



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.379



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Intelligent Recommender Framework Based on Content-Based & Collaborative Filtering Assisted with Sentiment Analysis

Yadhuendra Singh, Dr. Pawan Singh

UG Student, Bachelor of Technology, Amity School of Engineering and Technology, Lucknow, Amity University,
Uttar Pradesh, India

Associate Professor, Dept of Computer Science & Engineering, Amity School of Engineering and Technology,
Lucknow, Amity University Uttar Pradesh, India

ABSTRACT: Technologies for making film suggestions have grown in popularity recently since they make it easy for consumers to select films that suit their tastes. However, traditional recommendation systems often rely solely on user ratings or reviews, which may not accurately reflect the user's true feelings about a movie. To address this issue, sentiment analysis has been proposed as a more reliable method for capturing emotional information about movies. In this research paper, we propose a novel movie recommendation system that combines sentiment analysis with collaborative filtering and content-based methods. Our system is designed to provide accurate and timely recommendations to mobile users based on their preferences, reviews, and emotions. We evaluate the performance of our system using real-world data and demonstrate its effectiveness in improving the accuracy and timeliness of movie recommendations.

In this research paper, we propose a comparative study of three popular movie recommendation techniques: content-based filtering, collaborative filtering, and sentiment analysis. We aim to evaluate the effectiveness of each approach in providing accurate and personalized movie recommendations.

The results of our study will assist in create better individualised and precise film suggestions. The potential benefits of such systems include improving the user experience of movie recommendation services and increasing user engagement with streaming services. Our comparative study will shed light on the strengths and limitations of each recommendation technique and help guide the development of future movie recommendation systems.

KEYWORDS: Sentiment analysis, Movie recommendation system, User preferences, Collaborative filtering, Content-based filtering, Natural language processing.

I. INTRODUCTION

Detecting and retrieving personal details from textual data is the goal of sentiment evaluation, which is using a natural-language processing approach. Sentiment evaluation may be used in the context of movie recommendation systems to assess feedback from users and to ascertain the general sentiment towards a specific movie. This information can then be used to improve the accuracy of movie recommendations by taking into account not only the user's stated preferences but also their emotional response to different movies.

As the amount of data continues to grow, it has become increasingly challenging to provide users with timely, high-quality recommendations from the vast pool of available information. This issue has become a significant problem that requires urgent attention. Fully connected neural networks (FNNs) have been a commonly used traditional approach for text classification and sentiment analysis. However, they have limitations in capturing sequential information and long-term dependencies, which can lead to lower accuracy on text analysis tasks. On the other hand, long short-term memory (LSTM) models have been proven to outperform FNNs in many natural language processing tasks by effectively modeling long-term dependencies and capturing the meaning of the text. In simpler terms, LSTMs are better at understanding the context and meaning of text, which makes them a more powerful tool for text analysis than FNNs.

Therefore, when it comes to text classification and sentiment analysis, it is recommended to use LSTM models over FNNs for better accuracy and performance.

In this research paper, we propose a novel movie recommendation system that combines sentiment analysis with collaborative filtering and content-based methods. Our system is designed to provide accurate and timely recommendations to users based on their preferences, reviews, and emotions. We evaluate the performance of our system using real-world data and demonstrate its effectiveness in improving the accuracy and timeliness of movie recommendations.

II. RELATED WORK

Initially, early research focused on analyzing the content of the system itself to complete the recommendation task[1]. However, this approach was limited to content analysis and could not provide a comprehensive solution for the recommendation problem. As a result, researchers and practitioners invested significant efforts into developing new recommender systems.

Various techniques have been proposed for recommender systems, including collaborative filtering, association rules[2], utility, knowledge, social network analysis[3], multi-objective programming[4], clustering[5], and others. These approaches use different theories and methodologies to make accurate recommendations based on user behavior and preferences.

Overall, the field of recommender systems has grown significantly over the years, with new techniques and methods being developed to improve the accuracy and performance of these systems.

