BERT-enhanced sentiment analysis for personalized e-commerce recommendations

Ikram Karabila[1] · Nossayba Darraz[1] · Anas EL-Ansari[2] · Nabil Alami[3] · Mostafa EL Mallahi[1]

Abstract

Recommendation systems (RS) play a crucial role in enhancing conversion rates in e-commerce by offering personalized product recommendations based on customer preferences. However, traditional RS heavily rely on numerical ratings, which might not fully capture the subtle nuances of user preferences. To overcome this limitation, the integration of textual data, such as reviews using sentiment analysis (SA), has gained considerable significance. Nevertheless, effectively analyzing and comprehending unstructured review data presents its own set of challenges. In this work, we propose a novel RS that synergizes collaborative filtering with sentiment analysis to deliver precise and individualized recommendations. Our approach encompasses three main steps: (1) Developing a BERT fine-tuned model for accurate sentiment classification, (2) Creating a hybrid collaborative filtering-based Recommendation Model, and (3) Improving the product selection process in the RS using BERT insights for enhanced recommendation accuracy in the e-commerce domain. Notably, our SA model exhibits remarkable accuracy, achieving 91%, and outperforming state-of-the-art models on a benchmark dataset. Through extensive experimentation and evaluation, we demonstrate that our method significantly improves the accuracy and personalization of the RS, thereby providing customers with a tailored and reliable recommendation service in the e-commerce domain.

Keywords BERT · Collaborative filtering · Recommendation system · Sentiment analysis · Ensemble learning

1 Introduction

The proliferation of online information has given rise to a formidable challenge for users who must navigate through an abundance of options across diverse services like products, hotels, and restaurants. This phenomenon, termed as information overload, can significantly intricate decision-making [1]. In response, recommendation systems (RS) have emerged to

---

✉ Ikram Karabila
  ikram.karabila@usmba.ac.ma

Extended author information available on the last page of the article

filter and deliver personalized suggestions based on individual preferences [2]. These systems aim to alleviate the time users spend in information search and offer items aligned with their interests, ultimately enhancing the quality of information access services. The applicability of RS spans various sectors including social networking (Facebook), music streaming (Spotify), and e-commerce (Amazon).

The development of E-commerce platforms has underscored the indispensability of recommender systems in assisting customers to discover products that match their specific requirements and preferences [3].

Multiple techniques exist for item recommendations within RS, with the prominent methods being content-based filtering (CBF), collaborative filtering (CF), and hybrid approaches [4]. CF techniques, widely adopted in numerous online shopping platforms, tailors recommendations to a user based on the preferences of similar users [5]. CF can be categorized into memory-based and model-based systems. model-based creates models to predict user preferences, while memory-based relies on user or item similarities for recommendations [6]. Among CF techniques are Item-based and User-based methods [7]. On the contrary, CBF assesses the semantic substance of items [8]. A final approach, the hybrid [9], Integrates diverse algorithms into a unified RS.

Nevertheless, traditional recommender systems often rely on a single standard rating for recommendations, which might not provide comprehensive insights into user behavior [10]. Furthermore, CF encounters challenges like Sparsity and gray-sheep [11], diminishing its reliability in some scenarios. This underscores the ongoing significance of RS as a pivotal research domain, necessitating innovative solutions to enhance its efficacy.

Notably, customer reviews have substantially influenced consumer decisions, leading to a surge in online reviews. These reviews mirror individual experiences with services or products, and sentiment analysis techniques can be employed to infer customer opinions and emotions on diverse subjects.

The primary aim of Sentiment Analysis is to ascertain the emotional sentiment or attitude conveyed within user-generated text pertaining to a particular entity or subject [12]. This objective is achieved through the automated identification and extraction of relevant information concerning the subject under discussion, followed by an evaluation of whether the language utilized in the text portrays a sentiment that is positive, negative or neutral in nature. SA can be executed at various levels of data extraction [13], including document, sentence, and aspect levels. The resolution of the SA problem encompasses three key methodologies [14]: Lexicon-based, Machine Learning-based, and Hybrid methods. The earliest techniques employed for SA were lexicon-based methods, which rely on established linguistic rules and lexicons, and can be categorized into two principal categories: dictionary-based and corpus-based strategies [15]. Machine Learning (ML) driven methodologies encompass both traditional and deep learning (DL) techniques [16]. In a complementary manner, hybrid methodologies amalgamate lexicon-based approaches with machine learning techniques [12]. The domain of DL has exhibited superior efficacy compared to conventional methodologies in the realm of SA [17]. Notably, DL models, such as Recurrent Neural Networks (RNNs), Deep Neural Networks (DNNs), and Convolutional Neural Networks (CNNs), find utility in the task of sentiment classification.

Integrating SA methods into recommendation systems is a way to improve the performance and quality of recommendations. By incorporating sentiment analysis into recommendation systems, it becomes possible to gather additional information about user feedback and preferences on items. This integration can significantly enhance the performance of the system by providing a deeper understanding of user sentiments and preferences.

In this study, we leverage the capabilities of fine-tuned BERT (Bidirectional Encoder Representations from Transformers), a potent language model pre-trained on unlabeled data and further refined on labeled data, for sentiment analysis. Fine-tuning BERT for sentiment classification enhances its accuracy in analyzing sentiments within textual data. BERT's fine-tuning process empowers it to learn from labeled data, enabling it to understand intricate sentiment patterns. This surpasses traditional models, providing more reliable sentiment analysis. We integrate BERT-based SA into our recommender system, aiming to improve recommendation precision. We conduct experiments on a benchmark dataset to evaluate our approach. The results show that by combining BERT-based sentiment analysis with our recommender system, we achieve superior performance compared to state-of-the-art models.

The organization of this paper is outlined as follows: Section 2 furnishes an extensive review of preceding research works. Section 3 expounds upon the approach and methodologies adopted for the creation and refinement of our recommender system. Section 4 scrutinizes and presents an in-depth analysis of the outcomes derived from the experimental endeavors conducted. Delving into the interpretive domain, Section 5 engages in a comprehensive discussion of the acquired results. Lastly, Section 6 culminates the paper by furnishing a conclusive summary and accentuating prospective avenues for future research exploration.

2 Related work

The academic community is currently focused on collaborative filtering research. Traditional methods, however, have limitations such as sparse data for many items (data sparsity), difficulty in providing recommendations for new users or those with little interaction (cold start), and the inability to make recommendations for users with unique interests and preferences (grey sheep) that can make it difficult to capture accurate user profiles. In light of these limitations, researchers are currently working on developing new ways to improve conventional CF methods to address these limitations. One of the ways in which collaborative filtering techniques are being improved is by combining them with sentiment analysis.

For example, in a work by Rayan et al. [18], a novel method was introduced that integrates content-based and collaborative filtering to deliver personalized recommendations for personal well-being services. The researchers aimed to address the challenges of traditional CF, especially the cold-start problem. Through experimentation on a data of personal well-being services, the proposed technique demonstrated superiority over conventional collaborative filtering methods. Similarly, in another work [19], the authors proposed a hybrid recommender system that incorporates content-based and CF while incorporating sentiment analysis based on the Naive Bayes method. This system was developed using microblogging data to improve the overall recommender system.

In the study of Osman et al. [20], the authors suggested a RS that utilizes CF methods and incorporates contextual SA to enhance the personalization and accuracy of recommendations. They developed a system that takes into consideration the context of a user's feedback and preferences, and used it to generate more relevant recommendations. The authors aimed to address the drawbacks of conventional collaborative filtering methods, which ignore the context of a user's feedback, by integrating a contextual sentiment model in the RS.

In the study [21], the authors aimed to resolve the problem of cold-start in RS. They suggested a method that integrates SA of textual data from online communities (Facebook and Twitter) to overcome this problem. ML methods such as Naive Bayes (NB) and Support Vector Machine (SVM) were utilized to conduct SA on the data, which then provided

additional information about new items or users. This information was utilized to enhance the personalization and accuracy of recommendations by incorporating SA of social network data in the recommendation process.

Ziani et al. [22] have proposed a recommendation system that uses user-based CF and combines it with sentiment analysis to generate highly accurate recommendations. They use a semi-supervised SVM as a sentiment analysis method to analyze the sentiment of reviews and ratings offered by users. The combination of CF and sentiment analysis allows the system to take into consideration both the preferences of users and the sentiment of their feedback, resulting in more personalized and accurate recommendations. Additionally, the use of a multilingual approach allows the system to cater to users who speak different languages, making it more inclusive and accessible to a wider range of users.

The authors of study [23] suggested a technique that combines collaborative filtering and SA to enhance the performance of a RS for groups of users. They utilized Linear Support Vector Classification (LSVC) and Naive Bayes Multinomial (NBM) classification methods for SA and Singular Value Decomposition (SVD) to enhance the scalability of the RS. The results of the research show that the suggested technique enhances the performance of the RS, providing more personalized and precise recommendations.

The authors in [24] proposed a recommender system, which integrates item-based collaborative filtering with sentiment analysis using logistic regression. This combination aims to improve the recommendation process by enhancing the suggestions and excluding items with negative sentiment.

