IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 2, April 2026, pp. 1851~1862
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i2.pp1851-1862

Sentiment-aware user-item recommendation combining weighted XGBoost and optimized similarity metrics

Snehal Bhogan, Vijay S. Rajpurohit, Sanjeev S. Sannakki
Department of Computer Science and Engineering, Karnatak Law Society Gogte Institute of Technology, affiliated to Visvesvaraya Technological University, Belagavi, India

Article Info

Article history:

Received May 26, 2025
Revised Jan 5, 2026
Accepted Jan 25, 2026

Keywords:

Collaborative filtering
Deep learning
Ecommerce
Parameterized BERT
Sentiment analysis
Similarity metrics
Weighted extreme gradient boosting

ABSTRACT

User-item recommendation systems play a vital role in enhancing personalized digital experiences across e-commerce and social media platforms. Traditional recommendation approaches, such as collaborative filtering (CF) and content-based filtering (CBF), often suffer from challenges like data sparsity, cold-start issues, and limited contextual understanding. Sentiment-aware recommendation systems have emerged as a promising solution by incorporating emotional insights extracted from user reviews, thereby improving recommendation accuracy and personalization. This study proposes a novel sentiment-aware user-item recommendation system (SAUIRS) framework that integrates optimized term frequency-inverse document frequency (O-TF-IDF), parameterized bidirectional encoder representations from transformers (P-BERT), weighted extreme gradient boosting (WXGBoost), and an optimized similarity metrics model. The optimized TF-IDF enhances feature selection, reducing dimensionality while preserving relevant textual information. P-BERT, a fine-tuned BERT model, improves sentiment classification accuracy by leveraging deep contextual embeddings. WXGBoost further refined sentiment predictions, addressing class imbalance and enhancing model robustness. The extracted sentiment information is incorporated into an optimized similarity metrics model to improve recommendation precision by aligning user preferences with sentiment-driven insights. Extensive experiments conducted on Amazon benchmark datasets demonstrate the superior performance in terms of accuracy, root mean square error (RMSE), and mean absolute error (MAE) of the proposed framework compared to state-of-the-art recommendation models.

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

Corresponding Author:

Snehal Bhogan
Department of Computer Science and Engineering, Karnatak Law Society Gogte Institute of Technology affiliated to Visvesvaraya Technological University, Belagavi, India
Belagavi-590018, India
Email: snehalrbhogan@gmail.com

1. INTRODUCTION

With the rapid expansion of e-commerce, social media, and online content platforms, user-item recommendation systems have become a crucial component of personalized digital experiences [1], [2]. Traditional recommendation algorithms, such as collaborative filtering (CF) and content-based filtering (CBF), often suffer from challenges like data sparsity, cold-start problems, and an inability to incorporate contextual and sentiment-driven insights [3], [4]. Sentiment-aware recommendation systems aim to bridge this gap by integrating users' emotional inclinations extracted from textual reviews, leading to more accurate, and personalized recommendations [5], [6]. Natural language processing (NLP) techniques have played a transformative role in improving sentiment analysis for recommendation systems [7], [8]. Conventional term frequency-inverse document frequency (TF-IDF) methods [9], [10] often fall short in capturing contextual meanings, whereas advanced deep learning models [11], [12] such as bidirectional encoder representations from transformers (BERT) [12], [13] have significantly enhanced sentiment classification. However, optimizing these models for effective user-item recommendation [14] remains a challenge due to computational complexity and the need for efficient feature representation [15], [16].

