IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 6, December 2026, pp. 5968∼5979
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i6.pp5968-5979

Sentiment–topic driven recommendation system for enhanced personalization

El Mehdi Lghaouch[1], Soumaya Ounacer[1], Soufiane Ardchir[2], Mohamed Azzouazi[1]
[1]Laboratory of Information Technologies and Modeling, Faculty of Sciences Ben M'sik, Hassan II University, Casablanca, Morocco
[2]National School of Business and Management, Hassan II University, Casablanca, Morocco

Article Info

Article history:

Received Nov 19, 2025
Revised Aug 19, 2026
Accepted Sep 20, 2026

Keywords:

BERTopic
Collaborative filtering
DistilBERT
Recommendation systems
Sentiment analysis
Singular value decomposition++
Topic modeling

ABSTRACT

Personalized recommendations are an important factor in enhancing user satisfaction on the digital platforms. This research introduces a hybrid recommendation system that combines sentiment analysis and topic modeling with collaborative filtering to improve both accuracy and personalization. By leveraging distilled bidirectional encoder representations from transformers (DistilBERT) for sentiment analysis and bidirectional encoder representations from transformers topic modeling (BERTopic) to uncover key themes in user reviews, the proposed sentiment-topic singular value decomposition++ (SVD++) model captures both the emotional tone and specific preferences expressed in textual feedback. Experimental results demonstrate significant enhancements against traditional approaches, delivering more relevant and context-aware recommendations. Looking ahead, future work will focus on integrating real-time feedback, enhancing interpretability, and extending the framework to multi-domain scenarios, further refining the system's capabilities and adaptability.

This is an open access article under the CC BY-SA license.

Corresponding Author:

El Mehdi Lghaouch
Laboratory of Information Technologies and Modeling, Faculty of Sciences Ben M'sik, Hassan II University
Casablanca, Morocco
Email: elmehdi.lghaouch-etu@etu.univh2c.ma

1. INTRODUCTION

A recommendation system is a form of technology that analyzes user preferences and behaviors for the purpose of recommending certain products, services, or content [1]. These recommendation systems leverage data, such as user ratings, clicks, or browsing history, to offer personalized experiences. They play an important role in many fields: in e-commerce [2], they suggest products customers might be interested in; on streaming platforms they recommend movies or music [3]; and on social media [4], they allow finding relevant content or profiles. Consequently, recommendation systems not only improve the relevance of suggested items but also contribute to a more satisfying user experience.

Traditional recommendation techniques generally rely on two major paradigms. Collaborative filtering [5] exploits similarities among users or items to generate recommendations, whereas content-based filtering [6] recommends items according to characteristics shared with those previously preferred by the user. Although these approaches have demonstrated their effectiveness in many recommendation scenarios, they often fail to fully exploit the rich semantic information embedded in textual reviews. User reviews frequently express detailed opinions about multiple product characteristics, such as quality, price, durability, or performance [7]. Since users may assign different levels of importance to these characteristics,

relying solely on numerical ratings may overlook valuable preference information and consequently limit recommendation quality.

To overcome these limitations, increasing attention has been devoted to combining sentiment analysis with topic modeling. Sentiment analysis aims to determine the polarity and emotional orientation conveyed in textual opinions; however, when used independently, it provides only a global assessment of user sentiment without explicitly identifying the product aspects responsible for that opinion [8], [9]. In contrast, topic modeling is an unsupervised learning technique which allows for the automatic discovery of latent semantic topics from large collections of unstructured documents [10]. These latent topics often correspond to meaningful product attributes, including durability, design, functionality, or cost, thereby offering a structured representation of users' concerns.

The integration of sentiments analysis and topic models allows recommendation systems to capture both users' opinions and the underlying product aspects discussed in textual reviews, resulting in richer user preference representations and more context-aware recommendations. For example, if a user consistently expresses positive opinions regarding product durability, the recommender system can assign greater importance to products whose reviews exhibit positive sentiment toward this particular aspect identified through topic modeling [11]. Furthermore, recent progress in deep learning, particularly transformer-based models such as bidirectional encoder representations from transformers (BERT), has significantly enhanced the performance of sentiment analysis and topic extraction, allowing recommendation systems to exploit textual information with greater precision [12], [13]. As a result, the combination of these techniques has demonstrated its effectiveness in improving recommendation accuracy across various applications, including e-commerce, customer review analysis, and social media services [13].

The remainder of this paper is organized as follows. Section 2 reviews the existing literature on recommendation systems integrating sentiment analysis and topic modeling. Section 3 describes the proposed framework, including data preprocessing, sentiment extraction using distilled bidirectional encoder representations from transformers (DistilBERT), topic discovery through bidirectional encoder representations from transformers topic modeling (BERTopic), and the hybrid singular value decomposition++ (SVD++) recommendation model. Section 4 presents the experimental results together with a comparative discussion. Finally, section 5 summarizes the main findings of this work and outlines potential directions for future research.

2. RELATED WORK

In today's e-commerce and streaming service landscape, recommender systems have become essential to customizing user-based experiences. Earlier recommender approaches rely mainly on explicit user ratings. However, explicit ratings alone may not capture the underlying opinions and context behind user preferences. As a result, many researchers have used sentiment analysis and topic modeling as complementary processes to traditional recommender system methods.