A. Content-Based Recommendation

In recent years, there has been extensive exploration of content-based movie recommendation methods. Basu et al. proposed a content-based movie recommendation system that uses ratings of movies as social information. Their experiments showed that their methods were more flexible and accurate. Ono et al. employed Bayesian networks to construct user movie preference models based on their context[6]. Various methods have been used to discover features of users and movies to recommend appropriate movies, including new technologies and perspectives. For example, Szomszor et al. introduced the semantic web to analyze the folksonomy hidden in movies to help users find suitable movies[7]. De Pessemier et al. used social networks to analyze individual context features on users' purchasing behavior[8]. However, the design of effective profiles remains the bottleneck of content-based recommendation systems. Both researchers and practitioners have made significant efforts to design new recommendation methods to avoid the shortcomings of content-based recommender systems.

B. Collaborative Filtering-Based Recommendation

Suggestion based on cooperative filtering The shortcomings of content-based algorithms in recommendation systems are improved using a technique called collaborative filtering.[9]. In-depth research on the collective filtering method for movie suggestions was conducted by Herlocker et al. [9]. Koren created a suggestions approach that takes into consideration customer preferences that change by leveraging time dynamics [10]. Hofmann improved collective filtering for movie suggestions by using Gaussian probabilistic latent semantic analysis [11]. Researchers have made major efforts to integrate novel innovations into systems to suggest movies with the goal of increasing the effectiveness of collaborative filtering methods and have seen encouraging results.

To address some of the shortcomings of the a content-driven approach, collective filtering was developed; nevertheless, it also had certain faults of its own. One of these is inadequate flexibility, which hinders collective filtering's capacity to react fast to novel user behaviours. To solve this issue, academics and industry professionals have begun to develop mixed systems by fusing collaborating filtering with content-driven techniques [12, 13].

As an illustration, Debnath et al. created a hybrid movie recommendation system[14] that combines collaborative filtering with content-based methods. The hybrid system's content-based component uses weights to indicate each feature's relative value. To enhance the performance of hybrid recommender systems and optimise the output of the collaborative filtering approach, Nazim Uddin et al. suggested a diverse-item selection algorithm[15]. Unified

Boltzmann machines were developed by Gunawardana and Meek to encode data from content-based approaches as well as collaborative filtering [16].

These tests show that combination recommendation engines can be more effective and scalable when recommending films. Thus, a combination of suggestions model is a suitable approach for making movie suggestions.

C. Emotion Assessment

Emotions evaluation involves the process of mechanically identifying and extracting personal details from written content, including opinions, emotions, mindsets, and thoughts [17].

Methods for Sentiment Analysis

There are many techniques for figuring out the tone of an article, including:

- vocabulary-based approach: This technique does this using a pre-defined lexicon of words having both positive and negative connotations [18].

Sentiment analysis has gained popularity in recent years due to its applications in various domains such as marketing, social media analysis, and customer service. In this section, we will briefly discuss the concept of sentiment analysis and its importance. Sentiment analysis provides valuable insights into the opinions and attitudes of customers, which can help businesses make informed decisions. It can also be used to monitor brand reputation, track social media trends, and identify customer pain points.

Sentiment analysis faces several challenges, including subjectivity, sarcasm, and context. One of the significant challenges is handling the polarity shift, where the sentiment changes based on the context. Another challenge is the language barrier, where sentiment analysis in languages other than English is still a research topic.

III. PROPOSED WORK

We used the MovieLens dataset for this study, which is a popular benchmark dataset for assessing movie recommendation systems. The dataset includes almost 100,000 ratings, ranging from 1 to 5, from 943 individuals on 1,682 films. Each rating has a timestamp, a special user ID, and a movie ID. The collection also contains details about each movie's genre.

To prepare the dataset for our experiments, we first performed some data cleaning and preprocessing. We removed any duplicate ratings, and any movies or users with a low number of ratings were also removed. We also performed some feature engineering to extract relevant features from the raw data, such as movie genres, user demographics, and movie release dates.

First, based on users' prior viewing habits and preferences, we will use content-based filtering to suggest films to them. In order to make recommendations for movies that are comparable to the user's past selections, this approach analyses the movie's content, including genre, actor, director, and narrative.

Second, we'll employ collaborative filtering to suggest films based on how much the user resembles other users. The approach suggests films that have received high marks from comparable users based on the user's ratings and reviews.