In the study conducted by Nabil et al. [25], they created a novel recommendation system that relies on SA applied to tweets. The research aimed to improve the accuracy of recommendations by utilizing ML techniques (Naive Bayes and SVM), to categorize tweets into negative, positive, or neutral sentiments. The sentiment information extracted from the tweets was then utilized to provide more tailored and precise recommendations based on the interests and preferences of users. To address scalability challenges, the authors integrated Singular Value Decomposition into the recommendation system. The findings of the study indicated that the suggested approach significantly enhanced the performance of the recommender system, offering more personalized and accurate recommendations to users based on the SA.

The authors of the research [26] have introduced an approach to augment the performance of CF methods. They have incorporated lexicon-based SA to capture the emotional content within recommended items, leading to more accurate predictions of users' preferences.

The assimilation of SA has proven to yield promising outcomes in augmenting the efficacy of recommender systems. In Table 1, a comparative analysis is presented, juxtaposing our proposed approach with the most relevant studies that integrate SA in RS. To underscore the novelty of our work and its principal contributions, we elucidate distinctive aspects in our methodology that advance the current state of research in recommender systems.

The studies presented in the table differ in terms of the combination of techniques used, the data sources, and the goals of the research. However, all the studies aim to enhance the performance and personalization of recommendations. And each study has its own set of advantages and limitations.

With a focus on bridging the gap between advanced SA and CF methodologies, our research aims to redefine the contours of personalized recommendations. The following contributions encapsulate the unique elements that set our study apart in the realm of RS :

- Innovative Sentiment Analysis Model: Our study introduces a state-of-the-art sentiment analysis model based on contextual features extracted using BERT. Through meticulous

Table 1 Comparative analysis of our proposed approach with the most relevant studies that integrate SA in RS

Study	RS technique	SA Method	Advantages	Limitations
Contratres et al. [21]	CF	SVM, NB	- Addresses the problem of cold-start in RS. - Uses ML methods for automated sentiment analysis.	- Focuses only on SVM and NB, other ML or DL methods could be considered.
Ziani et al. [22]	CF (User-based)	Semi-supervised SVM	- Semi-supervised SVM is a flexible approach that can handle different types of input data and feature representations. - Leveraging both labeled and unlabeled data for sentiment analysis improves performance.	- The study only focuses on user-based CF and does not explore other CF algorithms, such as item-based or model-based CF, which may provide better performance in certain scenarios.
Sallam et al. [26]	CF (Item-based, SVD)	Lexicon-based	- Lexicon-based sentiment analysis is cost-effective and efficient for emotional content incorporation. - Simple and easy to integrate into existing CF algorithms.	- The study only focuses on lexicon-based sentiment analysis. Other sentiment analysis techniques such as machine learning-based approaches may also be useful in improving recommendation quality.
Our proposal	Hybrid CF (User-based, Item-based)	BERT Fine-tuned	BERT captures contextual and semantic information from text, enabling precise sentiment classification. - Combining item-based and user-based CF methods into a single hybrid approach that captures the advantages of both methods. - Contextual understanding of user feedback improves recommendation accuracy and personalization.	- BERT's computational complexity and resource requirements can be challenging for certain environments.

fine-tuning, our BERT model surpasses traditional deep-learning models like BiLSTM and BiGRU, establishing its superiority in sentiment classification tasks. This novel architecture significantly advances the understanding and representation of sentiments, contributing a unique dimension to the field.

- Proprietary Fusion Method: Another key contribution lies in our novel fusion method that merges predictions from User-Based and Item-Based CF approaches, employing Support Vector Regression. Unlike conventional methods, our fusion technique strategically leverages the strengths of both CF approaches, enhancing precision and personalization in recommendations. This innovative fusion method distinguishes our work and offers a fresh perspective on optimizing recommendation systems.

- Integration of Sentiment Analysis and Hybrid CF: Our study seamlessly integrates BERT-based sentiment analysis with a hybrid CF approach, specifically tailored for the

e-commerce domain. This integration leads to substantial enhancements in recommendation precision and personalization. By harmonizing sentiment analysis with CF, our approach provides users with a more individualized and reliable recommendation service, representing a novel and impactful contribution to recommender system literature.

- Thorough Evaluation and Superiority: To establish the robustness of our contributions, we conducted comprehensive testing, evaluating various well-known algorithms and techniques. Our results unequivocally demonstrate the superiority of our method in terms of prediction accuracy, surpassing alternative collaborative filtering methods. These findings underscore the practical significance of our approach in generating precise and relevant recommendations for e-commerce users.

3 Methodology

This research proposes a recommender system that utilizes sentiment classification based on BERT. The system predicts the sentiment of user reviews as either negative or positive, and leverages this information to enhance the accuracy of user recommendations. By integrating SA into recommender system, it aims to improve the user experience and provide more personalized and accurate recommendations.

The proposed system comprises a comprehensive approach consisting of six key stages. Firstly, relevant information is gathered and selected for analysis. It's crucial to choose the right data sources to ensure the effectiveness of the subsequent stages. Subsequently, a data preprocessing phase is carried out to prepare the data for subsequent analysis. The core component of the system involves sentiment classification utilizing a BERT fine-tuning approach. This enables the accurate categorization of user sentiments as positive or negative. The next stage involves the creation of a recommendation system that leverages the insights extracted from BERT to select suitable products for recommendation. Lastly, the system's performance is evaluated using metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Figure 1 provides an illustrative depiction of the system's overall architecture.

3.1 Dataset

The data selection process was carefully considered, taking into account factors such as data availability and accessibility. For our study, we opted to utilize the Amazon Musical Instruments reviews dataset [27], which proved to be an excellent fit for our purposes. This dataset comprises a diverse range of reviews for various Music products offered on Amazon. It provides detailed information about the users and the Music items, including both numerical ratings and textual reviews. With a substantial amount of data containing 10,262 rows and 9 columns, this dataset offered an ideal resource for conducting our analysis and evaluation. Its large number of reviews for Music products enabled us to conduct a comprehensive evaluation of both our recommendation system and sentiment analysis model. The dataset's diverse range of reviews allowed us to accurately assess the effectiveness and performance of our models. Furthermore, the dataset's significant size and variety provided us with valuable insights and a robust validation of our approach in the context of Music products. Table 2 presents the column names along with a brief description for each.
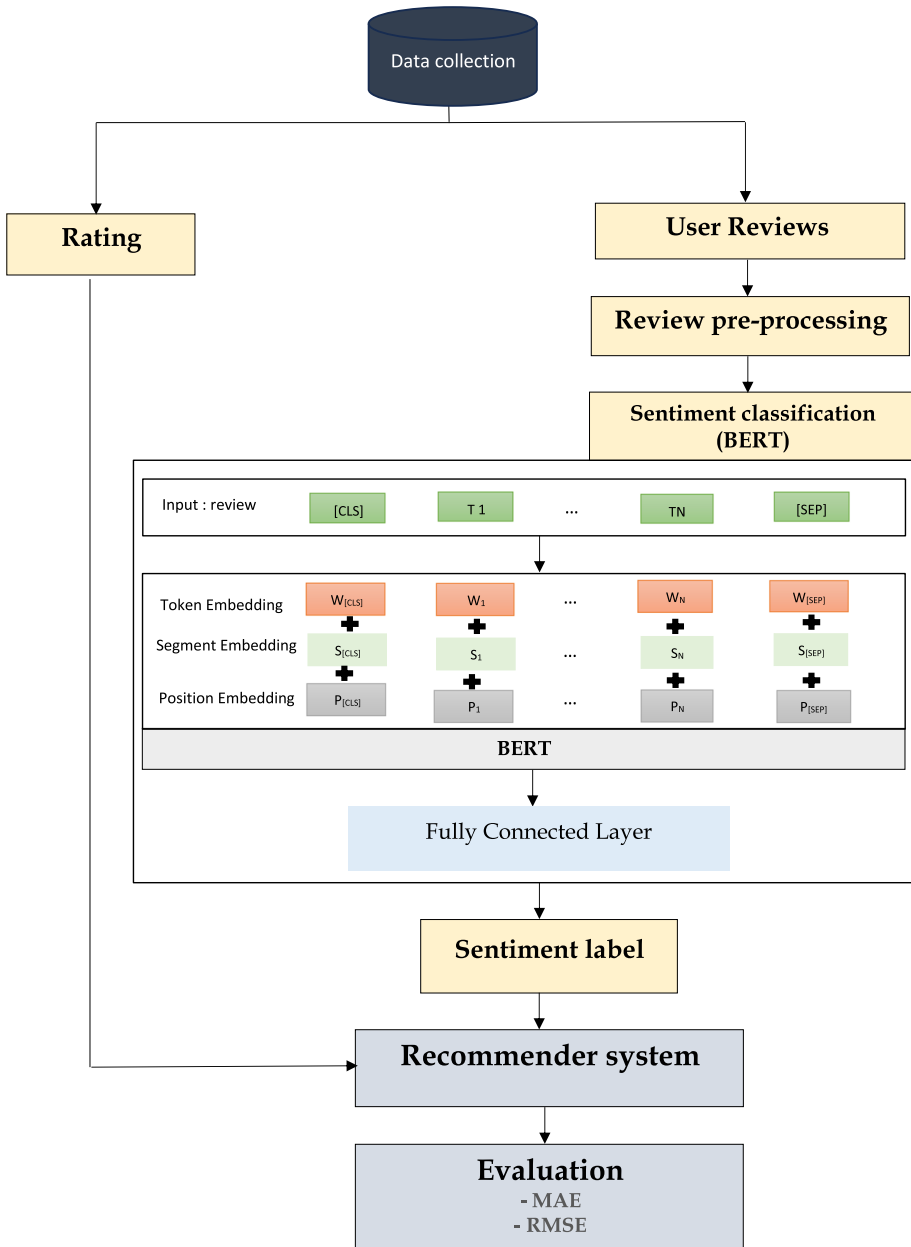
Fig. 1 Overall structure of the proposed system

3.2 Data preprocessing

During the preprocessing stage, we apply several steps to prepare the review text for input into our model. Firstly, we perform canonicalization by removing all digits, punctuation symbols,

Table 2 Description of the used dataset columns

Column name	Description
Asin	Unique product identifier
ReviewerID	Distinct identifier for the reviewer
ReviewerName	Full name of the reviewer
ReviewText	Content provided in the user's review
Helpful	Rating indicating the review's usefulness
Overall	User's rating of the product
Summary	Brief title or synopsis of the user's review
ReviewTime	Date and time of review submission
UnixReviewTime	Unix timestamp corresponding to the review's submission date and time

and accent marks, and converting the text to lowercase. This step helps standardize the text and reduce the impact of noise.