This study proposes a novel sentiment-aware user-item recommendation system (SAUIRS) that leverages an optimized TF-IDF for initial feature extraction, a parameterized BERT (P-BERT) for enhanced sentiment classification, and a weighted extreme gradient boosting (WXGBoost) model for robust sentiment prediction. The sentiment information is further integrated into an optimized similarity metrics model to refine user-item recommendations. The proposed hybrid model effectively captures user preferences, enhances interpretability, and improves recommendation accuracy by incorporating both sentiment-driven and content-based features. The primary contributions of this research include:

i) Development of an optimized TF-IDF technique to enhance feature selection and reduce dimensionality.
ii) Implementation of a P-BERT model to improve sentiment classification accuracy.
iii) Integration of a WXGBoost model to enhance classification robustness and mitigate class imbalance.
iv) Proposal of an optimized similarity metrics model to refine user-item recommendations using sentiment-aware insights.
v) Empirical validation through extensive experiments demonstrating improved accuracy and efficiency over existing recommendation frameworks.

Manuscript organization: section 2 reviews existing sentiment analysis–based recommendation methods used in e-commerce platforms. Section 3 presents a hybrid sentiment analysis and enhanced CF approach, while section 4 discusses experimental results and comparative analysis, followed by conclusions and research significance.

2. LITERATURE SURVEY

The integration of deep learning techniques, particularly large language models (LLMs) and sentiment analysis, has significantly advanced intelligent product recommendation systems. This literature survey reviews recent studies by analyzing their methodologies, datasets, optimized metrics, contributions, limitations, and future research directions. Thomas and Jeba [17] proposed a bigram-based deep learning framework that incorporates sentiment analysis to enhance product recommendation accuracy. By extracting sentiment scores from user reviews, the system refines recommendation precision and recall. The model was trained and validated using e-commerce datasets consisting of user reviews and ratings. Incorporating sentiment information enables the system to capture nuanced user preferences, resulting in improved personalization. However, the study highlights challenges related to large-scale data processing and recommends future exploration of efficient algorithms and real-time processing mechanisms. Abdalla *et al.* [18] introduced a hybrid recommendation model combining self-attention mechanisms with bidirectional long short-term memory (Bi-LSTM) networks and incentive learning-based CF. The model was evaluated using mean absolute error (MAE) and root mean square error (RMSE) metrics on e-commerce datasets containing user-item interactions and review texts. Self-attention enables effective contextual representation, improving recommendation accuracy. Nevertheless, the model's architectural complexity increases computational cost, suggesting the need for optimization and domain generalization in future research. Ibrahim *et al.* [19] presented a hybrid neural CF approach that integrates neural networks with traditional CF techniques. Performance was evaluated using precision, recall, and F1-score on publicly available datasets such as MovieLens. The hybrid framework effectively models both linear and nonlinear user-item interactions, improving recommendation accuracy. However, performance degradation under sparse data conditions remains a challenge, motivating future work on data augmentation and contextual feature integration. Sami *et al.* [20] proposed a hybrid recommendation model that combines deep learning with traditional recommendation algorithms to enhance personalization. The system was evaluated using accuracy and coverage metrics on datasets containing user interaction logs and profile information from online platforms. While the hybrid approach improves recommendation quality, the study emphasizes the need for scalable architectures and real-time data processing for practical deployment. Guo and Zhang [21] employed deep learning-based embeddings to predict customer satisfaction in cross-border e-commerce environments. The model optimized classification accuracy and F1-score using datasets comprising customer reviews and satisfaction ratings. Accurate satisfaction prediction supports strategic decision-making and customer retention. However, potential biases in customer feedback limit model generalizability, and the authors recommend incorporating diverse data sources in future studies. Yao and Zheng [22] proposed an enhanced sentiment analysis model combining transformers with conditional random fields (CRFs) to capture contextual dependencies in text. Evaluated on Yelp and Amazon review datasets, the model achieved an accuracy of 93.4%. While the transformer-CRF integration improves linguistic understanding, its computational complexity motivates future research on model optimization and real-time applicability.