Numerous studies have investigated the contribution of sentiment analysis to recommendation performance. Al-Ajlan and Alshareef [14] proposed a recommender system for Arabic content that exploits sentiment extracted from user reviews. Their findings demonstrated that incorporating users' emotional expressions improves item relevance beyond what can be achieved using numerical ratings alone. Likewise, Gao and Li [15] developed a personalized recommendation framework based on semantic sentiment analysis, reporting improved recommendation accuracy through the integration of sentiment-aware representations.

Kumar *et al.* [16] introduced a hybrid recommendation framework that integrates collaborative filtering, content-based filtering, and sentiment analysis using the naive Bayes algorithm. Their approach exploits sentiment information extracted from microblogging data to generate more refined recommendations. Experimental results showed that incorporating user opinions into the recommendation process leads to more relevant and personalized suggestions.

Topic modeling, beyond sentiment analysis, is also effective in identifying hidden topics in review text and expanding knowledge of user preferences. Liang *et al.* [17] constructed a joint sentiment-topic model with rating and review (JST-RR) model that jointly models ratings and reviews to predict sentiment and topic distributions. Their results showed improved recommendations due to better alignment between sentiments and topics. In another contribution, Wang *et al.* [18] proposed the matrix factorization fusing reviews (MFFR) model to address rating sparsity in recommender systems. Latent Dirichlet allocation (LDA) was used to

generate a preference matrix from user reviews, which was then jointly factorized with the original rating matrix. By directly integrating topic features from reviews into the rating model, MFFR improved rating prediction accuracy, particularly for sparse datasets.

Liu and Zhao [19] developed the sentiment analysis and matrix factorization (SAMF) model to overcome issues of data sparsity and credibility. They used LDA to construct a preference matrix from latent topics in reviews, then combined these matrices with the original rating data via matrix factorization. Furthermore, sentiment information obtained with BERT was used to refine ratings, resulting in better recommendation performance.

Hybrid approaches that combine multiple techniques to capture various dimensions of user preferences have become increasingly popular. Dang *et al.* [20] used deep learning models to integrate item genres (as aspects) with sentiment-driven insights to provide more comprehensive recommendations. Kanwal *et al.* [21] leveraged convolutional neural networks (CNNs) and topic modeling to refine user-specific recommendations by combining structured (rating) and unstructured (text) data. Xie *et al.* [22] combined sentiment analysis and topic models of reviews in their study of movie recommendations, showing that joint use of these techniques improves results compared to using either alone.

Despite these advancements, challenges remain. One of the main limitations lies in the interpretability of complex models, particularly deep learning-based approaches, where understanding how recommendations are generated is not always straightforward. Additionally, while sentiment analysis and topic modeling improve recommendation accuracy, their effectiveness varies across datasets and domains. The impact of sentiment polarity on recommendation quality is not always well-quantified, raising questions about how best to weight sentiment scores in predictive models. Furthermore, many current systems still struggle to incorporate user context dynamically, which is crucial for making real-time, highly relevant recommendations.

3. METHOD

Figure 1 presents the proposed recommendation system, which integrates sentiment analysis, topic modeling, and collaborative filtering to enhance rating prediction and recommendation accuracy. The methodology is organized into several stages: data preprocessing, sentiment extraction, topic identification, hybrid rating estimation, and performance evaluation using standard metrics. Unlike conventional collaborative filtering methods that depend exclusively on numerical ratings, the proposed approach also exploits textual reviews to represent sentiment polarity and semantic user preferences through sentiment analysis and topic modeling, respectively. Combining these complementary information sources supports more personalized and interpretable recommendations while considering both explicit ratings and implicit user preferences.

3.1. Data preprocessing

The experiments are conducted on the Amazon fine food reviews dataset, which provides user identifiers, product identifiers, numerical ratings, and textual reviews collected over a ten-year period ending in October 2012. It comprises 568,454 reviews associated with approximately 74,000 products and 260,000 users, reflecting substantial diversity among consumers interacting with the products. Reviews contain an average of 95 words, offering sufficient textual information for sentiment analysis. The mean product rating is 4.18, while 5 stars represents the most frequent score, revealing an overall positive rating tendency commonly observed in e-commerce data.

Text preparation constitutes the initial stage of the analysis. Reviews are cleaned by eliminating stop words, punctuation marks, special characters, and other unnecessary textual noise while retaining informative content. The cleaned reviews are subsequently divided into individual tokens through tokenization. Words are then standardized toward their basic forms by applying lemmatization and stemming. These preprocessing operations produce a more consistent textual representation and prepare the review content for the subsequent analytical stages.

Following textual preprocessing, additional cleaning operations are applied to the remaining dataset. Missing information is addressed through either imputation or removal of incomplete records, while duplicated instances are eliminated to reduce redundancy. Users contributing fewer than 10 reviews and products receiving fewer than 5 reviews are excluded to retain interactions involving sufficiently active users and represented items. Numerical attributes, including ratings, are normalized when appropriate, sentiment-related features are extracted, and the resulting dataset is finally partitioned into training and testing subsets for model assessment. After preprocessing, the dataset contains numerical attributes, including score for explicit ratings, together

with textual attributes such as text and CleanedText, which support sentiment and topic extraction. Table 1 summarizes the resulting dataset structure.