Next, we tokenize the text using the 'BertTokenizer', which is a specialized tokenizer designed specifically for BERT models.

Finally, the tokenized sequence includes the special tokens [SEP] and [CLS]. The [CLS] token is placed at the beginning of the sequence to represent the classification task, while the [SEP] indicates sentence boundaries. These special tokens are essential components of the BERT model and serve specific functions in tasks like sentiment analysis.

By following these preprocessing steps, we ensure that the review text is appropriately prepared and tokenized, ready to be fed into the BERT model for SA.

3.3 Sentiment analysis approach

Transformer networks have revolutionized the field of Natural Language Processing (NLP) by introducing a mechanism to effectively capture contextual information in large sequences of data, such as sentences or paragraphs. This architecture employs self-attention mechanisms to weigh the importance of different words in a sequence relative to each other, enabling the model to understand complex relationships and dependencies within the text. Transformer networks are built around the core idea of attention. Attention allows the model to focus on specific parts of a sequence based on their relevance to the current task. Transformer networks use self-attention mechanisms, where each word (or token) in a sequence is compared to all other words in the same sequence to determine its relative importance.

In particular, the self-attention mechanism calculates three vectors for each word: query, key, and value. These vectors are used to compute an attention score between the word in question and all other words in the sequence. The scores are then normalized to obtain attention weights, indicating the relative importance of other words with respect to the word in question. These weights are used to weight the values of the other words and obtain an enhanced contextual representation of the word in question [28]. Figure 2 illustrates the architecture of the self-attention mechanism.

One of the major advantages of attention is its ability to capture nonlinear relationships between words. In contrast to traditional sequential processing models, transformer networks excel at identifying long-range dependencies and intricate relationships within text. This
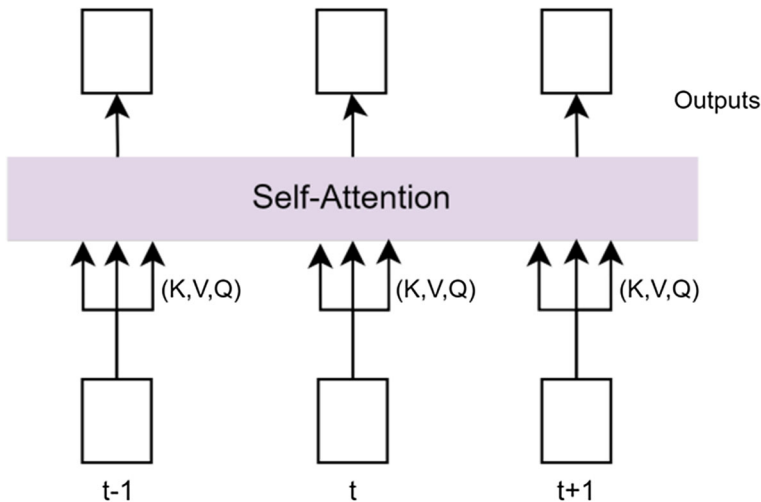
Fig. 2 Self-attention architecture

feature is particularly valuable for NLP tasks, where the meaning of a word can be significantly influenced by words located far away in the sequence.

In recent years, significant advancements in the field of NLP have been driven by the adoption of pre-trained models, including ELMo (Embeddings from Language Models) [30], BERT [29], and GPT (Generative Pretrained Transformer) [31]. These models have showcased substantial improvements in NLP performance, primarily owing to their ability to fine-tune after extensive pre-training on vast textual data. This approach has greatly enhanced the capabilities of NLP systems, resulting in more precise and effective text analysis.

BERT stands out as an exemplary NLP model designed to excel across a range of NLP tasks. It achieves this through fine-tuning with labeled text subsequent to pre-training on unlabeled text using deep bidirectional representations. Its exceptional performance has garnered significant recognition in the recent landscape of NLP [29].

In our study, we opted to utilize the BERT model, which empowers us to conduct comprehensive analysis of the contextual nuances within review text. This, in turn, results in more accurate predictions of sentiment attributes. BERT's bidirectional representation enables it to grasp intricate relationships and dependencies among words, capturing the intricate subtleties of language usage in expressing sentiments.

The decision to employ BERT for sentiment analysis within our study is well-grounded, bolstered by its established success in the field [32, 33]. Its impressive efficacy in enhancing the precision and effectiveness of text classification tasks underscores its potential for sentiment analysis. In a comprehensive comparative assessment carried out by the authors [34], which evaluated diverse state-of-the-art language models, BERT emerged as the preeminent choice for text classification tasks. The investigation underscored BERT's exceptional performance and highlighted its supremacy over other transformer-based models like ELMO and GPT.

The creators of BERT [29] introduced two models with different configurations: $BERT_{BASE}$ and $BERT_{LARGE}$. $BERT_{BASE}$ has a hidden layer size of 768, 12 layers, and 12 attention heads. On the other hand, $BERT_{LARGE}$ has a hidden layer size of 1024, 24 layers, and 16 attention heads.

The development process of our sentiment analysis model involved several steps to ensure robust performance. Firstly, to ensure a robust evaluation of our sentiment analysis model, we partitioned our dataset into distinct training and testing sets. The training set comprises 70% of the data, equivalent to 7,183 rows, while the testing set accounts for 30%, amounting to 3,078 rows. This division allows us to train our SA model on a portion of the data and assess its performance on unseen data, enabling a reliable evaluation of its effectiveness.

Due to computational constraints, we opted for a train/test split, balancing efficiency and robustness. We initially explored k-fold cross-validation with small k values and data portions. The results were not significantly different from the train/test split approach. This approach aligns with common practices and provides meaningful insights within our resource limitations. By training on the training set, our model learns patterns and relationships from the labeled data, enabling them to make predictions. On the other hand, the testing set serves as a benchmark for evaluating how well our model performs on new, unseen data.

To perform sentiment analysis, we utilized the $BERT_{BASE}$ model and fine-tuned it with our labeled dataset. This fine-tuning process involves training the BERT model on our specific dataset, enabling it to adapt to the sentiment classification task. By fine-tuning the BERT model with our dataset, we enhance its ability to capture the nuances and patterns in sentiment expressed in the text. This, in turn, improves its understanding of the context and sentiment of the given text data, resulting in more accurate sentiment classification predictions.

Through this approach, we aim to demonstrate the effectiveness of the BERT model for SA, particularly in the domain of e-commerce. Our goal is to enhance the performance of sentiment classification, especially within the context of recommendation systems. By fine-tuning the $BERT_{BASE}$ model with our labeled dataset (Fig. 6), we seek to leverage its pre-trained knowledge and language understanding capabilities to better understand and classify sentiments expressed in customer reviews related to various e-commerce products. The improved sentiment classification will then be integrated into our recommender system to deliver highly precise and personalized product recommendations based not only on numerical ratings but also on the sentiment expressed in user reviews.

The process of fine-tuning consists of three principal components, as illustrated in Fig. 3. Firstly, the review text undergoes tokenization using the 'BERT Tokenizer', which breaks the text into individual tokens, taking into account the subword-level structure of words. These tokens are then encoded using the pre-trained BERT model. This encoding process captures
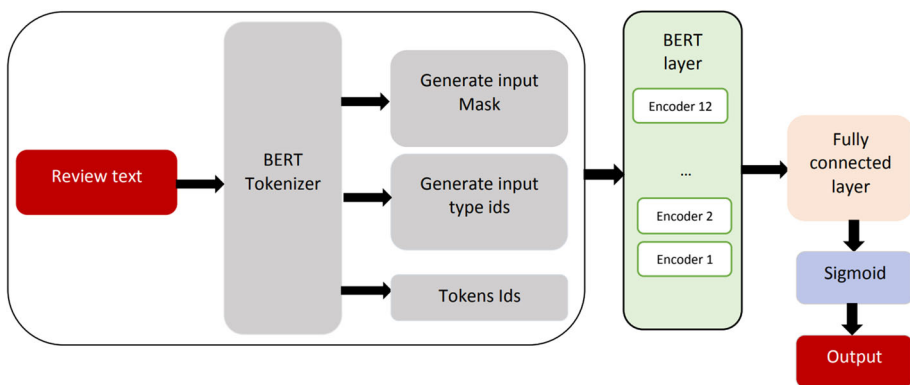

Fig. 3 BERT fine tuning process

the semantic meaning and contextual information of the text, resulting in the generation of the Final Hidden State. The Final Hidden State represents the encoded representation of each token in the review, encompassing important details such as word order, dependencies, and relationships. The second stage of fine-tuning involves introducing a fully connected layer. This layer receives the encoded representations from the Final Hidden State as input and performs sentiment prediction. By applying a set of weights and biases, the fully connected layer transforms the input into a suitable format for the binary classification task (positive or negative). Through this layer, the model learns to identify and leverage complex patterns and features extracted from the BERT model's representations. Following the fully connected layer, a sigmoid activation function is applied to generate the final sentiment predictions. The sigmoid function ensures that the output falls within the range of [0, 1], allowing the model to estimate the probability of the positive sentiment class. In our binary classification task, "0" represents the negative sentiment class, while "1" corresponds to the positive sentiment class. These labels serve as the target values for the model during the fine-tuning process, guiding it to learn the distinction between positive and negative sentiments effectively. Finally, during the training process, the model is optimized using the 'AdamW' optimization algorithm [35]. The 'cross-entropy' loss function is used to measure the discrepancy between predicted and actual sentiment values.