Roy *et al.* [23] introduced a hybrid recommendation system integrating content-based and item-based CF with cascaded LLMs. The model processes product metadata and user reviews to enhance recommendation accuracy, precision, and recall. Experiments conducted on online apparel retail datasets, particularly men's shirt categories, demonstrate improved personalization through nuanced preference modeling. However, scalability and real-time processing remain open challenges. Garapati and Chakraborty [24] proposed the review text granularity (RTG) model to improve sentiment analysis and rating prediction by analyzing review texts at varying levels of detail. Evaluated using MAE and RMSE on e-commerce datasets, the model captures subtle opinion nuances. Despite its effectiveness, computational complexity remains a limitation, highlighting the need for efficient architectures and contextual feature integration. Yang *et al.* [25] developed an attentive aspect-based recommendation network (AARN) that employs attention mechanisms to capture aspect-level user preferences from reviews. The model optimized precision, recall, and F1-score using annotated e-commerce review datasets. Aspect-level modeling enables more personalized recommendations, though accurately extracting and weighting aspects remains challenging. Future work suggests refining aspect extraction and integrating external knowledge sources.

In summary, these studies demonstrate that incorporating advanced NLP techniques such as LLMs, attention mechanisms, and aspect-level sentiment modeling significantly improves recommendation accuracy and personalization. However, challenges related to data imbalance, multi-domain e-commerce environments, computational complexity, and real-time processing persist. These limitations motivate the development of the proposed SAURIS framework discussed in the subsequent section.

3. PROPOSED METHOD

The SAUIRS presented in Figure 1 is designed considering recommending Top N items to users and identifying users with similar interests. The SAUIRS first collects realistic data from the Amazon website, considering multiple products of different domains, which include cosmetics, food, and electronics. More details of the datasets and their attribute are provided in sub-section 3.1. Then, the SAUIRS performs sentiment analysis using optimized TF-IDF and a transformer model employing a P-BERT model. More details of the process are discussed in sub-sections 3.2 and 3.3. Then, introduces WXGBoost to perform the sentiment classification task and optimized similarity metrics. A model is introduced to perform Top N recommendations. More details of this are presented in sub-sections 3.4 and 3.5.

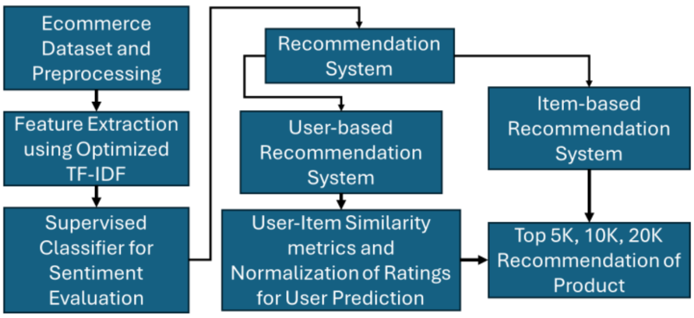

Figure 1. Architecture of SAUIRS

3.1. Dataset and preprocessing

The study uses the Amazon dataset which spans multiple product categories like product reviews, laptop reviews, and tablet reviews [24], [25], maintained by the University of California San Diego. Due to the dataset's large size [25], a subset of 30,000 rows having multiple categories were extracted to the men's shirt category under clothing, shoes, and jewelry were selected randomly for creating a novel dataset for training, testing, and validation. Further, the Amazon laptop and tablet datasets [25] were also collected from the publicly released Amazon product reviews 2 (5-core) collection, which ensures that each user and each item have at least five associated reviews. From the electronics category, laptop- and tablet-related entries were selected, preprocessed, and subsequently divided into training, validation, and test partitions following an 80:20 split. Only relevant fields necessary for constructing the sentiment-based recommendation system were retained from the original dataset. These fields include: Id (asin): a unique identifier for each product, Brand (details): details of the product's brand. Categories (main_category): the primary category of the product. Manufacturer name (Manufacturer): the name of the manufacturer. Reviews_date (timestamp): the date when the review was posted. Reviews_didPurchase (verified_purchase): indicates whether the product was purchased by the reviewer. Reviews_doRecommend (bought_together): recommended bundles from the websites. Reviews_rating (rating): numerical ratings given by users. Reviews_text (text): the textual review content. Reviews_title (title): titles of the reviews. Reviews_username (reviewerName): the username of the reviewer. User_sentiment: the manually annotated sentiment label for each review (positive or negative). The preprocessing step ensured the dataset is clean and structured for analysis. Key tasks included are as follows: handling missing data: null values in the dataset were addressed by removing that particular row. Data cleaning (removing unnecessary columns): irrelevant data columns were removed. Only the essential columns from the dataset were retained, based on exploratory data analysis (EDA). Moreover, the text-based review data underwent the following steps: punctuation removal: non-alphanumeric characters were eliminated. Stopword removal: common words (e.g., "and" and "the") that did not contribute to the sentiment were filtered out. Lemmatization: words were reduced to their root form to standardize textual content.

