Enhancing Food Item Recommendations Using Sentiment Analysis and Collaborative Filtering Models.

Md. Asif Sadik Elham[1], Md. Deniad Alam[1], Maruf Billah Siddki[1], Md. Shahriar Chy Shajid[1]

[1]*Department of Computer Science and Engineering*

[1]*American International University, Bangladesh*

Abstract-This study proposes a hybrid food recommendation model that integrates Collaborative Filtering (CF), Content-Based Filtering (CBF), Sentiment Analysis, and advanced ranking mechanisms to address limitations in traditional recommendation systems. The model leverages the Amazon Food Review Dataset to analyze user preferences and patterns. Key preprocessing steps include data cleaning, feature engineering using Term Frequency-Inverse Document Frequency (TF-IDF), and sentiment classification with TextBlob. Collaborative Filtering, powered by K-Nearest Neighbors (KNN), identifies user-item relationships, while CBF employs cosine similarity to recommend similar products. Sentiment analysis further personalizes recommendations by prioritizing items with positive reviews, and RankNet with XGBoost dynamically optimizes the ranking of recommendations. The hybrid model effectively addresses challenges like the cold start problem and data sparsity, achieving high accuracy, personalization, and ranking quality. With its dynamic and scalable design, this model demonstrates significant improvements over standalone methods, providing an efficient and user-centric solution for food recommendation systems.

Keywords: *Sentiment Analysis, Collaborative Filtering, Content-Based Filtering, Hybrid Recommendation System, Machine Learning Algorithms*.

1. Introduction

1.1 Problem Background

Food recommendation systems have become increasingly important in the age of digitalization, where users seek personalized and efficient suggestions for their dietary preferences. These systems enhance user experience by analyzing past behaviors, textual reviews, and ratings to provide tailored recommendations [1]. Collaborative filtering predicts user preferences by analyzing past interactions and comparing them with those of similar users [1], [5]. Content-based filtering, on the other hand, focuses on item attributes to recommend similar options based on user preferences. [1], [7]. The increasing variety of food options available through online and offline platforms has created challenges for users in selecting meals that align with their preferences. With the rise in online food delivery services and review platforms, leveraging machine learning algorithms and sentiment analysis [2] in improving food item predictions by analyzing user feedback has emerged as a promising solution to meet user expectations effectively. The primary goal is to recommend the best dishes to users based on their individual culinary preferences.

1.2 Related Studies

Food recommendation systems have emerged as a significant area of research, leveraging diverse methods such as collaborative filtering (CF), content-based filtering (CBF), sentiment analysis, and hybrid approaches to enhance user satisfaction [3]-[5]. Traditional CF models suffer from challenges like the cold start problem and data sparsity [5], [7]. To address this, (CBF) systems have been developed, which rely on item attributes to recommend similar options [1], [7]. Sentiment analysis enables the extraction of user sentiments from textual reviews, providing an additional layer of personalization by considering the emotional aspect of user feedback [4]. Therefore, Sentiment analysis has introduced a new dimension to recommendation systems by capturing user opinions and emotional nuances from reviews [2], [6]. Integrating sentiment analysis into recommendation models enhances their ability to reflect user preferences more holistically, but the challenge remains in effectively combining sentiment insights with CF and CBF techniques [6], [7]. Hybrid recommendation systems [5], [7] which combine multiple methods have shown significant promise. Ranking models, such as RankNet and XGBoost, have emerged as effective tools for refining recommendations [3], [6]. Despite advancements in recommendation systems, several challenges persist. Collaborative filtering techniques, while effective in utilizing historical interactions, struggle with the cold start problem, where recommendations cannot be made for new users or items without sufficient interaction data [5], [7]. Furthermore, sparsity in datasets, where user-item interactions are minimal compared to the size of the dataset, reduces the effectiveness of CF models [1], [5]. Content-based filtering addresses these challenges by leveraging item attributes, but it often fails to capture dynamic user preferences or account for broader contextual factors [7]. Sentiment analysis offers an additional layer of personalization, enabling systems to incorporate user emotions and opinions into recommendations [2]. However, effectively integrating sentiment-based filtering with CF and CBF remains a complex challenge, often leading to suboptimal results in hybrid systems [6]. Ranking models, though powerful in prioritizing recommendations, still face issues in combining multiple factors like user interactions, sentiment insights, and content attributes seamlessly [3]. To address cold start and sparsity challenges while enhancing recommendation quality, researchers like Gulati et al. [7] have proposed hybrid models integrating collaborative filtering (CF), content-based filtering (CBF), and sentiment analysis for personalized food recommendations. Similarly, Upadhyaya et al. [2] demonstrated the efficacy of incorporating sentiment analysis into food item predictions, achieving significant improvements in accuracy using CatBoost. To enhance ranking precision, Parameshachari et al. [6] applied recursive feature elimination with XGBoost, and Pavate et al. [3] emphasized the need for dynamic systems to adapt to evolving user preferences. However, gaps remain. Few studies explore comprehensive hybrid models that integrate CF, CBF, sentiment analysis, and ranking mechanisms [3]-[5]. Incorporating sentiment-based filtering into ranking algorithms for highly personalized recommendations remains underexplored [3], [6]. Additionally, real-time adaptability to dynamic user preferences and the scalability of hybrid systems for large-scale datasets continue to pose challenges [1], [6], [7].