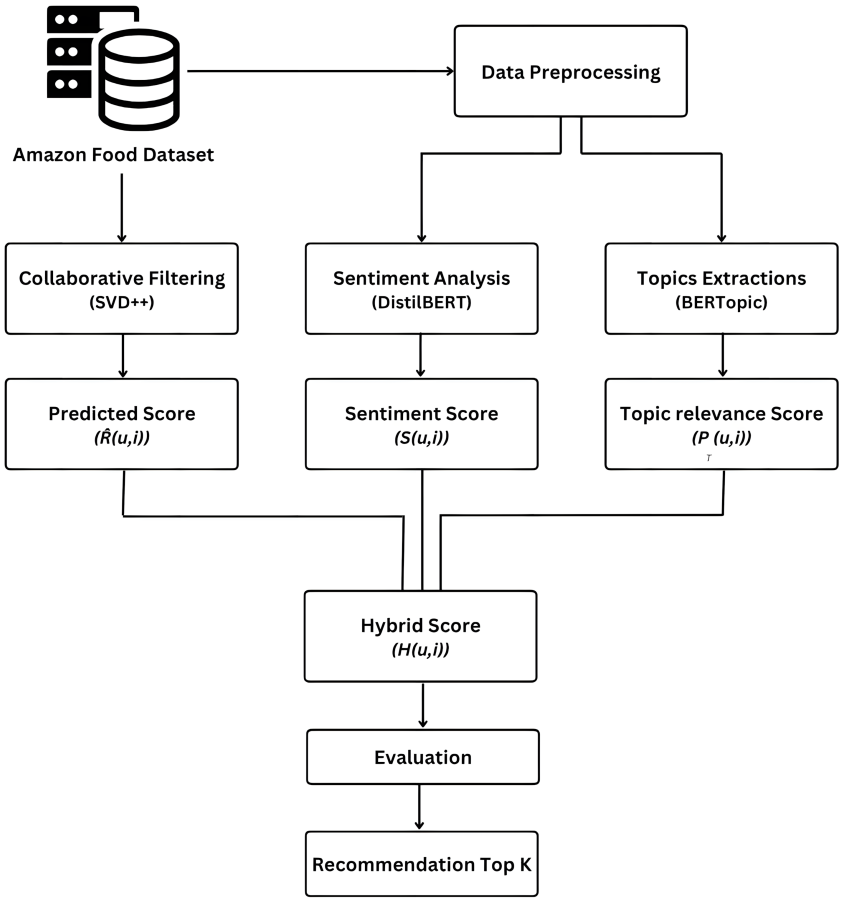

Figure 1. Proposed recommendation system

Table 1. Dataset structure after preprocessing

Feature name	Description
Id	Unique identifier for each record
ProductId	Unique identifier for products
UserId	Unique identifier for users
Score	Explicit user rating (1 to 5)
CleanedText	Preprocessed and cleaned text from reviews

3.2. Sentiment analysis with DistilBERT

Sentiment information is extracted from user reviews using DistilBERT, a pre-trained transformer language model subsequently fine-tuned for sentiment classification. DistilBERT is a compact version of BERT that maintains most of its predictive performance while reducing computational requirements [23]. Introducing sentiment analysis into the recommendation process makes it possible to capture implicit emotional information expressed in textual feedback, thereby complementing the information available from explicit numerical ratings.

Before classification, each review is processed through tokenization, lowercasing, stop-word removal, and punctuation normalization. The resulting text is supplied to the DistilBERT classifier, where self-attention mechanisms are used to model contextual relationships and linguistic phenomena including negation, irony, and comparative expressions. This preparation allows the classifier to emphasize semantically informative content while retaining the contextual information required for reliable sentiment identification.

For fine-tuning, the model uses a batch size of 64 and a learning rate of 2×10^{-5} to maintain stable convergence. Every review receives one of three sentiment labels—positive, negative, or neutral—along with its corresponding confidence score. These categorical predictions are then converted into numerical representations for subsequent integration within the recommendation framework.

For a given user–item pair (u, i), the sentiment score is defined as (1).

$$S(u,i) = (-1)^{1_{\text{neg}}} \cdot P_{\text{sentiment}} \tag{1}$$

Where $P_{\text{sentiment}}$ represents the confidence associated with the predicted sentiment, and $1_{\text{neg}} = 1$ when the review is classified as negative and 0 otherwise. Under this formulation, positive opinions are mapped to $[0, 1]$, whereas negative opinions fall within $[-1, 0]$. The absolute value reflects prediction confidence, allowing polarity and sentiment strength to be represented simultaneously.

For instance, when a positive review has $P_{\text{sentiment}} = 0.85$, the resulting score is:

$$S(u,i) = (-1)^0 \cdot 0.85 = 0.85$$

whereas a negative review associated with $P_{\text{sentiment}} = 0.75$ produces:

$$S(u,i) = (-1)^1 \cdot 0.75 = -0.75.$$

These two examples show how polarity and confidence are represented jointly through a single scalar sentiment value.

The resulting sentiment values are incorporated into the hybrid recommendation model to enrich rating prediction with emotional information obtained from textual feedback. Accordingly, users' subjective opinions can be considered alongside their explicit ratings. This additional signal helps the model distinguish between products that may receive comparable numerical scores while being described through different emotional responses in their reviews.

3.3. Topic modeling with BERTopic

Topic modeling comprises unsupervised learning methods intended to identify latent thematic structures from recurring word patterns in large collections of unstructured text [24]. Conventional approaches, such as LDA mainly depend on word co-occurrence information and may therefore provide a limited representation of deeper semantic relationships. BERTopic addresses this limitation by combining transformer-derived embeddings with clustering techniques, offering a modern approach for extracting semantically meaningful topics [25].