By fine-tuning the BERT model with the addition of the fully connected layer and the sigmoid activation function, the model effectively capitalizes on the pre-trained knowledge embedded within BERT. This enables the model to make accurate predictions for sentiment classification, effectively distinguishing between positive and negative sentiments.

Overall, this fine-tuning process leverages the power of BERT's pre-training and adapts it to the specific task of sentiment classification, resulting in enhanced performance and improved sentiment prediction accuracy.

In our study, our primary objective was to showcase the enhanced performance of the fine-tuned BERT model in the SA task. To achieve this, we conducted an extensive comparative analysis, pitting the fine-tuned BERT against other widely recognized algorithms. Among the algorithms subjected to evaluation were Bi-directional Long Short-Term Memory (BiLSTM) and Bi-directional Gated Recurrent Unit (BiGRU), both renowned for their prowess and success in SA tasks.

In the context of our study, prior to integration into deep learning models (BiLSTM and BiGRU), the textual data underwent a rigorous process of cleansing and lemmatization. This transformation was pivotal to facilitate the conversion of textual data into a suitable numerical format compatible with the employed deep learning models. Leveraging GloVe embeddings (Global Vectors for Word Representation), which have demonstrated superior performance compared to other methods like Word2vec across a spectrum of benchmark tasks, further elevates the quality of word representations in our study.

GloVe [36] stands as a robust and impactful word embedding technique, serving to translate words into real-valued vectors situated within a higher-dimensional space. The primary goal of GloVe is to capture the broader statistical characteristics of words, achieved by meticulously scrutinizing their co-occurrence frequencies across an expansive textual corpus. The underlying principle hinges on the understanding that words frequently appearing within akin contexts are prone to share semantic meanings or establish intricate relationships.

The models were evaluated using multiple classification metrics, including accuracy, Matthew's Correlation Coefficient, F1-score, and Area Under the Curve (AUC). These metrics offer a comprehensive understanding of the models' performance in sentiment classification tasks, covering different aspects such as overall performance, prediction accuracy, and the trade-off between false positive and true positive rates. By analyzing these metrics,

we can compare the effectiveness and efficiency of the different models and make informed decisions about their suitability for sentiment classification tasks.

3.4 Proposed recommendation system

Figure 4 depicts the development phases of our suggested system. These phases outline the key steps involved in building and implementing the recommendation system. The first step involves building a BERT-based sentiment analysis model. This model is trained on a labeled dataset to accurately classify the sentiment of textual reviews. The next step involves the creation of a recommendation system that utilizes hybrid collaborative filtering techniques, specifically combining both item-based and user-based methods. In the item-based CF method, the system identifies similarities between items based on the ratings given by users. It then recommends items that are similar to the ones that a user has previously rated highly. In the user-based CF approach, the system identifies users with similar rating patterns and preferences. It then recommends items that have been highly rated by similar users but not yet rated by the target user. Both item-based and user-based CF methods have proven to be successful in capturing user preferences and interests by analyzing historical interactions and ratings, enabling the RS to provide personalized and accurate recommendations for each user. However, CF-based RSs solely rely on numerical ratings, which may not fully capture the underlying sentiment and nuances expressed in user reviews. Integrating supplementary textual information from user reviews has the potential to considerably elevate the system's accuracy while offering a more holistic comprehension of user preferences.
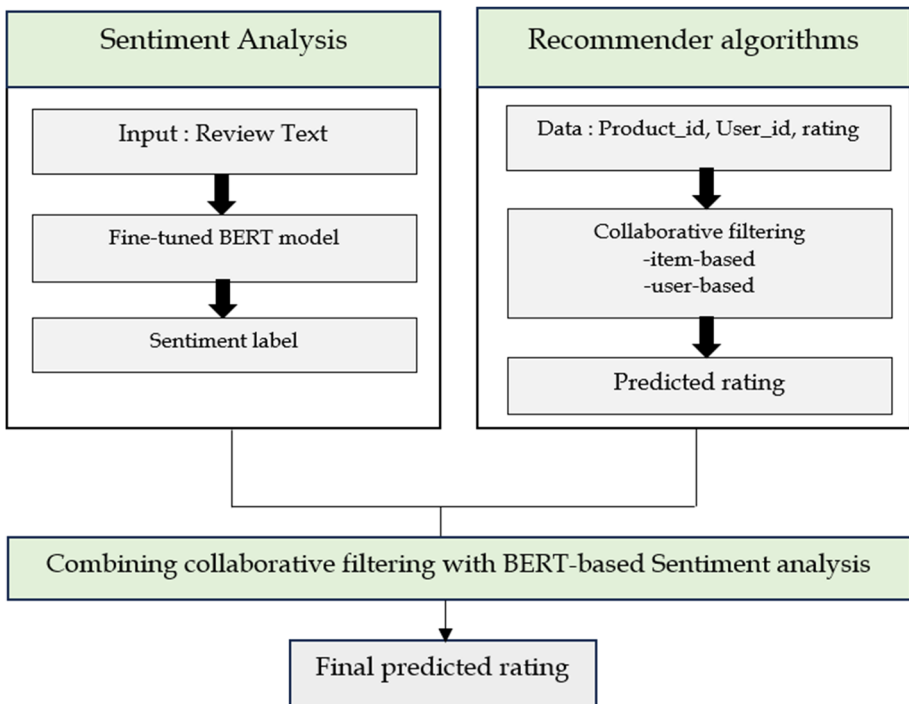

Fig. 4 Development phases for our proposed system

In the third step of our proposed approach, we aim to leverage the power of sentiment analysis by combining it with the hybrid CF approach. By integrating the sentiment analysis results into the collaborative filtering process, we can enrich the recommendation process with valuable insights from user sentiments. This integration allows the system to consider not only the ratings but also the emotional expressions, attitudes, and opinions conveyed in the reviews. As a result, the RS can make more informed and context-aware recommendations, ensuring that the recommended products not only align with the user's numerical ratings but also resonate with their sentiment preferences. This synergy between collaborative filtering and SA aims to enhance the accuracy and relevance of the recommendations, offering users a more personalized and satisfying recommendation service in the e-commerce domain.

- **Recommender system algorithms** The recommendation system in this work was developed utilizing a hybrid collaborative filtering technique based on SVR. This time the dataset was split into two portions: 80% (8,209 rows) for training the RS model and 20% (2,052 rows) for evaluating its performance. To prepare the training data, it was transformed into a matrix format, where each row represented a product and each column represented a user. This matrix captured the interactions between users and products, forming the basis for the CF model.

In collaborative filtering methods, when aiming to predict the ratings for products that a specific user has not yet rated, it is essential to employ a similarity metric to measure the likeness between users (in the case of user-based CF) or between items (in the case of item-based CF). To address this requirement, we opted for the cosine similarity metric, a widely used approach for this task (1) [37].

$$\text{sim}\,(v_a, v_b) = \frac{\sum_j x_{jb} \cdot x_{ja}}{\sqrt{\sum_j x_{jb}^2} \cdot \sqrt{\sum_j x_{ja}^2}} \qquad (1)$$

Where v_a and v_b are two vectors, and x_{ja} and x_{jb} represent the components of vectors v_a and v_b, respectively. This equation calculates the similarity between vectors v_a and v_b based on their weighted components.

Subsequently, a cosine similarity-based approach was employed to predict ratings. This process involved harnessing the ratings of the most akin users (in the context of the user-based technique) or items (in the context of the item-based technique) to estimate the rating that a specific user would assign to a particular item that lacked a previous rating.

$$P_{ij} = \bar{r}_i + \frac{\sum_u \text{sim}(u, i) \left(r_{ij} - \bar{r}_i\right)}{\sum_u |\,\text{sim}(u, i)\,|} \qquad (2)$$

where:

- P_{ij}: represents the predicted rating for item j by user i.
- r_{ij}: is the rating provided by user i for item j.
- $\bar{r}_i$: is the average rating given by user i.
- $\text{sim}(u, i)$: indicates the similarity between users u and i.

$$P_{ui} = \frac{\sum_j r_{uj} \cdot sim(i, j)}{\sum_j sim(i, j)} \qquad (3)$$

where:

- P_{ui}: represents the predicted rating for user u and item j.

