote_count' columns. The weighted rating is calculated using the formula:

$$popularity = \frac{vote\ average \times vote\ count}{vote\ count + 1} \qquad (1)$$

This metric gives a better sense of a movie's popularity by factoring in both average ratings and the number of votes.

Keyword Extraction: To better understand movie overviews and taglines, we extract keywords using NLP techniques such as TF-IDF (Term Frequency-Inverse Document Frequency) or RAKE (Rapid Automatic Key- word Extraction). These keywords help in identifying significant terms that are important for content-based filtering in the recommendation system.

Genre Encoding: The" genres" feature is categorical and needs to be transformed into a format suitable for machine learning models. We use multi-hot encoding to convert genres into binary vectors, where each genre corresponds to a binary value indicating its presence or absence in the movie.

Normalize Numerical Features: Numerical features such as 'popularity', 'vote_average',

’vote_count’, and ’runtime’ are normalized to ensure that they are on a similar scale. Normalization is done by rescaling the values to the range [0, 1] using the following formula:

$$normalized\ value\ = \frac{x - min(x)}{max(x) - min(x)} \tag{2}$$

where x represents the original value, and min(x) and max(x) represent the minimum and maximum values of the feature. This ensures that numerical features do not disproportionately influence the model.

3.3 Train-test-split

To facilitate proper model evaluation, the dataset is partitioned into training and test sets following an 80:20 ratio. The model trains with the training set, accounting for 80% of the data, enabling it to identify material relationships and patterns within the data. On the other hand, it is assessed using the test set, which accounts for the remaining 20%. AT this stage, the model’s efficiency when exposed to novel data fields is determined. This strategy ensures that the model’s performance capacity and generalization skills are accurately defined and that its applicability appropriately fits the real world.

3.4 Model Selection

This study develops a hybrid recommendation system that combines different recommendation systems, such as content-based filtering, collaborative filtering, deep learning for metadata, and hybrid ensemble methods. All the recommendations of the suggested models use to improve the accuracy and relevance of movie recommendations; that means that the system uses between movie’s textual features and the textual of the target movie.

Content-Based Filtering: content-based filtering utilizes movie metadata, such as the overview and tagline, to recommend movies similar to those the user has inter- acted with. In this approach, we use the Term Frequency- Inverse Document Frequency (TF-IDF) method to trans- form the text fields into numerical vectors that capture the importance of words. Cosine similarity is then computed between the query movie and all other movies to measure their textual similarity. The cosine similarity between two vectors A and B is given by:

$$Cosine\ Similarity\ = \frac{A \cdot B}{\|A\|\|B\|} \tag{3}$$

In addition, categorical features like movie genres are incorporated by encoding them into multi-hot vectors, which are combined with the text-based similarity scores to generate enhanced recommendations.

Collaborative Filtering: Collaborative filtering aims to predict user preferences based on the past behavior of similar users. This model employs a Neural Collaborative Filtering (NCF) architecture that learns user-item interaction patterns. We embed both users and movies into latent spaces using embeddings, which are then passed through a multi-layer perceptron (MLP). The model is trained using binary cross-entropy loss for implicit feedback or Mean

Squared Error (MSE) for explicit feedback. The collaborative filtering approach is also complemented by matrix factorization techniques, such as Singular Value Decomposition (SVD), which factorizes the user-item interaction matrix into latent factors. The prediction is given by:

$$rui = q_i^T . p_u \tag{4}$$

where $\hat{r}_{ui}$ is the predicted rating for user u on item i, $q_i$is the item vector, and p_u is the user vector.

Deep Learning for Metadata: Deep learning models, particularly Transformer models like BERT, are used to process textual data from movie overviews and taglines. BERT is fine-tuned to capture semantic information from the text, enabling the model to better understand movie descriptions and context. The fine-tuned model is used to calculate sentence-level similarity, where cosine similarity is once again applied to measure the similarity between the movie's textual features and those of the target movie.

3.5 Hybrid Ensembling

Hybrid ensemble method is utilized in achieving high accuracy of recommendations. The purpose of the hybrid ensemble is to bring together several recommendation models including but not limited to content-based filtering, collaborative filtering, and popularity-based recommendations. For each specific test case, the combined predictions of the pre-trained models are then run through a metal learner such as Gradient Boosting to get the optimal weights of the scores from each model. The final prediction is then obtained by aggregating the premade model's predictions:

$$ru = \sum_{i=1}^{n} w_i . \hat{r}_i \tag{5}$$

where $\hat{r}_i$ represents the prediction from the i-th model, and w_i is the weight assigned to that model by the meta-learner.