In BERTopic, pre-trained language models such as BERT are first used to represent documents as dense contextual embeddings. Uniform manifold approximation and projection (UMAP) subsequently reduces the dimensionality of these representations, after which hierarchical density-based spatial clustering of applications with noise (HDBSCAN) groups semantically related reviews into topic clusters. Through this sequence of operations, semantic proximity and density information are preserved within the embedding representation, supporting the extraction of coherent and interpretable topics.

Each resulting topic is characterized by representative keywords, while the probability assigned to a document is determined from its position relative to the corresponding cluster. The resulting topic distributions provide a concise representation of the principal aspects discussed within user reviews. Hence, one or several dominant thematic dimensions can be associated with each review.

For each user–item pair, topic relevance is calculated as (2).

$$P_T(u,i) = \sum_T P(T \mid R) \cdot I_T \tag{2}$$

Where $P(T \mid R)$ is the probability of topic T given the review set R, while I_T denotes the importance assigned to topic T for item i. This expression aggregates topic information from the reviews related to the same user–item interaction. It consequently identifies the aspects, such as taste, packaging, or price, that contribute most strongly to the user's evaluation.

3.4. Hybrid rating prediction with SVD++

The collaborative filtering stage relies on SVD++, a matrix factorization extension that incorporates implicit feedback when estimating user ratings. In comparison with standard matrix factorization, SVD++ jointly exploits explicit rating information and implicit interaction signals when representing user preferences. This property makes the method particularly appropriate for recommendation settings characterized by sparse rating information.

The estimated rating for user u and item i is expressed as (3).

$$\hat{R}(u,i) = \mu + b_u + b_i + q_i^T \left(p_u + |N_u|^{-0.5} \sum_{j \in N_u} y_j \right) \tag{3}$$

Where μ denotes the overall mean rating, b_u and b_i correspond to the user and item bias terms, p_u and q_i represent their latent factor vectors, and N_u contains the items rated by user u. The implicit feedback term $|N_u|^{-0.5} \sum_{j \in N_u} y_j$ contributes additional information when user ratings are sparse. This representation therefore accounts for both long-term preference patterns and interaction-related information.

The final hybrid score combines collaborative prediction with sentiment and topic information as (4).

$$H(u,i) = \alpha \cdot \hat{R}(u,i) + \beta \cdot S(u,i) + \gamma \cdot P_T(u,i) \tag{4}$$

Where α, β, and γ specify the relative contribution of the three components. These coefficients combine explicit numerical information with sentiment polarity and topic relevance extracted from reviews. Their values are determined through grid-search optimization to obtain the strongest overall recommendation performance.

3.5. Evaluation metrics

Model performance is examined from two perspectives: numerical rating prediction and top-k recommendation quality. Rating-oriented measures quantify the difference between the predicted and actual ratings. In contrast, top-k measures evaluate the relevance of the ranked items returned to users.

3.5.1. Rating prediction accuracy

Rating prediction accuracy is evaluated using root mean squared error (RMSE) and mean absolute error (MAE), as defined in (5) and (6). Both metrics quantify deviations between estimated and observed ratings, although RMSE assigns greater influence to larger prediction errors. They therefore provide complementary measures of the accuracy of the predicted scores.

$$RMSE = \sqrt{\frac{1}{|T|} \sum_{(u,i) \in T} \left(R(u,i) - H(u,i) \right)^2} \tag{5}$$

$$MAE = \frac{1}{|T|} \sum_{(u,i) \in T} |R(u,i) - H(u,i)| \tag{6}$$

Where $R(u,i)$ denotes the observed user rating and $H(u,i)$ represents the hybrid prediction. Smaller RMSE and MAE values correspond to more accurate rating estimates. Both measures are commonly employed to quantify numerical prediction error in recommender-system evaluation.

3.5.2. Top-k recommendation metrics

The quality of ranked recommendations is measured using Precision@k, Recall@k, and F1-score@k. Precision@k quantifies how many items within the top-k list are relevant as defined in (7), whereas Recall@k determines how much of the relevant item set is recovered by the recommendation list as defined in (8). F1-score@k combines these two measures into a single balanced indicator as defined in (9).

$$Precision@k = \frac{1}{|U|} \sum_{u \in U} \frac{|R(u) \cap T(u)|}{|R(u)|} \tag{7}$$

$$Recall@k = \frac{1}{|U|} \sum_{u \in U} \frac{|R(u) \cap T(u)|}{|T(u)|} \tag{8}$$

$$F1 - score@k = \frac{2 \cdot Precision@k \cdot Recall@k}{Precision@k + Recall@k} \tag{9}$$

Relevant items $T(u)$ are determined according to three alternatives: i) median rating threshold, ii) mean rating plus standard deviation threshold, and iii) percentile-based threshold. These definitions make it possible to determine relevant items under different assumptions regarding rating distributions. Among the evaluated alternatives, the 75th percentile threshold provided the most consistent behavior across different rating distributions.

3.6. Experimental setup and implementation details

This subsection reports the experimental configuration and implementation parameters used to reproduce the proposed recommendation framework. The same settings were maintained throughout the experiments. The dataset was divided according to a fixed 70% training and 30% testing partition. This identical split and evaluation procedure was used for both the baseline SVD++ implementation and the proposed sentiment-topic SVD++ model. A fixed random seed was additionally maintained throughout the experiments to limit variations caused by random initialization.