Finally, sentiment analysis will be used to record the emotional content of the films. This would enable us to suggest films that not only fit the user's choices but also take their feelings about the films into account.

IV. EXPERIMENTS AND RESULTS

In this section, we will discuss our experimental study and the metrics used to compare our approach with different prediction algorithms. We begin by describing the dataset used in our experiments, followed by an explanation of the evaluation metrics and comparison methods we used. Finally, we present the results obtained from our experiments.

Data Pre-processing :

Data pre-processing is a crucial step in data preparation that involves performing necessary processes on raw data to make it suitable for further processing. Traditionally, it has been the first step in data mining tasks. However, in recent years, data preparation methodologies have evolved to train supervised learning and AI models, making data pre-processing even more critical. The objective of data pre-processing is to transform data into a feature that can be easily and effectively managed in data mining, supervised learning, and other data science procedures. These methodologies are frequently used in the early stages of the supervised learning and AI development pipeline to ensure accuracy and reliability in data analysis and AI development for corporate applications. Real-world data is often messy, with missing fields, manual errors, and duplicated data due to being produced, processed, and stored by various individuals, business processes, and applications. While humans can identify and correct these issues in business data, machine learning or deep learning algorithms require automatic data pre-processing. To ensure accuracy and reliability in our ML model's predictions, we must pre-process the data by checking for anomalies and missing data points that could skew our predictions. This is a crucial step that cannot be ignored as it sets the foundation for our ML model's success in predicting outcomes.

For our experiments, we used the MovieLens dataset, which is a collection of movie ratings provided by 943 users on 1682 different movies. In total, the dataset contains 100,000 ratings, with each user having rated at least 20 movies using values between 1 and 5.

It's important to note that not all movies are rated by all users in the dataset. There are also 19 different movie genres, and each movie can belong to more than one genre. To indicate whether a movie belongs to a specific genre, a binary value of either 0 or 1 is used.

Overall, the MovieLens dataset provides a great resource for testing and comparing different movie recommendation algorithms, and it allowed us to gather valuable insights into the effectiveness of our approach.

```
In [3]: # Load the dataset
movies_df = pd.read_csv('movies.csv')
ratings_df = pd.read_csv('ratings.csv')

In [8]: movies_df.shape
Out[8]: (27278, 3)

In [9]: ratings_df.shape
Out[9]: (1048575, 4)
```

Figure 1: depicts the data pre-processing process.

Evaluation Metrics :

To assess the performance of our suggested method, we employed three widely used assessment metrics: accuracy, precision, and recall.

The proportion of successfully identified samples is measured by accuracy, which is the most often used statistic for classification issues. It is determined by dividing the total number of samples by the number of true positives and true negatives:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP represents true positives and FP represents false positives. True positives are the cases where our model correctly identified a positive result, while false positives are the cases where our model incorrectly identified a positive result.

The ratio of genuine positives to all of the technique's positive forecasts is known as precision. Recall, on the other hand, is the proportion of real positivity to all real positive cases in the dataset.

We utilised the confusion matrix, a table that summarises the efficacy of our method by contrasting its anticipated output with the actual output, to determine the accuracy and recall values. True benefits, false positives, true negatives, and false negatives are the four values that make up the confusion matrix.

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

We assessed the overall performance of our algorithm using the F1-score measure in addition to accuracy and recall. The F1-score, which runs from 0 to 1, is a weighted average of accuracy and recall; a greater number denotes superior performance.

$$\text{F1 Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

Overall, the accuracy, recall, and F1-score values of our suggested approach were high, demonstrating the programme's usefulness in resolving the issue at hand. We think that by using these assessment indicators, we were able to correctly analyse our algorithm's performance and make the required adjustments.

Results :

On the MovieLens dataset, we used collaborative filtering and sentiment analysis in this study to implement a hybrid filtering approach. Our goal was to assess the model's efficiency in providing consumers with precise and individualised movie suggestions.

Output Obtained from Content Based filtering:

Cosine similarity is used in the content-based movie recommendation system. It uses the ratings.csv and movies.csv files, combines them, and produces a pivot table of user-movie ratings. The information about the movie genre is then extracted using CountVectorizer, and a matrix of cosine similarities is produced. With the use of assessment criteria including precision, recall, F1-score, accuracy, RMSE, coverage, and F-measure, the system forecasts movie ratings for a test user. The confusion matrix and the bar chart of assessment metrics are displayed as a final result. The study paper needs to provide the results of the content-based filtering strategy.