- r_{ui}: is the actual rating that user u gave to item j.
- $sim(i, j)$: represents the similarity between items j and i.

In the context of user-based CF, the prediction of ratings is carried out using (2). Conversely, for item-based CF, the prediction process is governed by the utilization of (3).

In this research, we introduce a fusion technique that combines predictions from User-Based and Item-Based using Support Vector Regression models. The conceptual representation of this fusion approach is illustrated in Fig. 5.

The principle of fusing Item-Based and User-Based CF lies in combining the predictions generated by these two collaborative filtering approaches. In the context of CF, User-based offers recommendations based on similar user preferences, while Item-based focuses on similarities between items to generate recommendations. The key idea of this approach is to leverage the predictions from these two distinct approaches and fuse them using SVR model. By integrating the strengths of both collaborative filtering techniques, our fusion method aims to enhance the accuracy and personalization of recommendations, providing users with a more tailored and customized recommendation experience that aligns with their specific needs and preferences.

It is imperative to highlight that integrating sentiment analysis can enhance the accuracy and personalization of the recommendations. By taking into account the emotions conveyed in user reviews, the system can better understand user preferences and make more informed recommendations. The inclusion of SA can provide valuable insights into the sentiment associated with each item, allowing the system to recommend products that align not only with user ratings but also with their sentiment preference.

- **Enhancing recommendations: a fusion of collaborative filtering and BERT-based sentiment analysis**

  The final result is presented as a weighted combination of the predicted rating using hybrid CF technique (R) and the predicted sentiment score using the BERT-based model (S). The weight assigned to each term in the equation is denoted by a. The purpose of combining the predicted rating and sentiment score is to incorporate both the explicit
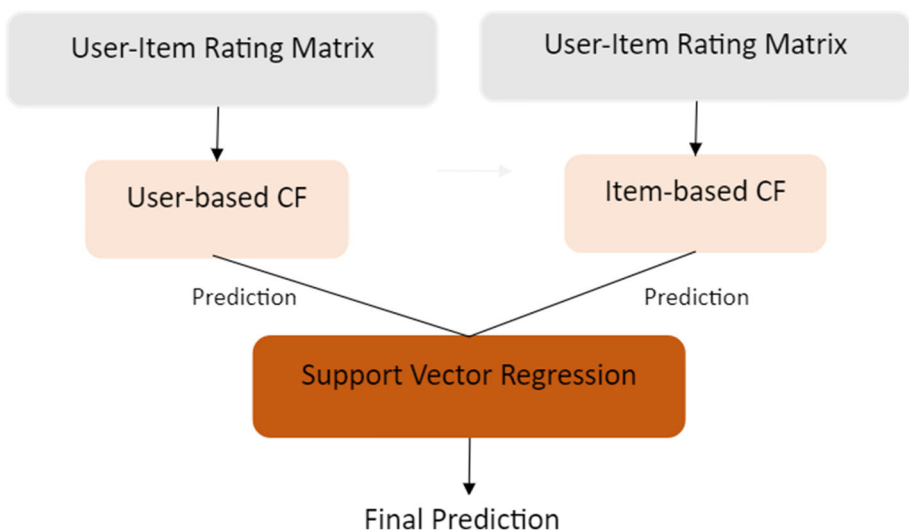

Fig. 5 The fusion of item-based and user-based approaches

user preferences captured by hybrid CF approach and the implicit sentiment expressed in the textual reviews. By assigning an appropriate weight to each term, we can balance the influence of user ratings and sentiment in the final recommendation.

$$H = a \times R + (1 - a) \times S \tag{4}$$

Equation (4) represents a linear combination of the two terms, where a determines the relative importance of the rating and sentiment score. A higher value of a places more emphasis on the predicted rating, while a lower value places more emphasis on the sentiment score.

By combining the hybrid approach of CF-based predicted rating and the BERT-based sentiment score, the proposed system aims to provide more comprehensive and accurate recommendations. It leverages both implicit and explicit user feedback to enhance the accuracy and relevance of the recommended items.

- **Evaluation:**
  The evaluation of the recommender system's effectiveness is accomplished by employing two assessment metrics: MAE and RMSE. These metrics yield values within the interval $[0, +\infty]$ and are inversely oriented, meaning that lower values indicate better performance.

$$MAE = \frac{\sum_{i=1}^{N} |(Predicted_{ij} - Actual_{ij})|}{N} \tag{5}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Predicted_{ij} - Actual_{ij})^2}{N}} \tag{6}$$

In all of the aforementioned equations, N denotes the total number of ratings within the test partition. $Predicted_{ij}$ signifies the predicted rating for item j and user i, while $Actual_{ij}$ corresponds to the actual real vote.

4 Results

To validate and assess the efficacy of our system, we undertook experiments employing two distinct approaches: the first entailed the absence of sentiment analysis, while the second involved BERT-based sentiment analysis.

In the first scenario, recommendations were generated through the application of recommender system techniques, with no consideration given to sentiment analysis. This approach served as a baseline for comparison, allowing us to assess the impact of incorporating our proposed sentiment analysis model. In the second scenario, we incorporated the sentiment analysis results of the reviews into the recommendation process. By leveraging BERT's sentiment analysis capabilities, we aimed to enhance the accuracy and relevance of the recommendations. This approach enabled the system to consider the sentiment expressed in the reviews when generating recommendations. By comparing the results of the two approaches, we were able to assess the effectiveness of integrating our proposed sentiment analysis model into the recommendation system.

	reviewerID	asin	overall	review_prepared	sentiment
1832	AOMEH9W6LHC4S	B0002E3CHC	5.0	long partial daddario stringsi guitar without ...	1
9616	ANADY67KNW4Z	B0079KQEOK	5.0	buy classical guitar realize need around neck ...	1
2721	A2IFKMKKXV8AVU	B0002GZLZQ	5.0	use one forgot mine love much get small take o...	1
1234	A3SGZO3N7MXKPV	B0002E1G5C	4.0	always keep around string change prefer bass v...	1
5266	ACVSCM9GWLNLS	B000P5LVSK	5.0	ever since see vid mark tremonti say hear lot ...	1
4224	AKGWRO6IC0VIE	B000B6DHB2	1.0	want like really idea amp portable cheap make ...	0

Fig. 6 Preprocessed data result

4.1 Data preparation

The initial phase of creating our recommendation system and sentiment model involved data preprocessing. Upon obtaining the data, our process commenced with the acquisition of the dataset and the elimination of irrelevant attributes that held no relevance to our analysis. Furthermore, we executed preprocessing procedures on the textual data, as detailed in Section 3.2, prior to the construction of our BERT Fine-Tuned model. This paved the way for the eventual development of our recommender system. Figure 6 displays the preprocessed data.

where:

– reviewerID: Unique identifier for the reviewer.
– Asin: Unique identifier for the product.
– Overall: Rating of the product given by the user.
– review_prepared: Text submitted in the review by the user after preprocessing.
– sentiment: the polarity of the review; negative (0) or Positive (1).

The dataset was divided into two distinct segments: a training set and a test set. The fundamental objective of the training set was to facilitate model training, enabling the acquisition of data patterns and relationships. Conversely, the test set was reserved to evaluate the model's effectiveness on fresh and autonomous data, previously unencountered during the training stage.

4.2 Sentiment model

An integral aspect of any deep learning solution involves the careful selection of hyperparameters. Hyperparameter selection and optimization can be done through two distinct approaches: manual and automatic. Both methods have their merits and trade-offs. While the manual approach allows for informed decisions based on a deep understanding of the model, automatic approaches can efficiently explore a broader range of possibilities but come with higher computational costs. In our study, we opted for a manual approach to select hyperparameter values for our proposed approach. We conducted experiments with different hyperparameter configurations, including those recommended, and assessed their impact on accuracy. By carefully testing and comparing various settings, we chose the hyperparameters that yielded the best accuracy for our sentiment analysis task. By taking this manual approach, we could leverage our understanding of the model and the specific requirements of the sentiment classification task, ultimately leading to a well-optimized system without incurring excessive computational overhead.

Table 3 Hyper-parameters used in the experiment for our proposed model (Fine-tuned BERT)

Hyper-parameter	Value
Activation function	Sigmoid
Epochs	10
The optimizer	AdamW
Dropout rate	0.2
Batch size	32
Learning rate	2e-5
max_seq_length	128
Loss function	Crossentropy

Table 3 shows the hyperparameters used in the fine-tuned BERT model for our approach in the experiment.

In Figs. 7 and 8, the outcomes of the sentiment classification experiments conducted with the BERT model are displayed. The figure illustrates that the sentiment classification model utilizing BERT performs admirably, achieving high accuracy, F1-Score, and reasonably strong MCC and AUC values. These results serve as validation for the effectiveness of the proposed SA model.

To demonstrate the superiority of our sentiment model, we conducted a comprehensive comparison with other popular algorithms known for their success in sentiment classification tasks, including Bi-LSTM and Bi-GRU. These algorithms have been widely used and proven effective in sentiment analysis. By comparing our results with these established methods, we can assess the performance of our suggested method.

Tables 4 and 5 show the hyperparameters used for BiLSTM and BiGRU models in the experiment.