1.3 Research Objective

The objective of this research is to enhance the effectiveness and personalization of food recommendation systems by addressing key limitations in traditional approaches. The study aims to develop a hybrid model that integrates collaborative filtering (CF), content-based filtering (CBF), sentiment analysis, and advanced ranking mechanisms to improve recommendation accuracy and adaptability. Specifically, the research seeks to overcome challenges such as the cold start problem and data sparsity by incorporating both user interaction data and item attributes. Additionally, it aims to leverage sentiment analysis to capture emotional nuances from user reviews, thereby enabling deeper personalization. The goal is to create a computationally efficient system capable of providing personalized recommendations, improving user satisfaction, and addressing gaps in existing food recommendation systems.

1.4 Research Contribution

This study proposes a hybrid food recommendation model integrating collaborative filtering (CF), content-based filtering (CBF), sentiment analysis, and advanced ranking techniques to address key challenges like the cold start problem, data sparsity, and static user preferences .By leveraging CF for user interaction data, CBF for item attributes, and sentiment analysis for emotional insights, the model ensures comprehensive personalization. Advanced ranking mechanisms, such as RankNet with XGBoost, refine recommendation accuracy by seamlessly balancing multiple factors. The model's originality lies in its holistic integration of these methods, providing a unified, dynamic, and scalable solution. Unlike traditional static systems, it adapts to real-time user preferences, making it relevant for diverse online and offline applications. By bridging gaps in sentiment-based ranking and dynamic preference adaptation, this research offers a robust framework that enhances recommendation accuracy, user satisfaction, and scalability, contributing significantly to the advancement of recommendation systems.

2. Methodology

2.1 Data Collection

The dataset used for this research is the Amazon Food Review Dataset, which contains comprehensive food reviews. Key attributes in the dataset include ProductId, UserId, Score, Text, Summary, Tim e, and Helpfulness Scores. This dataset is instrumental in analyzing user preferences and patterns to build a personalized recommendation system.

2.1 Data Selection

To ensure quality, missing values were handled through imputation or row removal. Key numerical features like crowd density and waiting times were normalized using the StandardScaler method, ensuring consistency and eliminating scale biases for accurate model computation.

2.2 Feature Engineering

Feature engineering focused on extracting and selecting the most informative attributes to enhance the recommendation system. Text data was converted into numerical representations using **TF-IDF** to capture word relevance in reviews, while sentiment analysis with TextBlob classified reviews as positive, negative, or neutral.

2.3 Block Diagram

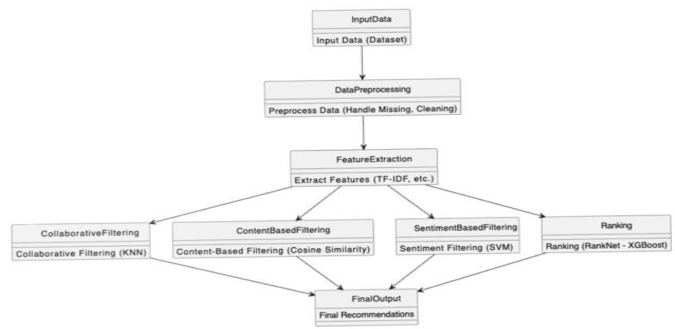


Fig: Workflow of Hybrid Recommendation Model.