3.2. Feature extraction

To make the cleaned dataset suitable for machine learning, feature extraction was carried out using a newly introduced optimized term frequency-inverse document frequency (O-TF-IDF) method. TF-IDF is a statistical technique that reflects the relevance of a term within a specific document relative to an entire collection. It is commonly applied in information retrieval and text analytics to transform textual input into meaningful numerical representations that can be utilized by learning algorithms. Term frequency (TF) reflects how often a particular word occurs within a document compared with the total number of words in that document as obtained in (1).

$$TF(t) = \frac{Number\ of\ times\ term\ t\ appears\ in\ a\ document}{Total\ Number\ of\ terms\ in\ the\ document} \tag{1}$$

In contrast, inverse document frequency (IDF) indicates the significance of a term by lowering the influence of words that appear frequently across many documents, such as common stopwords as in (2).

$$IDF(t) = \log\left(\frac{Total\ number\ of\ documents}{Number\ of\ documents\ containing\ term\ t}\right) \tag{2}$$

The TF-IDF value is obtained by multiplying these two components as in (3).

$$TF - IDF(t) = TF(t) \times IDF(t) \tag{3}$$

This combined measure emphasizes terms that are more distinctive or meaningful within a specific document while downplaying words that are commonly found throughout the entire corpus. Although TF-IDF remains a widely used and effective method for transforming text into numerical features, it still presents certain shortcomings that may limit its effectiveness for specific applications. The main problem identified during recommendation system design using sentiment analysis are as follows:

i) TF-IDF treats words or n-grams as independent features and does not consider the semantic or syntactic relationships between them. Hence, due to this it cannot understand the meaning of words or phrases, leading to a lack of context in feature representation. For example, "not bad" and "bad" may be treated similarly despite having opposite sentiments.
ii) The feature space grows rapidly with the size of the vocabulary and the use of n-grams (e.g., bigrams or trigrams). Hence, resulting in large, sparse matrices, increasing computational costs and memory requirements. It also makes models prone to overfitting, especially with small datasets.
iii) Words that are very common across documents (e.g., "product" and "review") may still have non-zero weights due to term frequency, even if they add little value. This can introduce noise into the feature set.

iv) Rare but meaningful terms (e.g., "awesome" and "disastrous") may be discarded if the mined threshold is too high. This can lead to a loss of critical information.
v) The performance of TF-IDF heavily depends on the selection of parameters like minds, max_df, and ngram_range. Poor parameter choices can result in irrelevant features being included or significant features being excluded.

For solving the above issues, the O-TF-IDF has been enhanced by selecting the min_df, max_df and n_gram range. As TF-IDF vectorization is used to convert text data into numerical features for model training. The following parameter optimized metrics been used for enhancing the TF-IDF. That is, the configuration of parameters are tfidf=TfidfVectorizer(min_df =5, max_df =0.95, ngram_range =(1,2)), min_df =5: ignores terms appearing in fewer than 5 documents, max_df =0.95: excludes terms that appear in more than 95% of documents, and ngram_range =(1,2): the advantage of optimized TF-IDF are as follows:
i) By including bigrams (ngram_range =(1,2)), this work significantly increases the number of features, leading to a high-dimensional sparse matrix. This provides better features.
ii) The min_df =5 and max_df =0.95 thresholds are chosen according to the dataset. These parameters provide a balance between retaining important terms and removing noise.
iii) By optimizing the TF-IDF approach, better performance is achieved providing better results for sentiment-based product recommendation.