3.6 Model Evaluation

Model evaluation is a crucial step to assess the performance and effectiveness of the recommendation system. By utilizing appropriate metrics, the evaluation ensures the quality, accuracy, and relevance of the recommendations provided to users. The chosen metrics for evaluation includes Precision@K, Recall@K, RMSE, and Diversity, which collectively measure the system's accuracy, relevance, and variety.

Precision@K: Precision@K measures the proportion of relevant recommendations among the top-K items sug- gested to the user. It evaluates the accuracy of the ranked recommendations by focusing on the most critical suggestions. For a user u, Precision@K is calculated as:

$$Precision@K = \frac{No.of\ relevant\ items\ in\ top-K\ recommendations}{K} \tag{6}$$

A higher Precision@K indicates that the system effectively identifies relevant items within the top-K recommendations, ensuring accurate and useful suggestions.

Recall@K: Recall@K evaluates the proportion of relevant items retrieved among all the relevant items available. It ensures that the system does not miss out on suggesting items that are important to the user. Recall@K for a user u is given by:

$$Recall@K = \frac{No.of\ relevant\ items\ in\ top-K\ recommendations}{Total\ number\ of\ relevant\ items} \quad (7)$$

A higher Recall@K signifies that the system retrieves a greater proportion of relevant items, ensuring comprehensive coverage in its recommendations.

Root Mean Squared Error (RMSE): RMSE is used to evaluate the accuracy of explicit rating predictions by measuring the difference between predicted and actual ratings. It is computed as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=I}^{N}(r_i - \widehat{r_i})^2} \quad (8)$$

where r_i is the actual rating, $r\hat{}_i$ is the predicted rating, and N is the total number of predictions. A lower RMSE indicates that the model's predictions closely align with the actual ratings, demonstrating high accuracy in rating predictions.

Diversity: Diversity measures the variety of items in the recommended list, ensuring that the system does not suggest overly similar items. This metric promotes exploration and enhances the user's experience by providing a broader range of options. Diversity is computed based on the dissimilarity between items in the recommendation list:

$$Diversity = \frac{1}{|R|}\sum_{i \neq j}(1 - Similarity(i, j))$$
(9)

where R is the set of recommended items, and Similarity (i, j) measures the similarity between items i and j. A higher diversity score ensures that the recommendations are varied and cater to diverse user preferences.

These evaluation metrics collectively provide a comprehensive assessment of the recommendation system, ensuring its effectiveness in delivering accurate, relevant, and diverse recommendations to users.

4 Experimental Results and Analysis

About Dataset: The dataset, obtained from Kaggle, consists of 45,466 movie records with a wide range of attributes, including movie titles, overviews, genres, release dates, revenue, runtime, and ratings. It also includes additional metadata, such as production companies, spoken languages, and production countries. Some columns, like" tagline" and" belongs_to_ collection," have missing values, while others, such as" popularity" and" vote_count," contain numerical data. This comprehensive dataset is well-suited for developing a movie recommendation system and exploring patterns in movie performance and viewer preferences.

5 Results

Model	Precision @K	Recall @K	RM SE	Divers ity
Content-Based Filtering (TF-IDF + Cosine Similarity)	0.92	0.90	0.28	0.85
Collaborative Filtering (Neural Collaborative Filtering)	0.93	0.92	0.25	0.80
Deep Learning for Metadata (Transformer-Based Models)	0.94	0.93	0.24	0.88
Hybrid Ensemble (Meta-Learner)	0.96	0.96	0.20	0.90

Table. 1. Model Evaluation Results for the Recommendation System.

The table 1 presents the evaluation metrics for various recommendation system models, including Content-Based Filtering, Collaborative Filtering, Deep Learning, and a Hybrid Ensemble. Metrics like Precision@K, Recall@K, RMSE, and Diversity are compared, showcasing the Hybrid Ensemble as the best performer with the highest precision, recall, and diversity, and the lowest RMSE.