Figure 9 and Table 6 illustrate the comparison between the proposed model and state-of-the-art methods in terms of metrics such as Accuracy, F1-Score, MCC, and AUC.

– BiLSTM: The incorporation of BiLSTM [38, 39] in sentiment analysis aims to enhance its performance. This is attributed to the BiLSTM model's ability to utilize two independent
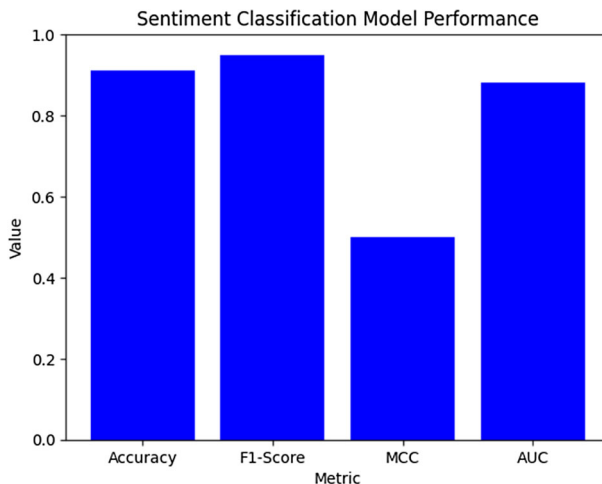
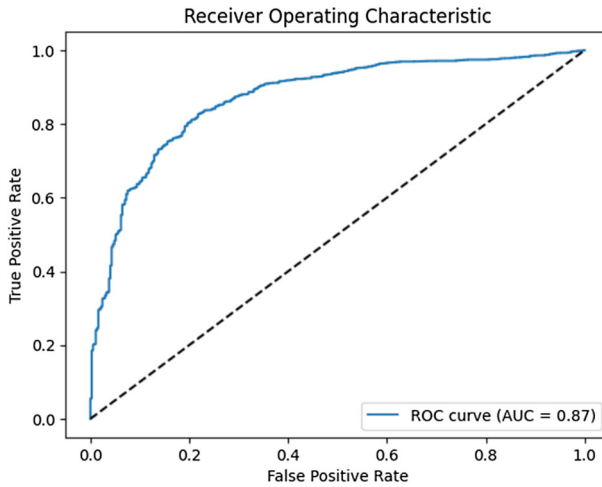
Fig. 7 Sentiment classification model performance

Receiver Operating Characteristic

Fig. 8 Performance evaluation using receiver operating characteristic (ROC)

Table 4 Hyper-parameters used in the experiment for BiLSTM models

Hyper-parameter	Value
Activation function	Sigmoid
Epochs	10
Dropout rate	0.5
Loss function	Crossentropy
The optimizer	Adam
Batch size	64

Table 5 Hyper-parameters used in the experiment for BiGRU models

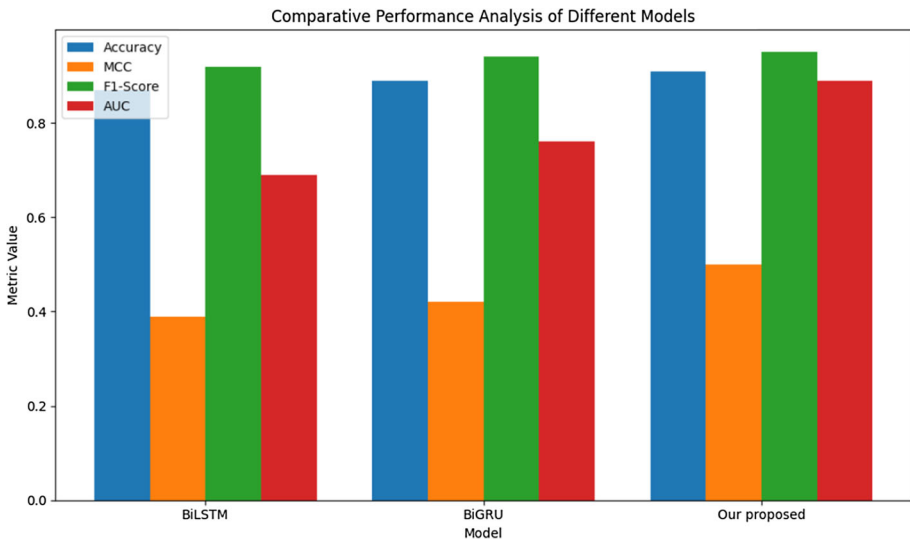| Hyper-parameter	Value
Activation function	Sigmoid
Epochs	10
Dropout rate	0.3
Loss function	Crossentropy
The optimizer	Adam
Batch size	32

Fig. 9 Comparative evaluation of the proposed model and baseline models on our dataset

LSTM models to capture contextual information from both right to left and left to right, subsequently merging them. In our study, we employed GloVe embeddings [36] to create feature vectors, which served as input for the BiLSTM model. The BiLSTM model achieved an Accuracy of 0.87, F1-Score of 0.92, AUC of 0.69, and an MCC of 0.39 on our dataset.

– BiGRU: The integration of Bi-GRU [40, 41] in sentiment analysis task aims to improve its performance similarly to BiLSTM. The Bi-GRU model, like BiLSTM, leverages bidirectional processing to capture contextual information from both directions, allowing it to effectively understand the text's underlying patterns. The principal difference between BiGRU and BiLSTM is that BiGRU uses GRU (Gated Recurrent Unit) cells instead of LSTM (Long Short-Term Memory) cells. GRU cells have fewer parameters, making them computationally more efficient than LSTM cells. In our experimentation, we used GloVe embeddings [36] to create feature vectors, which served as input for the Bi-GRU model. The Bi-GRU model achieved notable results, with an Accuracy of 0.89, F1-Score of 0.94, AUC of 0.76, and an MCC of 0.42 on our dataset.

According to the results depicted in Fig. 9 and Table 6, our proposed model outperformed other deep-learning models on the Amazon Music Reviews dataset. The proposed model achieved a maximum accuracy of 0.91, an F1-score of 0.95, MCC of 0.50, and AUC of 0.87, demonstrating its superiority compared to the alternative models.

Table 6 Comparative evaluation of the proposed model and baseline models on our dataset

Model	Accuracy	MCC	F1-Score	AUC
BiLSTM	0.87	0.39	0.92	0.69
BiGRU	0.89	0.42	0.94	0.76
Our proposal	0.91	0.50	0.95	0.87

Table 7 Comparative performance on the Amazon Musical Instruments reviews dataset

Authors	Method	Accuracy (%)	F1 score (%)
Elmurngi and Gherbi [42]	Naïve Bayes	80.61	–
	Decision tree	81.45	–
	Logistic regression	81.61	–
	Support vector machine	80.90	–
Tilloo et al. [43]	CNN	–	79.04
Our proposed model	BERT	91	95

Table 7 presents our proposed SA model's performance compared with other recently published articles that used the same dataset (Amazon Musical Instruments Reviews), focusing on F1-Score and accuracy as the evaluation metrics.

The proposed model, which utilizes BERT, achieved remarkable results with an accuracy of 91% and an F1 Score of 95% on a benchmark dataset. These outcomes clearly demonstrate that our proposed approach outperformed the other sentiment analysis models, showcasing its superior performance in accurately predicting sentiments within the given dataset. The high accuracy and F1 Score values serve as strong evidence for the effectiveness of the BERT-based approach, affirming its capability to capture intricate nuances of sentiments and deliver more precise predictions compared to alternative methods.

Given the compelling evaluation metrics, the BERT-based model was selected as the preferred model for sentiment analysis. This model will be utilized for sentiment prediction as a preliminary step before integrating it with recommendation approaches. This integration aims to amplify the overall performance and personalization capabilities of our recommender system.

4.3 Recommender system

During the evaluation of our recommendation model, we conducted a thorough series of experiments that encompassed user-based CF and item-based techniques, along with the fusion of these two approaches. We then compared the results obtained from traditional CF methods with those from the fusion CF methods, further improved by the integration of our proposed sentiment analysis model. The main objective behind these comparisons was to reveal the noticeable improvements resulting from this novel combination of methodologies.

Figures 10 and 11 visually depict the outcomes obtained from our analysis, centering on the RMSE and MAE measurements within the context of the Amazon Musical Instruments reviews dataset. These metrics serve as tangible benchmarks of prediction accuracy, constituting the fundamental criteria for assessing the efficacy of our proposed system.