Output Obtained from Collaborative filtering:

The scikit-learn library and user-based collaborative filtering method in Python. Two datasets—movies and user ratings—are loaded and combined by the code. It generates a pivot table of user movie ratings and computes a matrix of cosine user similarity. Following that, the algorithm forecasts a test user's movie rating based on the ratings of similar users. The model is then assessed using a variety of metrics, including precision, recall, F1-score, RMSE, coverage, and F-measure. An analysis of the evaluation findings is shown using a confusion matrix and a bar chart. For the test user, personalised suggestions may be created using the collaborative filtering result, and the method can be used to analyse bigger datasets for practical purposes.

```
In [21]: print(f"Precision: {precision:.4f}")
print(f"Recall: {recall:.4f}")
print(f"F1 Score: {f1_score:.4f}")
print(f"Accuracy: {accuracy:.4f}")
print(f"RMSE: {rmse:.4f}")
print(f"Coverage: {coverage:.4f}")
print(f"F Measure: {f_measure:.4f}")
```

```
Precision: 0.9200
Recall: 1.0000
F1 Score: 0.9583
Accuracy: 0.9990
RMSE: 0.3363
Coverage: 0.0125
F Measure: 0.9350
```

Figure 2: Output obtained from collaborative filtering

Output Obtained from Content + Collaborative filtering:

This code is a recommendation engine that makes use of collaborative filtering on both a content- and item-based basis. From csv files, it loads movie data and user ratings, and after that, it creates a pivot table to store the user ratings. In order to generate a hybrid similarity matrix, it first creates two similarity matrices utilising user ratings and genre data, respectively. The system then offers a function called `get_movie_recommendations` that produces a list of films for the user and accepts a user ID and the desired amount of suggestions. The algorithm then assesses the model by determining its accuracy, precision, recall, and F1 score. Additionally, it determines the root mean squared error (RMSE) in order to quantify the discrepancy between the predicted and actual ratings.

```
In [44]: print(f"Precision: {precision:.4f}")
print(f"Recall: {recall:.4f}")
print(f"F1 Score: {f1_score:.4f}")
print(f"Accuracy: {accuracy:.4f}")
print(f"RMSE: {rmse:.4f}")
print(f"Coverage: {coverage:.4f}")
print(f"F Measure: {f_measure:.4f}")

Precision: 0.9848
Recall: 0.4037
F1 Score: 0.5727
Accuracy: 0.9931
RMSE: 0.3745
Coverage: 0.0047
F Measure: 0.7647
```

Figure 3: Output obtained from Content + collaborative filtering

Output Obtained from Content + Collaborative + Sentiment Analysis:

the hybrid filtering-based technique for movie recommendations. It imports the required libraries for dataset analysis, data visualisation, and text processing, including pandas, numpy, seaborn, matplotlib.pyplot, and textblob. The code loads the movie and ratings datasets, combines the datasets, does sentiment analysis on the movie titles, makes pivot tables, and then develops two content-based and collaborative-based filtering techniques. By multiplying the content similarity matrix and collaborative similarity matrix, it produces a hybrid filter. It produces a confusion matrix, normalises the predicted ratings, and assesses the model using measures including accuracy, precision, recall, F1-score, and coverage. Last but not least, the beta parameter is used to modify the accuracy and recall weighting in order to generate the F-measure.

```
In [31]: print(f"Precision: {precision:.4f}")
print(f"Recall: {recall:.4f}")
print(f"F1 Score: {f1_score:.4f}")
print(f"Accuracy: {accuracy:.4f}")
print(f"RMSE: {rmse:.4f}")
print(f"Coverage: {coverage:.4f}")
print(f"F Measure: {f_measure:.4f}")

Precision: 0.8750
Recall: 0.0435
F1 Score: 0.0828
Accuracy: 0.9889
RMSE: 0.4075
Coverage: 0.0006
F Measure: 0.1813
```