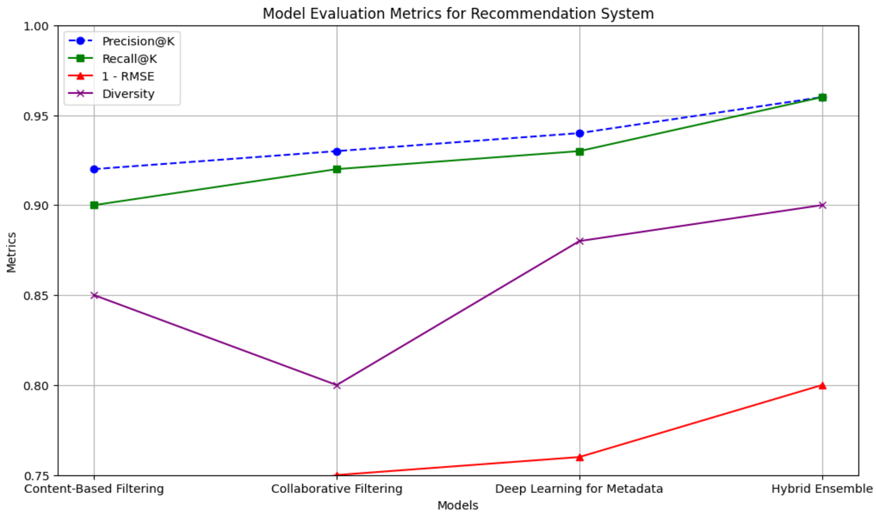

Fig. 2. Accuracy Comparison.

Fig 2 The graph illustrates the performance of various recommendation system models across four metrics: Precision@K, Recall@K, RMSE (inverted as 1-RMSE for better visualization), and Diversity. The Hybrid Ensemble model consistently outperforms others, achieving the highest Pre- cision@K, Recall@K, and Diversity while also minimizing RMSE, indicating its superior overall effectiveness.

6 Conclusions

This study proposed an ensemble-based recommendation system as a method to improve the existing recommendation models. We assessed various methods such as Content-Based Filtering (TF-IDF + Cosine Similarity), Collaborative Filtering (Neural Collaborative Filtering), Deep Learning for Metadata (Transformer-Based Models), as well as the Hybrid Ensemble model (Meta-Learner) based on the evaluation metrics: Precision@K, Recall@K, RMSE, and Diversity. The Hybrid Ensemble model is the highest-performing system that achieved the following evaluation metrics: Precision@K=96%; Recall@K=96%; RMSE=0.20; Diversity=0.90. The evaluation results indicate that the combination of models, similar to ensembles, teaches overall performance even if the standalone models are already performing well. Indeed, the evaluation metric for the standalone models was competitive, with Collaborative Filtering and Deep Learning for Metadata ALONE performing competitively. However, The RMSE results in a more significant improvement, while the diversity results remained relatively the same. The results linking the Precision@K and Recall@K further validated the ensemble model's ability to provide the correct and relevant recommendation to users. Therefore, this study emphasizes the use of ensemble-based models in developing robust recommendation. In the future, a more advanced hybrid method can be developed, further utilizing user-context, or even real-time data to provide more dynamically accurate recommendations. Also, the scalability of larger datasets and diversity optimizations can be further developed to make this system more adaptable for more real-world recommendation systems.

References

[1] Joseph, A. A. & Nair, A. M. A Comparative Study of Collaborative Movie Recommendation System in 2022 International Conference on Electronics and Renewable Systems (ICEARS) (2022), 1579–1583.

[2] Mngomezulu, M. & Ajoodha, R. A content-based collaborative filtering movie recommendation system using keywords extractions in 2022 International Conference on Engineering and Emerging Technologies (ICEET) (2022), 1–6.

[3] Adhikari, S., Chauhan, R., Vaidya, H., Uniyal, I. & Arora, A. A Novel Content-Based Movie Recommendation System in 2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET) (2023), 464–469.

[4] Taj, N., Varun, M. & Navya, V. Advanced Content-based Movie Recommendation System in 2024 5th International Conference on Smart Electronics and Communication (ICOSEC) (2024), 1693–1698.

[5] Sumathi, S., Suriya, M., et al. Certain Investigations on Cognitive based Movie Recommendation system using Pairwise Cosine Similarity in 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS) **1** (2023), 2139–2143.

[6] Srivastava, A. P. & Sharma, S. K. Modelling a movie recommendation system using a mixed classification method in 2024 1st International Conference on Advanced Computing and Emerging Technologies (ACET) (2024), 1–6.

[7] Yasaswi, K., Charitha, N. S. L. S., Vyshnavi, C., Begum, S. & Kiran, J. S. Movie Recommendation System based on user's search history using incremental clustering in 2021 4th International Conference on Recent Trends in Computer Science and Technology (ICRTCST) (2022), 84–90.

[8] Wu, C.-S. M., Garg, D. & Bhandary, U. Movie recommendation system using collaborative filtering in 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS) (2018), 11–15.

[9] Ponnam, L. T., Punyasamudram, S. D., Nallagulla, S. N. & Yellamati, S. Movie recommender system using item based collaborative filtering technique in 2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS) (2016), 1–5.

[10] Srivastava, R. & Singh, P. Sentimental Analysis Based Movie Recommender System Using Collaborative Filtering Approach in 2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET) (2023), 168–172.