3.6.1. Sentiment analysis implementation

Sentiment extraction used the distilbert-base-uncased transformer checkpoint. The model was fine-tuned for three epochs using adaptive moment estimation with decoupled weight decay (AdamW) with a learning rate of 2×10^{-5} and a batch size of 64. Review sequences were padded or truncated to a maximum of 128 tokens. The classifier produced a polarity label and confidence value for every review, after which the result was converted into the numerical sentiment representation $S(u, i)$ given in (1).

3.6.2. Topic modeling implementation

Topic extraction was performed with BERTopic using the all-MiniLM-L6-v2 model for sentence embeddings. UMAP reduced the embedding dimensionality using $n_neighbors = 15$, $n_components = 5$, and cosine distance. HDBSCAN then performed clustering with a minimum cluster size of 30. The corresponding topic probabilities were subsequently employed to calculate the relevance measure $P_T(u, i)$ presented in (2).

3.6.3. Collaborative filtering implementation

The SVD++ implementation provided by the Surprise library was used as the collaborative filtering component. Its latent dimension was configured to 50 factors, with a learning rate of 0.005 and a regularization coefficient of 0.02. Training continued for 30 epochs. The baseline uses this SVD++ configuration alone, whereas the proposed system incorporates sentiment and topic information according to the hybrid expression in (4).

3.6.4. Weight optimization and evaluation protocol

The values of α, β, and γ were determined using grid search over $\{0.1, 0.2, \ldots, 0.9\}$ on validation data derived from the training portion. The configuration producing the best performance ($\alpha = 0.3062$, $\beta = 0.5870$, $\gamma = 0.1068$) was retained for the final experiments. All reported performance measures were subsequently calculated only on the independent test partition to avoid information leakage. Table 2 consolidates the main implementation settings and experimental parameters used throughout the evaluation.

Table 2. Implementation details and experimental parameters

Component	Parameters and settings
Data split	70% training / 30% testing; fixed random seed
Sentiment model	DistilBERT (distilbert-base-uncased); AdamW optimizer; learning rate 2×10^{-5}; batch size 64; epochs 3; max sequence length 128
Topic modeling	BERTopic with all-MiniLM-L6-v2 embeddings; UMAP ($n_neighbors = 15$, $n_components = 5$, cosine metric); HDBSCAN (min_cluster_size=30)
Collaborative filtering	SVD++ (Surprise library); latent factors 50; learning rate 0.005; regularization 0.02; epochs 30
Hybrid weighting	Grid search on α, β, $\gamma \in \{0.1, \ldots, 0.9\}$; selected values $\alpha = 0.3062$, $\beta = 0.5870$, $\gamma = 0.1068$
Evaluation	RMSE, MAE, Precision@k, Recall@k, F1-score@k computed on test set only

4. RESULTS AND DISCUSSION

This section presents and discusses the experimental results of the proposed sentiment-topic SVD++ model in comparison with the baseline SVD++ approach. The evaluation considers both rating prediction accuracy and top-k recommendation quality. The results are analyzed to assess the contribution of sentiment analysis and topic modeling to the performance of the hybrid recommendation framework.

4.1. Results

This subsection presents the experimental evaluation of the proposed sentiment-topic SVD++ model in comparison with the conventional SVD++ baseline. By integrating sentiment information and topic representations into the collaborative filtering process, the proposed approach consistently outperforms the baseline across all evaluation metrics. The hybrid prediction model combines collaborative filtering, sentiment analysis, and topic modeling using the optimized weighting coefficients $\alpha = 0.3062$, $\beta = 0.5870$, and $\gamma = 0.1068$, obtained through a grid search procedure. The observed improvements indicate that textual information extracted from user reviews provides complementary knowledge beyond explicit ratings, resulting in more accurate and personalized recommendations. A detailed comparison between the baseline SVD++ model and the proposed sentiment-topic SVD++ approach is presented in Table 3. Table 3 reports Precision@k, Recall@k, and F1-score@k for different recommendation list sizes, providing a comprehensive assessment of the recommendation quality achieved by both models.

Table 3. Comparison of recommendation quality for baseline SVD++ and sentiment-topic SVD++

k	Baseline SVD++			Sentiment-topic SVD++		
	Precision@k	Recall@k	F1-score@k	Precision@k	Recall@k	F1-score@k
5	0.790	0.839	0.792	0.838	0.937	0.856
10	0.782	0.880	0.811	0.818	0.988	0.871
15	0.780	0.887	0.814	0.815	0.995	0.873
20	0.780	0.889	0.816	0.814	0.998	0.873
25	0.780	0.890	0.816	0.813	0.999	0.874
30	0.780	0.890	0.816	0.813	0.999	0.874

Figure 2 shows a comparison of the predictive performance of the baseline SVD++ model and the proposed sentiment-topic SVD++ framework, using the RMSE and MAE evaluation metrics. RMSE highlights the largest prediction errors, while MAE measures the mean absolute deviation between the predicted ratings and the actual ratings. Together, these metrics help to provide complementary insights into the predictive capability of recommendation models and highlight the improvement achieved by the proposed approach.