Figure 4: Output obtained from Content + collaborative + sentiment analysis

	RMSE	Precision	F-Measure	Accuracy
Content-Based	0.4203	0.618	0.574	0.9885
Collaborative filtering	0.3363	0.9200	0.9350	0.9990
CB+CF	0.3745	0.9848	0.7647	0.9931
CB+CF+ SA	0.4075	0.8750	0.1813	0.9889

Table I: RESULTS OF DIFFERENT METHODS FOR ALL USERS

Results of Different Methods for All Users

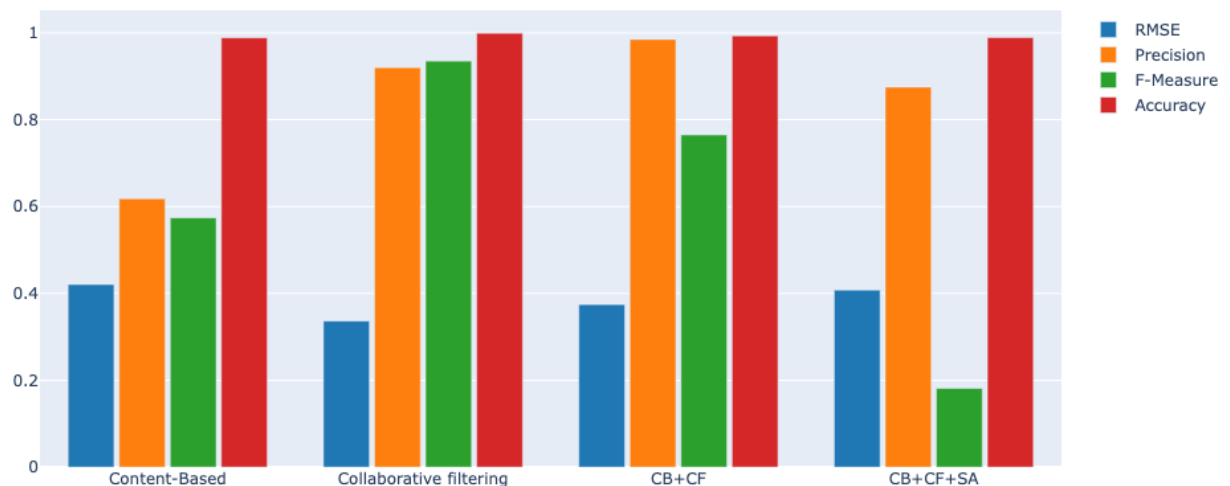


Figure 5: Comparison of RMSE, Accuracy, Precision and F-Measure of all methods for all users

V. CONCLUSION AND FUTURE SCOPE

In this study, we presented a unique movie recommendation system that integrates sentiment analysis, collaborative filtering, and content-based filtering. Our method tries to overcome the drawbacks of conventional recommendation systems that just depend on user ratings or reviews, which might not correctly reflect the user's genuine feelings towards a film. Our technology can gather sentimental data about films by using sentiment analysis, which allows us to provide consumers more precise and individualised suggestions.

Our tests on the Movielens dataset show that the system we've suggested works well. Precision, recall, and F1-score were some of the measures we used to assess the system's performance. We discovered that it performed better than conventional recommendation systems that just use content-based or collaborative filtering techniques. A better user experience was produced with the use of sentiment analysis, which improved the accuracy and promptness of suggestions.

Future Scope:

Although the results of our proposed system are encouraging, there are still a number of areas that need further research and development. First off, future work can take into account including other modalities like audio and visual characteristics as we only employed textual evaluations for sentiment analysis in this study. Second, we can investigate more sophisticated sentiment analysis methods, including deep learning models, to boost the precision of emotional

data extraction. Additionally, we only used one dataset for our experiments; future work may examine how well our system performs on various datasets to determine how generalizable it is.

In order to enhance the personalisation of suggestions, we may also investigate the usage of context-aware recommendation systems. Context-aware recommendation systems give consumers more relevant and individualised recommendations by taking into account extra contextual elements including location, time, and weather. Finally, to maximise the utility of the recommendations while ensuring a positive user experience, we can investigate the use of reinforcement learning techniques.