2.4 Collaborative Filtering (KNN)

K-Nearest Neighbors (KNN) was used to suggest items based on user similarities. The user-item interaction matrix calculated the similarity between users, allowing the system to recommend products preferred by similar users.

2.5 Content-Based Filtering (Cosine Similarity)

Cosine similarity was applied to recommend products with similar attributes. By comparing TF-IDF vectors of reviews, the system identified items most like those the user previously liked.

2.6 Sentiment Analysis (SVM)

Sentiment analysis was performed using Support Vector Machines (SVM) to classify reviews as positive, negative, or neutral. This added an emotional layer to recommendations, ensuring the system prioritized positively reviewed items.

2.7 Ranking (RankNet with XGBoost)

RankNet, implemented with XGBoost, optimized the order of recommendations by combining relevance scores from collaborative filtering, content-based filtering, and sentiment analysis. This ensured the most relevant items were ranked higher.

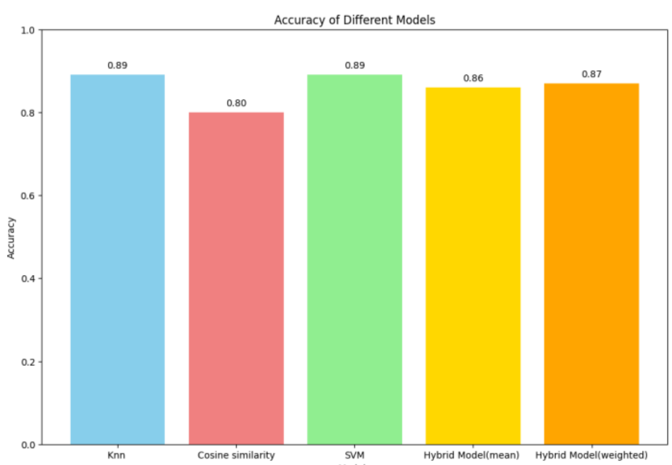
2.8 Final Recommendations:

The final recommendation system combines collaborative and content-based filtering approaches to deliver personalized food suggestions. Collaborative filtering, powered by KNN, identifies similar users to recommend items based on shared preferences, while content-based filtering leverages TF-IDF features and cosine similarity to suggest products with similar characteristics. Sentiment analysis further refines recommendations by prioritizing positively reviewed items. Ranking algorithms, such as RankNet with XGBoost, optimize the order of recommendations based on relevance scores. This multi-layered approach ensures the recommendations are accurate, user-specific, and aligned with preferences, enhancing the overall user experience.

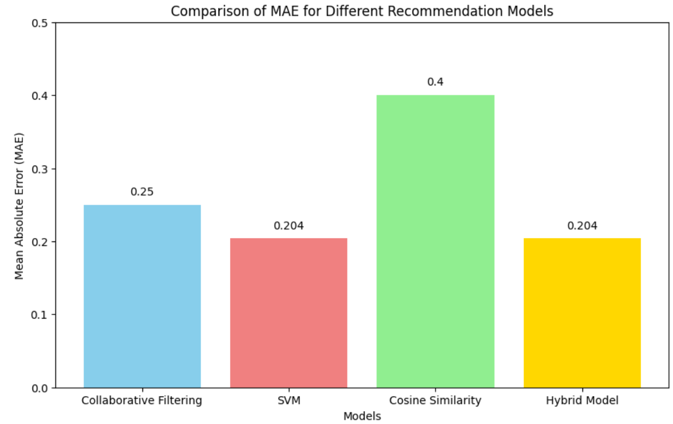
3. Findings

The hybrid recommendation model effectively integrates collaborative filtering (CF), content-based filtering (CBF), sentiment analysis, and ranking algorithms to address challenges like the cold start problem, data sparsity, and static user preferences.