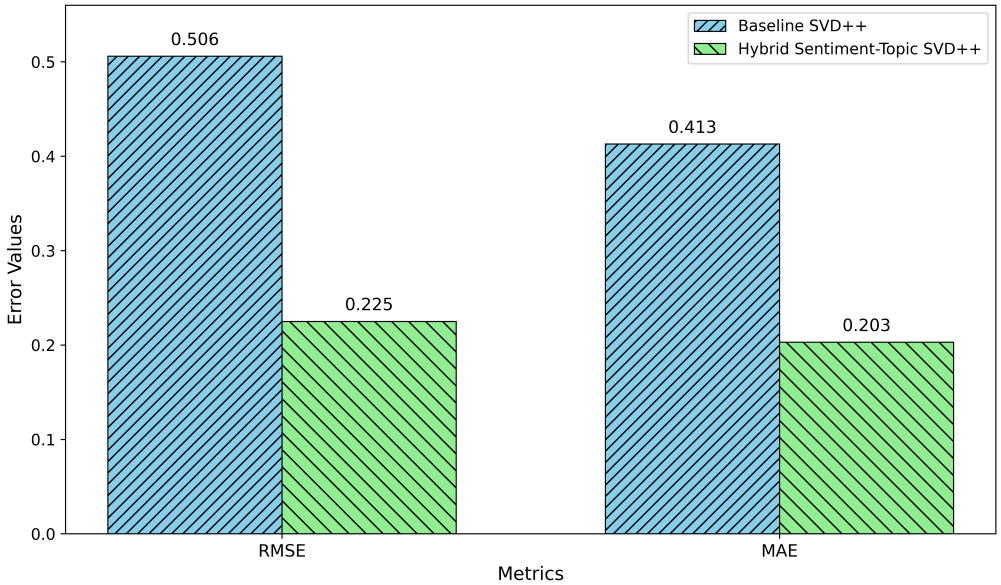

Figure 2. Comparison of prediction accuracy for the baseline SVD++ and sentiment-topic SVD++

To further position the proposed model with respect to representative recommendation approaches, its performance was also compared with the MFFR and SAMF models reported in [18], [19], both evaluated on the Amazon food dataset. The comparison presented in Table 4 shows that the proposed sentiment-topic SVD++ model consistently achieves the highest F1-score across all recommendation cutoffs ($k = 10$ to $k = 30$). For instance, at $k = 10$, the proposed model attains an F1-score of 0.871, outperforming SAMF (0.75) and MFFR (0.64). The same trend is observed for larger recommendation lists, where the proposed approach reaches an F1-score of 0.874 at $k = 30$, compared with 0.84 for SAMF and 0.81 for MFFR. These results demonstrate that integrating sentiments analysis and topic modeling into the collaborative filtering framework substantially enhances recommendation effectiveness.

Table 4. F1-score on Amazon food dataset

k	MFFR	SAMF	Sentiment-topic SVD++
10	0.64	0.75	0.871
15	0.66	0.76	0.873
20	0.67	0.80	0.873
25	0.80	0.82	0.874
30	0.81	0.84	0.874

4.2. Discussion

The experimental findings highlight the improvement of the sentiment-topic SVD++ model over the baseline SVD++ approach. The clearest gains concern precision and recall. For every evaluated value of k, the hybrid model obtains higher Precision@k and Recall@k scores. At $k = 5$, Precision@k rises from 0.790 to 0.838, while Recall@k increases from 0.839 to 0.937. These results show that the hybrid approach retrieves more relevant items while increasing the proportion of useful recommendations. Such improvements are valuable in practical settings, where presenting relevant items can improve user experience.

A similar pattern is observed for F1-score@k. At $k = 5$, F1-score@k increases by 8.1%, from 0.792 to 0.856, while the relative gain remains stable and reaches 7.1% at $k = 30$. The improvement across recommendation list sizes indicates that the hybrid model maintains reliable performance for varying values of k. Its F1-score of 0.856 at $k = 5$ is particularly relevant for practical applications, where only a limited number of items are typically presented to users.

Sentiment analysis contributes significantly to the hybrid model's performance. By examining the emotional orientation expressed in user reviews, the system can favor items consistent with users' preferences and attitudes. For instance, when a user repeatedly emphasizes durability and expresses positive opinions about this aspect, the model can prioritize products associated with favorable sentiment regarding durability. This behavior is particularly useful when reviews contain detailed comments that provide information beyond numerical ratings.

Topic modeling also contributes to representing specific user interests when reviews emphasize particular product characteristics. By identifying latent themes in textual reviews, the model can determine repeatedly discussed aspects such as price, design, or performance. For example, when a user frequently refers to "battery life," the system can associate this preference with items whose reviews express positive sentiment toward that aspect. This mechanism supports recommendations aligned with interests expressed in textual feedback.

The model's effectiveness is also influenced by the weighting configuration obtained through grid search. The final combination assigns $\alpha = 0.3062$ to collaborative filtering, $\beta = 0.5870$ to sentiment analysis, and $\gamma = 0.1068$ to topic modeling. These values represent the contribution of each component to the hybrid score, with sentiment analysis receiving the largest weight. This balance allows the model to combine collaborative information, emotional signals, and topic-related preferences.

The prediction-error measures provide additional evidence of the model's effectiveness. The proposed approach achieves an RMSE of 0.225 and an MAE of 0.203, indicating accurate estimation of user ratings. These values show that the hybrid method improves not only ranking-oriented recommendation quality but also numerical prediction of user preferences. Overall, the results support its use in scenarios requiring relevant recommendations and reliable rating estimates.