Figure 10 offers a comprehensive comparative insight into MAE and RMSE values, encompassing both user-based and item-based CF methodologies. Moreover, the hybrid approach, which amalgamates the predictive outcomes of these two techniques, is included in the analysis. It is crucial to emphasize that this evaluation transpires prior to the incorporation of our sentiment analysis model. In specific numerical terms, the MAE values for the Item-Based CF and User-Based CF stand at 2.06 and 2.50, respectively. Concurrently, the RMSE values for Item-Based CF and User-Based CF are recorded as 2.60 and 2.80, respectively. Notably, upon the implementation of the hybrid approach, a significant decrease in
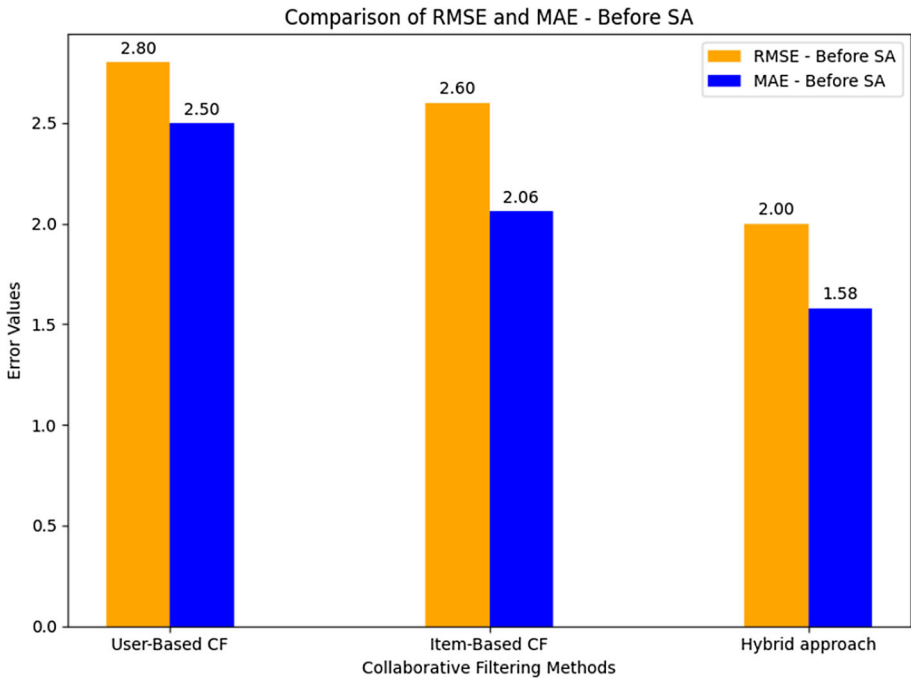
Fig. 10 Comparison of MAE and RMSE values before integrating sentiment analysis
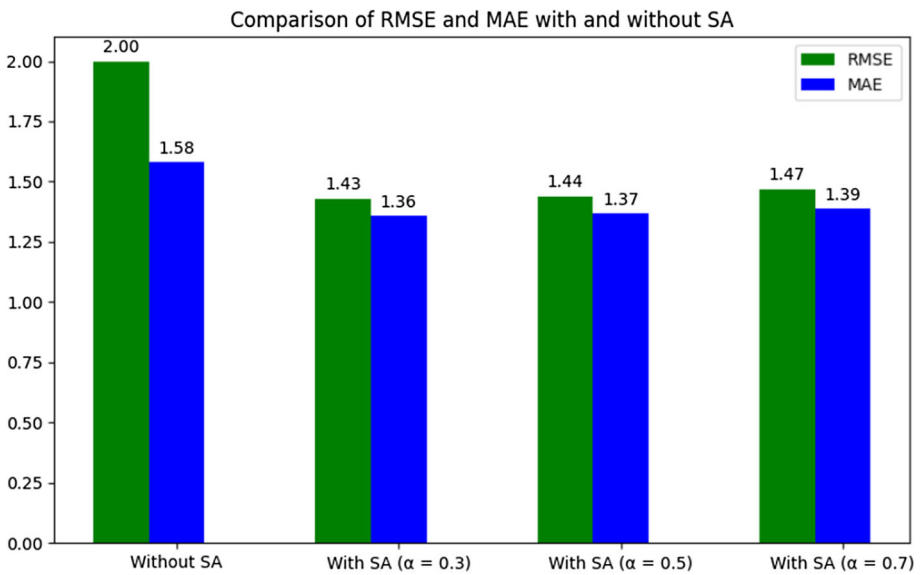

Fig. 11 Comparison of MAE and RMSE values with and without using the proposed sentiment analysis model

the MAE value to 1.58 is observed, coupled with a decline in the RMSE value to 2. This outcome underscores the substantial predictive enhancement achieved solely through the fusion methodology, demonstrating its efficacy in comparison to the individual collaborative filtering techniques.

Figure 11 offers a comprehensive comparative analysis that investigates the effectiveness of the hybrid collaborative filtering approach, considering both its utilization with and without the proposed sentiment analysis model. In the context of sentiment analysis integration, various weighting factors denoted by a (such as 0.3, 0.5, and 0.7) are introduced. This integration reveals significant insights: the MAE values exhibit a substantial reduction, indicative of enhanced predictive accuracy. For instance, when the weighting factor a is set at 0.3, the MAE values experience notable minimization, resulting in a decrease to 1.36 for the hybrid approach. A similar trend is observed for RMSE, which also experiences a decrease to 1.43. This unequivocally underscores the positive impact of integrating sentiment analysis into the hybrid CF framework. Importantly, this enhancement is uniformly manifested across the range of examined scenarios, emphasizing the consistent and advantageous impact of sentiment analysis on predictive accuracy.

In conclusion, the presented results collectively underscore the tangible benefits of integrating SA into the hybrid CF approach. The reduction in both RMSE and MAE values substantiates the notion that sentiment analysis serves as a valuable enhancement, leading to more accurate and reliable recommendations.

5 Discussion

The results obtained from our comprehensive comparison of the suggested model based on BERT with other state-of-the-art models commonly used for sentiment classification clearly demonstrate the superiority of BERT. The evaluation metrics, including an accuracy of 91%, and an F1-score of 95%, showcase the model's remarkable capability in accurately predicting sentiment. These impressive metrics highlight the performance of our suggested method and emphasize its potential in outperforming alternative methods for sentiment analysis.

The conducted experimental investigations on the amalgamation of the hybrid Collaborative Filtering approach with Sentiment Analysis have yielded compelling outcomes, signifying substantial and consistent enhancements in the evaluation metrics in comparison to the hybrid CF method devoid of SA. Across a range of weight coefficients (a), the hybrid collaborative filtering approach enriched by sentiment analysis consistently exhibits superior performance compared to its counterpart that lacks the sentiment analysis component. This discernible improvement can be attributed to the invaluable insights furnished by SA, which adeptly captures the emotional nuances embedded within user reviews. This augmentation elevates the system's comprehension of user inclinations and sentiments pertaining to specific items. Remarkably, the optimal or most favorable performance is observed with a weight coefficient value of $a = 0.3$. This outcome underscores the pivotal role played by the sentiment analysis model, underscoring its substantial influence on the recommendation process. This influence translates to more accurate and tailored suggestions, thereby accentuating the significance of sentiment analysis in enhancing the overall performance of the recommendation system.

Overall, our study provides valuable insights into the effectiveness of the suggested BERT-based model for sentiment analysis and its successful integration into the collaborative filtering recommender system. These findings contribute to the advancement of SA

and RSs, showcasing the potential for more accurate and personalized recommendations for users based on sentiment insights.

While our study presents promising results and novel contributions, it's important to acknowledge certain limitations and potential weaknesses that should be considered:

1. Computational Complexity: The integration of BERT-based sentiment analysis introduces computational challenges due to the model's large size and resource-intensive nature. Fine-tuning BERT and processing large volumes of text data require significant computational resources and time, which might limit the feasibility of real-time applications or implementation in resource-constrained environments.
2. Domain Specificity: Our study's findings are based on experiments conducted on a specific Domain, which might limit the generalizability of the results to other domains or datasets. Different domains may have varying sentiment expression patterns, affecting the performance of the sentiment analysis model.
3. Data Imbalance: Imbalanced sentiment distribution within the dataset can impact the model's ability to accurately predict minority sentiment classes. Addressing class imbalance, particularly in scenarios with a limited number of negative or positive samples, is a challenge that might affect the model's overall performance.

Addressing these issues and weaknesses requires a thoughtful approach. While computational complexity and domain specificity pose challenges, optimizing resources and employing domain adaptation techniques can enhance the model's practicality. For data imbalance, strategic data handling techniques can contribute to more balanced sentiment prediction and improve the overall robustness of the proposed system.

6 Conclusion & future work

In this research, we have proposed a comprehensive system architecture encompassing two core constituents: sentiment analysis of user reviews and an ensemble learning framework that seamlessly integrates collaborative filtering and sentiment analysis, culminating in personalized recommendations. Our approach harmoniously combines the efficacy of the hybrid collaborative filtering technique with the precision of BERT-based sentiment analysis, resulting in dependable and tailor-made suggestions for users.

The pivotal role of sentiment analysis is highlighted, as it underpins our investigation with the objective of categorizing user reviews into positive or negative sentiments. To achieve this, we meticulously constructed diverse models, striving to identify the most optimal approach amidst the inherent variability characteristic of such evaluations. The judicious selection of an appropriate classification model emerged as a critical focal point in our research journey.

The validation and rigor of our system were demonstrated through an exhaustive analysis, involving the rigorous comparison of diverse recommendation algorithms. This endeavor spanned across three distinctive scenarios: User-based CF, Item-based CF, and our hybrid CF method that adeptly blends both User-based and Item-based techniques. A noteworthy nuance of our evaluation was the exploration of the hybrid method's performance in two distinct scenarios: one driven solely by ratings for recommendations and the other synergistically fusing ratings and sentiments. The former configuration was established as a baseline to discern the discernible impact of sentiment analysis infused into the latter scenario.

Empirical findings unequivocally affirmed the potency of incorporating sentiment analysis within the hybrid CF approach, leading to a marked enhancement in performance metrics. This substantiates the empirical utility of harnessing sentiment information to augment and

refine the recommendation process, underscoring a transformative effect on the overall efficacy of the system.