This optimized TF-IDF method assigned importance to words considering unigrams and bigrams based on their frequency across reviews, ensuring better representation of textual content for the supervised machine learning classifier.

3.3. Sentiment analysis with transformer-based model

The extracted feature set was subsequently fed into a BERT-based model [15] to obtain sentiment scores. Among transformer-driven techniques, BERT [15] remains one of the most widely adopted due to its strong contextual understanding. Transformers, a class of deep learning models, rely on a self-attention mechanism that identifies the relevance of each feature and its contextual meaning within a review. Unlike traditional neural network architectures, transformer models support parallel computation, enabled by self-attention and positional embeddings, which capture long-range dependencies efficiently. For this reason, BERT is selected for sentiment evaluation. However, given the size of the dataset used in this study—approximately 30,000 reviews standard BERT models and their common variants struggle to deliver optimal performance. To address this limitation, this work proposes a parameter-optimized version of BERT, referred to as P-BERT. The architecture of the proposed P-BERT model is illustrated in Figure 2.

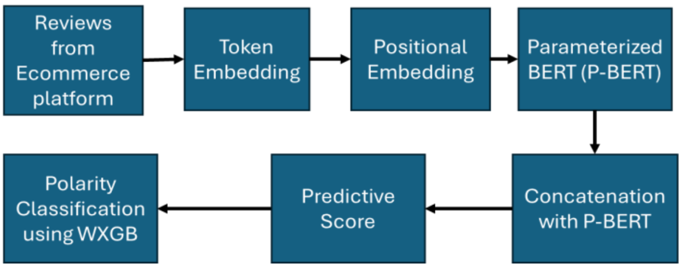

Figure 2. Sentiment predictive score, polarity labeling, and polarity classification

In extracting polarities at a given review considering the contextual feature level, let the collection of review-text in the dataset be x and N denote words in each review-text. This is mathematically denoted as (4).

$$X = \{x^N\} \tag{4}$$

By using X, overall polarity of review-text is identified using (5), in a tuple-set format represented as T.

$$T = \{(a_i, o_i, s_i)\} \tag{5}$$

In (5), a_i denotes contextural features, o_i denotes polarity, and s_i denotes review-polarity for given X. The P-BERT approach considers position-tagging technique for representing contextual features; hence, tagging is done on complete review-text where the contextual-tags for starting-point and ending-point was denoted as A_s and A_e and its respective polarities were denoted as O_s and O_e. Moreover, position-encoding was incorporated in P-BERT for retrieving contextural feature position and its respective polarity as tokens.

In (6), the CR_e denotes overall tokens present in review-text, r_{el} denotes single tokens in the review-text. Using (6), the contextural feature polarity was extracted. The BERT approach was modified during fine-tuning process in this work. For modifying BERT, a SoftMax function was used for extracting review-level polarity and classifying polarity, to get the classification results achieved by the P-BERT. The output (tokens) from BERT were used as input for SoftMax function. The main aim of using SoftMax function was to use polarity given by BERT approach, and convert it to probabilities. The probabilities provided polarity likelihood for every class (neutral, negative, and positive), using which review-level polarity was extracted. The SoftMax probability was evaluated using (7).