5. CONCLUSION

The study proposed a hybrid recommendation approach that utilizes a combination of different techniques. The hybrid recommendation framework presented here employs sentiment analysis and topic modeling along with collaborative filtering to produce more personalized recommendations and greater accuracy in recommending products. The proposed sentiment-topic SVD++ model applies DistilBERT to perform sentiment extraction and utilizes BERTopic for discovering topics. By combining these methods, the model captures the emotional tone as well as the context of the user's review. The results from experimentation were successful in validating that the addition of textual signals resulted in recommendations that were not only more accurate but contextualized compared to traditional methods of collaborative filtering. The flexibility of the proposed framework allows for extension into other areas besides e-commerce (e.g., streaming services, social media platforms). The next steps in this research will be to improve interpretability of the models, incorporate real-time user feedback, and test model generalizability on additional datasets.

ACKNOWLEDGMENTS

The authors express their gratitude for the support received from the Laboratory of Information Technologies and Modeling at the Faculty of Sciences Ben M'sik, Hassan II University, Casablanca, Morocco, which provided the necessary facilities for conducting this research. Additionally, the authors acknowledge the support of the National Center for Scientific and Technical Research or CNRST under the "Ph.D.-AS-sociate scholarship – PASS" program. The authors acknowledge the use of artificial intelligence tools, specifically OpenAI's ChatGPT, to assist in improving the language clarity, grammar, and overall readability of the manuscript. All conceptual, analytical, and experimental content was developed by the authors.

FUNDING INFORMATION

The authors received no specific financial support or external funding for completion of this research.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
El Mehdi Lghaouch	✓	✓	✓						✓					
Soumaya Ounacer	✓			✓						✓				
Soufiane Ardchir				✓	✓					✓				
Mohamed Azzouazi		✓							✓			✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal Analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject Administration
Fu : **Fu**nding Acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest related to this work.

INFORMED CONSENT

Informed consent is not applicable, as the study does not involve human participants or identifiable personal data.

ETHICAL APPROVAL

Ethical approval is not applicable, as the research does not include experiments involving humans or animals. The work was conducted solely on computational models, publicly available datasets, or non-sensitive data that did not require intervention with living organisms. Therefore, ethical approval from an institutional review board or animal ethics committee was not necessary for this study.

DATA AVAILABILITY

This study uses the publicly available from Amazon fine food reviews dataset, accessible on Kaggle at https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews.

REFERENCES

[1] P. Resnick and H. R. Varian, "Recommender systems," *Communications of the ACM*, vol. 40, no. 3, pp. 56–58, Mar. 1997, doi: 10.1145/245108.245121.

[2] P. M. Alamdari, N. J. Navimipour, M. Hosseinzadeh, A. A. Safaei, and A. Darwesh, "A systematic study on the recommender systems in the e-commerce," *IEEE Access*, vol. 8, pp. 115694–115716, 2020, doi: 10.1109/ACCESS.2020.3002803.

[3] L. V. Nguyen, "Classifications, evaluation metrics, datasets, and domains in recommendation services: a survey," *International Journal of Hybrid Intelligent Systems*, vol. 20, no. 2, pp. 85–100, Jun. 2024, doi: 10.3233/HIS-240003.

[4] M. O. Ayemowa, R. Ibrahim, and Y. A. Bena, "A systematic review of the impact of auxiliary information on recommender systems," *IEEE Access*, vol. 12, pp. 139524–139539, 2024, doi: 10.1109/ACCESS.2024.3462750.

[5] X. Yu, F. Jiang, J. Du, and D. Gong, "A cross-domain collaborative filtering algorithm with expanding user and item features via the latent factor space of auxiliary domains," *Pattern Recognition*, vol. 94, pp. 96–109, Oct. 2019, doi: 10.1016/j.patcog.2019.05.030.

[6] S. M. Al-Ghuribi and S. A. M. Noah, "Multi-criteria review-based recommender system — the state of the art," *IEEE Access*, vol. 7, pp. 169446–169468, 2019, doi: 10.1109/ACCESS.2019.2954861.

[7] M. Loukili, F. Messaoudi, and M. El Ghazi, "Sentiment analysis of product reviews for e-commerce recommendation based on machine learning," *International Journal of Advances in Soft Computing and its Applications*, vol. 15, no. 1, pp. 1–13, 2023, doi: 10.15849/IJASCA.230320.01.

[8] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1–167, May 2012, doi: 10.2200/S00416ED1V01Y201204HLT016.

[9] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends® in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, Jul. 2008, doi: 10.1561/1500000011.

[10] D. M. Blei, "Probabilistic topic models," *Communications of the ACM*, vol. 55, no. 4, pp. 77–84, Apr. 2012, doi: 10.1145/2133806.2133826.

[11] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: a survey," *WIREs Data Mining and Knowledge Discovery*, vol. 8, no. 4, Jul. 2018, doi: 10.1002/widm.1253.

[12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.

[13] X. Li, L. Bing, W. Zhang, and W. Lam, "Exploiting BERT for end-to-end aspect-based sentiment analysis," in *5th Workshop on Noisy User-generated Text (W-NUT 2019)*, 2019, pp. 34–41, doi: 10.18653/v1/D19-5505.

[14] A. Al-Ajlan and N. Alshareef, "Recommender system for Arabic content using sentiment analysis of user reviews," *Electronics*, vol. 12, no. 13, Jun. 2023, doi: 10.3390/electronics12132785.