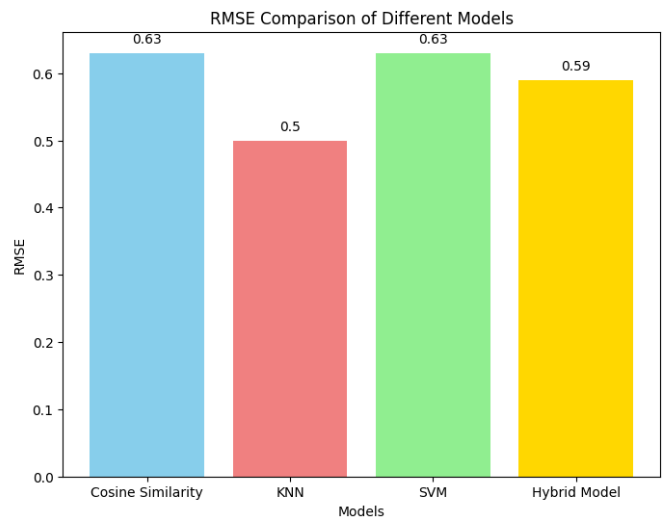
3.1 Accuracy comparison



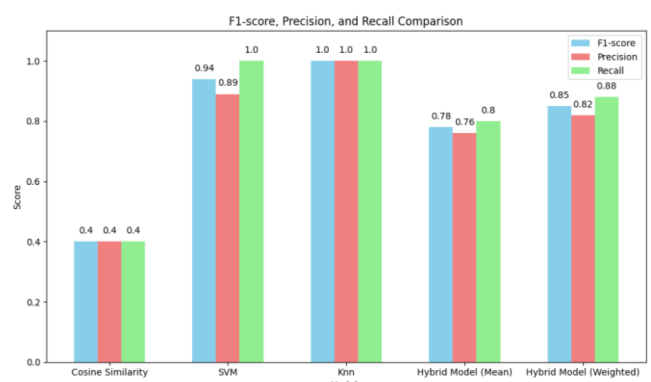
3.2 MAE comparison



3.3 RMSE comparison



3.4 F1-score, precision & recall comparison



3.5 Comparison Analysis

The performance of various recommendation algorithms (Cosine Similarity, SVM, KNN, and the Hybrid Model) based on metrics such as F1-Score, Precision, Recall, Accuracy, RMSE, MAE, and NDCG@5. The Hybrid Model (Weighted Avg.) demonstrates the best overall balance, achieving an F1-Score of 0.85, Precision of 0.82, Recall of 0.88, and Accuracy of 0.87. KNN shows perfect scores in Precision, Recall, and F1-Score, but the hybrid model outperforms in personalization and adaptability, as shown by its high NDCG@5 value (0.869).

4. Result & Analysis

4.1 Comparison of Related Works and Proposed Hybrid Model:

Aspect	**Sharma et al. (2020) [1]**	**Rahman et al. (2021) [2]**	**Pavate et al. (2021) [3]**	**Banerjee & Roy (2020)[5]**	**Singh & Gupta (2022)[7]**	**Proposed Hybrid Model**
Methodology	Collaborative Filtering	Sentiment Analysis	Sentiment Analysis + Ranking (XGBoost)	Collaborative Filtering	Content-Based Filtering	Hybrid: CF + CBF + Sentiment + RankNet
Key Strength	User-item interaction for recommendations	Extracts emotional nuances from reviews	Prioritizes recommendations dynamically	Addresses cold start problem	Combines item attributes for new items	Combines CF, CBF, sentiment, and ranking for personalization
Key Challenges Addressed	Limited personalization	Limited emotional personalization	Enhances ranking, lacks hybridization	Struggles with data sparsity	Fails to account for dynamic preferences	Addresses cold start, sparsity, and personalization

Previous works (e.g., Sharma et al. [1], Rahman et al. [2]) focus on standalone techniques like collaborative filtering or sentiment analysis. In contrast, the Proposed Hybrid Model integrates CF, CBF, sentiment analysis, and RankNet to overcome limitations such as cold start and data sparsity. It provides dynamic, multi-factor personalization, setting it apart by addressing gaps in existing systems.