$$CR_e = \{r_{e1}, r_{e2}, \dots, r_{el}\} \tag{6}$$

$$Y = Softmax(CR_{e1}W_{e1} + b_{e1}) \tag{7}$$

In (7), W_{e1} denotes a weight matrix, which was utilized for transforming contextualized tokens to format which was used for classification and b_{e1} denotes a bias vector which was added to linear transformed output for changing the outcome and enhancing BERT flexibility. In the P-BERT framework, *CLS* denotes the classification token, *TOK* refers to individual tokens, and *SEP* marks the separator. Each review is first processed through token and positional embedding layers, where every word is assigned a token identifier and a corresponding position index. This study employs only the encoder component of BERT, as the encoder alone is sufficient to generate rich contextual embeddings for the input text. The encoder contains a self-attention mechanism combined with a feed-forward network that transforms the review sequence into contextual representations. Unlike the original BERT model, which relies on static masking, P-BERT incorporates dynamic masking to produce more diverse and context-aware representations. Additionally, instead of using character-level byte-pair encoding, the proposed model adopts byte-level byte pair encoding (BPE) to accelerate computation. These enhancements allow P-BERT to handle larger datasets with more batches and longer sequences than standard BERT. Before constructing the attention masks and input IDs, each review is tokenized. During concatenation, the attention mask identifies the relative importance of each token, whereas input IDs convert the text into a sequential numeric format. Both components serve as inputs to the P-BERT architecture. The final P-BERT configuration includes 768 hidden units and 12 encoder layers. Its output layer categorizes each review as positive, neutral, or negative. The resulting classification scores are subsequently processed using a WXGBoost model, which is described in the following section.

3.4. Weighted extreme gradient boosting for sentiment classification

This model classified user reviews as either positive or negative, using accuracy, precision, and F-score performance metrics. The sentiment analysis step formed the foundation for generating user and item-based recommendations. For polarity using the proposed model, consider the e-commerce dataset X with $tThe$ number of samples rated by the P-BERT model is defined in (8).

$$X = \{(x_1, y_1), (x_2, y_2), \dots (x_t, y_t)\} \tag{8}$$

In (8), x_t denotes a feature vector for every t^{th} observation and $y_t \in \{0,1\}$ where $y_t = 1$ denotes positive sample and $y_t = 0$ denotes a negative sample. The supervised classifier model F maps input features X for getting output y, as presented in (9).

$$F = f(X) = \sum_{k=1}^{K} \theta_k \cdot h_k(X) \tag{9}$$

In (9), K denotes the total number of trees present, θ_k denotes weights assigned to k^{th} tree and $h_k(X)$ denotes output of k^{th} for input X. The WXGBoost main aim is to minimize log-loss while incorporating a regularization term for preventing overfitting. The loss function is evaluated using (10).

$$L = \sum_{t=1}^{N} [y_t \log(p_t) + (1 - y_t) \log(1 - p_t)] + \sum_{k=1}^{K} \Omega(h_k) \tag{10}$$

In (10), $p_t = \sigma(f(x_t))$ is predicted positive polarity probability, where sigmoid-activation σ is evaluated as presented in (11). Also, $\Omega(h_k)$ in (10) is evaluated using (12).

$$\sigma(z) = \frac{1}{1+e^{-z}} \tag{11}$$

$$\Omega(h_k) = \gamma N + \frac{1}{2}\lambda\left|\left|w_k\right|\right|^2 \tag{12}$$

In (12), N represents the total number of leaves in each decision tree, while w_k corresponds to the weight assigned to the k-th leaf. The parameters γ and λ serve as regularization terms that manage the overall complexity of the tree. During training, the WXGBoost model processes the e-commerce dataset by minimizing the loss function L, identifying the most effective tree structure through optimal node splitting, determining the ideal leaf weights w_k, and ensuring an appropriate balance between predictive accuracy and generalization by regulating $\Omega(h_k)$.