Moving forward, our future endeavors encompass expanding the scope of our system to encompass a diverse range of domains, accommodating various types of ratings and user comments while catering to a wide array of user preferences. Additionally, we aim to address the issue of imbalanced data by implementing strategies to handle this challenge effectively. We aspire to improve the SA component by integrating advanced algorithms and incorporating additional information to ensure a comprehensive evaluation across different domains. Through ongoing refinement and improvement, our ultimate goal is to develop a flexible and adaptable recommendation service that continually evolves to meet the dynamic needs of users and provides accurate and personalized recommendations while addressing the complexities of imbalanced data.

Data Availability Not applicable

Declarations

Conflicts of interest The authors declare that they have no conflict of interest

References

1. Lee SW, Jiang G, Kong HY, Liu C (2021) A difference of multimedia consumers rating and review through sentiment analysis. Multimedia Tools and Applications. 80:34625–34642
2. Ebadi A, Krzyzak A (2016) A hybrid multi-criteria hotel recommender system using explicit and implicit feedbacks. International Journal of Computer and Information Engineering. 10(8):1450–1458
3. Pu P, Chen L, Hu R (2012) Evaluating recommender systems from the users perspective: survey of the state of the art. User Model User-Adap Inter 22:317–355
4. Patel B, Desai P, Panchal U (2017) Methods of recommender system: a review. In: 2017 International conference on innovations in information, embedded and communication systems (ICIIECS) (pp 1–4). IEEE
5. Nilashi M, bin Ibrahim O, Ithnin N (2014) Hybrid recommendation approaches for multi-criteria collaborative filtering. Expert Syst Appl 41(8):3879–3900
6. Nilashi M, Ibrahim O, Bagherifard K (2018) A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques. Expert Syst Appl 92:507–520
7. Thakker U, Patel R, Shah M (2021) A comprehensive analysis on movie recommendation system employing collaborative filtering. Multimedia Tools and Applications. 80(19):28647–28672
8. Aggarwal CC, Aggarwal CC (2016) Content-based recommender systems. The textbook, Recommender systems, pp 139–166
9. Geetha G, Safa M, Fancy C, Saranya D (2018) A hybrid approach using collaborative filtering and content based filtering for recommender system. In: Journal of physics: conference series (vol 1000, No. 1, p. 012101). IOP Publishing
10. Jain A, Jain V, Kapoor N (2016) A literature survey on recommendation system based on sentimental analysis. Advanced Computational Intelligence. 3(1):25–36
11. Sánchez-Moreno D, Gil González AB, Muñoz Vicente MD, López Batista V, Moreno-García MN. Recommendation of songs in music streaming services: dealing with sparsity and gray sheep problems. InTrends in Cyber-Physical Multi-Agent Systems. The PAAMS Collection-15th International Conference, PAAMS 2017 15 2018 (pp 206-213). Springer International Publishing
12. Birjali M, Kasri M, Beni-Hssane A (2021) A comprehensive survey on sentiment analysis: approaches, challenges and trends. Knowl-Based Syst 226:107134
13. Sharma D, Kumar A (2021) Levels and classification techniques for sentiment analysis: a review. Advances in Communication and Computational Technology: Select Proceedings of ICACCT 2019:333–345
14. Bhavitha BK, Rodrigues AP, Chiplunkar NN (2017) Comparative study of machine learning techniques in sentimental analysis. In: 2017 International conference on inventive communication and computational technologies (ICICCT) (pp 216-221). IEEE

15. Salas–Zte MDP, Medina-Moreira J, Lagos-Ortiz K, Luna-Aveiga H, Rodriguez-Garcia MA, Valencia-Garcia R (2017) Sentiment analysis on tweets about diabetes: an aspect-level approach. Computational and mathematical methods in medicine, 2017

16. Zhang X, Zheng X (2016) Comparison of text sentiment analysis based on machine learning. In: 2016 15th International symposium on parallel and distributed computing (ISPDC) (pp 230–233). IEEE

17. Zhao W, Guan Z, Chen L, He X, Cai D, Wang B, Wang Q (2017) Weakly-supervised deep embedding for product review sentiment analysis. IEEE Trans Knowl Data Eng 30(1):185–197

18. Nouh RM, Lee HH, Lee WJ, Lee JD (2019) A smart recommender based on hybrid learning methods for personal well-being services. Sensors. 19(2):431

19. Kumar S, De K, Roy PP (2020) Movie recommendation system using sentiment analysis from microblogging data. IEEE Transactions on Computational Social Systems. 7(4):915–923

20. Osman NA, Noah SAM, Darwich M (2019) Contextual sentiment based recommender system to provide recommendation in the electronic products domain. International Journal of Machine Learning and Computing. 9(4):425–431

21. Contratres FG, Alves-Souza SN, Filgueiras LVL, DeSouza LS (2018) Sentiment analysis of social network data for cold-start relief in recommender systems. In: Trends and advances in information systems and technologies: vol 2 6 (pp 122–132). Springer International Publishing

22. Ziani A, Azizi N, Schwab D, Aldwairi M, Chekkai N, Zenakhra D, Cheriguene S (2017) Recommender system through sentiment analysis. In: 2nd International conference on automatic control, telecommunications and signals

23. Abbasi F, Khadivar A, Yazdinejad M (2019) A grouping hotel recommender system based on deep learning and sentiment analysis. Journal of Information Technology Management. 11(2)

24. Dubey Abhishek et al (2018) Item-based collaborative filtering using sentiment analysis of user reviews. International conference on application of computing and communication technologies. Springer, Singapore

25. Nabil S, Elbouhdidi J, Chkouri MY (2018) Recommendation system based on data analysis-application on tweets sentiment analysis. In 2018 IEEE 5th International congress on information science and technology (CiSt) (pp 155-160). IEEE

26. Sallam RM, Hussein M, Mousa HM (2022) Improving collaborative filtering using lexicon-based sentiment analysis. International Journal of Electrical and Computer Engineering. 12(2):1744

27. Amazon Musical Instruments http://jmcauley.ucsd.edu/data/amazon/ (Accessed 20 July 2023)

28. Rahali A, Akhloufi MA (2023) End-to-end transformer-based models in textual-based NLP. AI. 4(1):54–110

29. Devlin J, Chang MW, Lee K, Toutanova K (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805

30. Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L (1802) Deep contextualized word representations. CoRR abs/1802.05365 (2018)

31. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Amodei D (2020) Language models are few-shot learners. Adv Neural Inf Process Syst 33:1877–1901

32. Munikar M, Shakya S, Shrestha A (2019) Fine-grained sentiment classification using BERT. In: 2019 Artificial intelligence for transforming business and society (AITB) (vol 1, pp 1–5). IEEE

33. Gao Z, Feng A, Song X, Wu X (2019) Target-dependent sentiment classification with BERT. Ieee Access. 7:154290–154299

34. El-Ansari A, Beni-Hssane A (2023) Sentiment analysis for personalized chatbots in e-commerce applications. Wireless Pers Commun 129(3):1623–1644

35. Loshchilov I, Hutter F (2017) Decoupled weight decay regularization. arXiv:1711.05101

36. Pennington, J., Socher, R.,Manning, C. D. (2014, October). Glove: global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp 1532-1543)

37. Isinkaye FO, Folajimi YO, Ojokoh BA (2015) Recommendation systems: principles, methods and evaluation. Egyptian Informatics Journal. 16(3):261–273

38. Hameed Z, Garcia-Zapirain B (2020) Sentiment classification using a single-layered BiLSTM model. Ieee Access. 8:73992–74001

39. Arbane M, Benlamri R, Brik Y, Alahmar AD (2023) Social media-based COVID-19 sentiment classification model using Bi-LSTM. Expert Syst Appl. 212

40. Sachin S, Tripathi A, Mahajan N, Aggarwal S, Nagrath P (2020) Sentiment analysis using gated recurrent neural networks. SN Computer Science. 1:1–13

41. Pan Y, Liang M (2020) Chinese text sentiment analysis based on BI-GRU and self-attention. In: 2020 IEEE 4th Information technology, networking, electronic and automation control conference (ITNEC) (vol 1, pp 1983–1988). IEEE

42. Elmurngi EI, Gherbi A (2018) Unfair reviews detection on amazon reviews using sentiment analysis with supervised learning techniques. J Comput Sci 14(5):714–726
43. Tilloo, Pallavi, Gottimukkala, Raga, Mamidala, Sreeja (2021) Sentiment analysis for amazon musical instruments user reviews

Authors and Affiliations

Ikram Karabila[1] · Nossayba Darraz[1] · Anas EL-Ansari[2] · Nabil Alami[3] · Mostafa EL Mallahi[1]

Nossayba Darraz
nossayba.darraz@usmba.ac.ma

Anas EL-Ansari
anas.elansari@gmail.com

Nabil Alami
n.alami@ump.ac.ma

Mostafa EL Mallahi
mostafa.elmallahi@usmba.ac.ma

[1] IPI Laboratory, Department of Mathematics and Computer Science, High Normal School, Sidi Mohamed Ben Abdellah University, Fez, Morocco

[2] MASI Laboratory, Computer Science Department, Polydisciplinary Faculty, Mohammed First University, Nador, Morocco

[3] MASI Laboratory, Mohammed First University, Higher School of Technology, Nador, Morocco