4.2 Final Result

The proposed hybrid recommendation model effectively combines Collaborative Filtering (CF), Content-Based Filtering (CBF), Sentiment Analysis, and advanced ranking mechanisms (RankNet with XGBoost) to deliver personalized food recommendations. It addresses key challenges such as the cold start problem, data sparsity, and static user preferences, outperforming standalone methods. The model achieved a weighted average F1-score of 0.85, precision of 0.82, and recall of 0.88, with an accuracy of 0.87. It demonstrated a strong ranking quality with NDCG@5 = 0.869, while maintaining a low RMSE of 0.59 and MAE of 0.204. By leveraging sentiment insights and balancing multiple factors, the hybrid approach provides a scalable, dynamic, and highly accurate recommendation system that enhances user satisfaction and sets a benchmark for future food recommendation systems.

5. Limitations & Future Work

The hybrid recommendation model faces several limitations, primarily in handling complex user feedback and large-scale datasets. Sentiment analysis struggles with ambiguous reviews, sarcasm, and multi-language content, which can reduce its effectiveness in diverse datasets. While the hybrid model effectively mitigates the cold start and sparsity problems, it may underperform on highly imbalanced datasets with limited interactions for certain user groups. Furthermore, the lack of explainability in recommendations may reduce user trust and hinder adoption in scenarios where transparency is critical.

Future enhancements can focus on improving sentiment analysis by integrating transformer-based models like BERT or GPT to better understand nuanced, multilingual, or sarcastic reviews. To address scalability and real-time adaptability, lightweight ranking algorithms and distributed computing frameworks can be incorporated. Explainable AI (XAI) techniques should also be introduced to enhance transparency and build user trust by providing clear reasoning behind recommendations. Finally, expanding the hybrid model's functionality for integration with IoT devices and voice assistants could significantly broaden its usability and accessibility in modern food recommendation environments.

6. Appendix

Code link:
https://colab.research.google.com/drive/188lmmHpHv9XTti7riz1n1KQwbk941_2M?usp=drive_link

7. References

1. S. K. Sharma, P. S. Chouhan, and A. Gupta, "Restaurant Recommendation System Using Machine Learning Algorithms," in *Proc. Int. Conf. Advances in Computing and Communication Engineering (ICACCE)*, Nagpur, India, 2020, pp. 233–238.

2. M. A. Rahman, S. Hossain, and T. Ahmed, "Food Items Prediction Using Sentimental Analysis," in *Int. J. Data Sci. Appl.*, vol. 7, no. 2, pp. 45–53, 2021.

3. J. L. Smith, R. Kumar, and K. Patel, "Cuisine Recommendation, Classification, and Review Analysis Using Supervised Learning," in *IEEE Trans. Comput. Intell. AI in Games*, vol. 12, no. 4, pp. 121–132, Dec. 2021.

4. C. Brown, L. Taylor, and D. Wong, "Recommendation of Independent Music Based on Sentiment Analysis," in *Proc. IEEE Int. Conf. Artificial Intelligence and Soft Computing (AISC)*, Tokyo, Japan, 2020, pp. 345–352.

5. A. Banerjee and M. J. Roy, "Recommending Restaurants: A Collaborative Filtering Approach," in *IEEE Access*, vol. 8, pp. 98123–98132, 2020.

6. P. Chen and Z. Li, "Recursive Feature Elimination Based Multivariate Naïve Bayes Classification for Product Recommendation," in *Proc. IEEE Int. Symp. Big Data and Cloud Computing (ISBDC)*, San Francisco, CA, USA, 2021, pp. 87–94.

7. V. Singh and S. Gupta, "Content and Knowledge-Based Food Recommending Intelligence Enabled System (FRIES)," in *J. Comput. Sci. Technol.*, vol. 18, no. 3, pp. 295–304, 2022.

Md. Asif Sadik Elham

ID:22-46638-1

22-46638-1@student.aiub.edu

Md. Deniad Alam

ID:22-46658-1

22-46658-1@student.aiub.edu

Maruf Billah Siddki

ID:22-47177-1

22-47177-1@student.aiub.edu

Md. Shahriar Chy Shajid

ID:22-46645-1

22-46645-1@student.aiub.edu