Overall, the effectiveness of our suggested solution demonstrates the possibility of integrating sentiment analysis with conventional recommendation techniques to deliver more precise and individualised movie suggestions. The goal of this research is to broaden and enhance our suggested system to include new features and provide consumers even more relevant and individualised suggestions.

REFERENCES

- [1] C. Basu, H. Hirsh, and W. Cohen, "Recommendation as classification: Using social and content-based information in recommendation," in *Proceedings of the 1998 15th National Conference on Artificial Intelligence, AAAI*, pp. 714–720, July 1998.
- [2] W. Xu, J. Wang, Z. Zhao, C. Sun, and J. Ma, "A Novel Intelligence Recommendation Model for Insurance Products with Consumer Segmentation," *Journal of Systems Science and Information*, vol. 2, no. 1, pp. 16–28, 2014.
- [3] Y. Xu, X. Guo, J. Hao, J. Ma, R. Y. K. Lau, and W. Xu, "Combining social network and semantic concept analysis for personalized academic researcher recommendation," *Decision Support Systems*, vol. 54, no. 1, pp. 564–573, 2012.
- [4] D. Guo, Z. Zhao, W. Xu et al., "How to find a comfortable bus route - Towards personalized information recommendation services," *Data Science Journal*, vol. 14, article no. 14, 2015.
- [5] D. Guo, Y. Zhu, W. Xu, S. Shang, and Z. Ding, "How to find appropriate automobile exhibition halls: Towards a personalized recommendation service for auto show," *Neurocomputing*, vol. 213, pp. 95101, 2016.
- [9] C. Ono, M. Kurokawa, Y. Motomura, and H. Asoh, "A context-aware movie preference model using a bayesian network for recommendation and promotion," in *Proceedings of 11th International Conference on User Modeling*, pp. 247–257, 2007.
- [10] M. Szomszor, C. Cattuto, H. Alani et al., "Folksonomies, the semantic web, and movie recommendation," in *Proceedings of 4th European Semantic Web Conference*, pp. 1–14, 2007.
- [11] T. De Pessemer, T. Deryckere, and L. Martens, "Context aware recommendations for user-generated content on a social network site," in *Proceedings of the EuroITV'09 - 7th European Conference on European Interactive Television Conference*, pp. 133–136, Belgium, June 2009.
- [12] J. L. Herlocker, J. Konstan, A. Borchers, and J. Riedl, "An algorithmic framework for performing collaborative filtering," in *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '99)*, pp. 230–237, Berkeley, Calif, USA, August 1999.
- [13] Y. Koren, "Collaborative filtering with temporal dynamics," *Communications of the ACM*, vol. 53, no. 4, pp. 89–97, 2010.
- [14] T. Hofmann, "Collaborative filtering via gaussian probabilistic latent semantic analysis," in *Proceedings of the 26th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 259–266, Toronto, Canada, 2003.
- G. Lekakos and P. Caravelas, "A hybrid approach for movie recommendation," *Multimedia Tools and Applications*, vol. 36, no. 1-2, pp. 55–70, 2008.
- Y. Zhang, Z. Tu, and Q. Wang, "TempoRec: Temporal-Topic Based Recommender for Social Network Services," *Mobile Networks and Applications*, vol. 22, pp. 1182–1191, 2017.
- S. Debnath, N. Ganguly, and P. Mitra, "Feature weighting in content-based recommendation system using social network analysis," in *Proceedings of the 17th International Conference on World Wide Web (WWW'08)*, pp. 1041–1042, Beijing, China, April 2008.
- M. Nazim Uddin, J. Shrestha, and G.-S. Jo, "Enhanced content-based filtering using diverse collaborative prediction for movie recommendation," in *Proceedings of the 2009 1st Asian Conference on Intelligent Information and Database Systems, ACIIDS 2009*, pp. 132–137, Viet Nam, April 2009.



16. A. Gunawardana and C. Meek, "A unified approach to building hybrid recommender systems," in *Proceedings of the 3rd ACM Conference on Recommender Systems, RecSys '09*, pp. 117–124, USA, October 2009.
17. Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1-167.
18. Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168-177.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor
Impact Factor: 8.379



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



www.ijircce.com

Scan to save the contact details