[15] L. Gao and J. Li, "E-commerce personalized recommendation model based on semantic sentiment," *Mobile Information Systems*, vol. 2022, pp. 1–10, Aug. 2022, doi: 10.1155/2022/7246802.

[16] S. Kumar, K. De, and P. P. Roy, "Movie recommendation system using sentiment analysis from microblogging data," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 4, pp. 915–923, Aug. 2020, doi: 10.1109/TCSS.2020.2993585.

[17] Q. Liang, S. Ranganathan, K. Wang, and X. Deng, "JST-RR model: joint modeling of ratings and reviews in sentiment-topic prediction," *Technometrics*, vol. 65, no. 1, pp. 57–69, Jan. 2023, doi: 10.1080/00401706.2022.2063187.

[18] H. Wang, Z. Hong, and M. Hong, "Research on product recommendation based on matrix factorization models fusing user reviews," *Applied Soft Computing*, vol. 123, Jul. 2022, doi: 10.1016/j.asoc.2022.108971.

[19] N. Liu and J. Zhao, "Recommendation system based on deep sentiment analysis and matrix factorization," *IEEE Access*, vol. 11, pp. 16994–17001, 2023, doi: 10.1109/ACCESS.2023.3246060.

[20] C. N. Dang, M. N. M. -García, and F. D. L. Prieta, "Using hybrid deep learning models of sentiment analysis and item genres in recommender systems for streaming services," *Electronics*, vol. 10, no. 20, Oct. 2021, doi: 10.3390/electronics10202459.

[21] H. Kanwal, M. Assam, A. Jabbar, S. Khan, and Kalimullah, "Convolutional neural network and topic modeling based hybrid recommender system," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 7, 2020, doi: 10.14569/IJACSA.2020.0110775.

[22] J. Xie, M. F. Zolkipli, M. F. M. Mohsin, T. Qiu, F. Liu, and M. Hou, "Movie recommendation algorithm based on sentiment analysis and topic model," in *2024 7th International Conference on Internet Applications, Protocols, and Services (NETAPPS)*, Nov. 2024, pp. 1–7, doi: 10.1109/NETAPPS63333.2024.10823455.

[23] A. R. Nair, R. P. Singh, D. Gupta, and P. Kumar, "Evaluating the impact of text data augmentation on text classification tasks using DistilBERT," *Procedia Computer Science*, vol. 235, pp. 102–111, 2024, doi: 10.1016/j.procs.2024.04.013.

[24] D. M. Blei and J. D. Lafferty, "Topic models," in *Text Mining: Classification, Clustering, and Applications*, Boca Raton, Florida: Chapman & Hall/CRC, 2009, pp. 71–93.

[25] M. Grootendorst, "BERTopic: neural topic modeling with a class-based TF-IDF procedure," Mar. 2022, *arXiv:2203.05794*.

BIOGRAPHIES OF AUTHORS



El Mehdi Lghaouch received his bachelor's degree in Mathematics and Computer Science from the Faculty of Sciences Ben M'sik, Hassan II University, Casablanca, Morocco, in 2021. He obtained his master's degree in Data Science from the same faculty in 2023. He is currently pursuing a Ph.D. in Data Science at the Faculty of Sciences Ben M'sik, Hassan II University. His research interests include data science, machine learning, and recommender systems. He can be contacted at email: elmehdi.lghaouch-etu@etu.univh2c.ma.



Soumaya Ounacer received her bachelor's degree in Mathematics and Computer Science from the Faculty of Sciences Ben M'sik, Hassan II University, Casablanca, Morocco, in 2008. She obtained her master's degree in Computer Engineering from ENSIAS, Mohammed V University, Rabat, in 2010, and her doctorate in Big Data and Machine Learning from Faculty of Sciences Ben M'sik in 2020. She has been an associate professor at Faculty of Sciences Ben M'sik, Hassan II University, since 2018, and is a member of the LTIM Laboratory. She has authored over 20 scientific papers and supervised 4 theses. Her research interests include machine learning, data science, big data, education, and mathematics. She can be contacted at email: soumayaounacer@gmail.com.



Soufiane Ardchir received his bachelor's degree in Computer Science from the Faculty of Sciences Ben M'sik, Hassan II University, Casablanca, in 2003. He obtained his master's degree in Networks, Telecommunications, and Multimedia from ENSIAS, Mohammed V University, Rabat, in 2007, and his doctorate in Computer Engineering in 2020. He has been an associate professor at ENCG, Hassan II University, Casablanca, since 2020. He is affiliated with the LTIM Laboratory. He has published 17 scientific papers and supervised 2 theses. His research interests include artificial intelligence, machine learning, and deep learning. He can be contacted at email: s.ardchir@encgcasa.ma.



Mohamed Azzouazi received his bachelor's degree in Mathematics and Applications from the Faculty of Sciences Ben M'sik, Hassan II University, Casablanca, Morocco, in 1990. He obtained a DESA in Information Systems from ENSIAS, Mohammed V University, Rabat, in 1994, and his doctorate in Big Data Analytics from Faculty of Sciences Ben M'sik in 2017. He has been an associate professor at Faculty of Sciences Ben M'sik, Hassan II University, since 2000, and is a member of the LTIM Laboratory. His research interests include artificial intelligence, machine learning, deep learning, and big data solutions. He can be contacted at email: azouazii@gmail.com.