3.5. Recommendation using similarity metrics and CF

CF comes into play, which involves the creation of a cumulative product set based on other products deemed like the ones selected through content filtering, and using this approach, products that share commonalities with user preferences are identified. Subsequently, the similarity between the embeddings of the user-entered custom phrase and the products' description and review in the corpus using the cosine formula as in (13).

$$cosC = \frac{\langle a,b\rangle}{\|a\|\cdot\|b\|} \tag{13}$$

Where, a and b represent the cascaded embeddings of custom user phrases and concatenated product description and review information in the dataset respectively. $\langle a, b\rangle$ represents the dot product of vectors. The vector norm for a is shown in (14).

$$\|a\| \cdot \|b\| = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2} \tag{14}$$

In this study, the model evaluates the similarity between an active user $c_a$and each neighboring user $c_n$using the cosine similarity metric shown in (13). Here, the neighbors of c_a are defined as users who have provided ratings for the same items and whose rating patterns closely resemble those of the active user. To estimate the final predicted ratings, the system combines the criterion scores for each recommended item using a weighted aggregation strategy. The aggregated scores are used to compute the expected rating for each item, after which the items are sorted according to their predicted values. The highest-ranked items are ultimately recommended to the user.

$$S(c_a, c_n) = \frac{\sum_{i=1}^{N} Rt_{ai}Rt_{ci}}{\sqrt{\sum_{i=1}^{N} Rt_{ai}^2}\sqrt{\sum_{i=1}^{N} Rt_{ci}^2}} \tag{15}$$

Let the string representation of the active user be denoted as Sc_a and that of the neighboring user as Sc_n. Their vectorized forms are represented by Vc_a and Vc_n, respectively. For each item associated with the active user u_a, the corresponding genres are extracted, concatenated into a textual sequence St_a, and then transformed into a vector Vc_a. A similar process is applied for the neighboring user u_n, where the categories for each item are combined into St_n and converted into the vector form Vc_n. These representations are generated using the BERT-XGBoost model. Because a user's full rating history defines their behavioral profile, the BERT-XGBoost model processes all items rated by a user as the contextual input for each c_n. The similarity significance between Vc_a and Vc_n is further refined using a normalized Euclidean distance measure. In (16) integrates both the similarity score $S(c_a, c_n)$ and the corresponding weight $W(c_a, c_n)$ to compute the predicted rating for an unrated item ur_i. The preference of user c_a for item ur_i is inferred from the ratings of the K most similar neighbors.

$$P_{ai} = \bar{r}_a + \frac{\sum_{i=1}^{K} W(c_a,c_i)\cdot S(c_a,c_i)\cdot(r_{ai}-\bar{r}_a)}{\sum_{i=1}^{K}|W(c_a,c_i)\cdot S(c_a,c_i)|} \tag{16}$$

Here, r_{ai} denotes the rating assigned by c_a to the item ur_i, $\bar{r}_a$ is the average rating of user c_a, c_i represents the neighboring user, $S(c_a, c_i)$ is the similarity measure, and $W(c_a, c_i)$ is the weight associated with that similarity. The outputs obtained from CF and sentiment analysis are then combined to generate a unified recommendation list. Given a rating matrix $R_{k\times l}$, where k denotes the number of users and l denotes the number of items, each rating $r_{ai} \in R_{k\times l}$ represents the score assigned by user a to item i. The final predicted rating for item i by user $a$is obtained as in (17).

$$P_{final} = \alpha \cdot P_{ai} \tag{17}$$

Using the modelled similarity matrix in this work, user-item similarity matrix is constructed to identify similar interest users; then, user rating normalized to perform prediction of what items a user may purchase. Then, item-based recommendation is performed by identifying top-10K, Top-20K items for the users. This system generates recommendations for products that are similar to those previously bought or rated highly by the user. The model similarity metrics is designed in such a way when no rating is available for particular products, it can still make Top-N recommendations to users.

4. RESULTS AND DISCUSSION

This section studies the performance attained by proposed model and various other baseline models using multi-domain ecommerce Amazon dataset. The study includes studying performance for sentiment classification and Top-N recommendation using novel hybrid model combined with new similarity metrics-based collaborative metrics. The findings of polarity classification are evaluated using accuracy defined in (18).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{18}$$

To assess the effectiveness of the proposed model, two commonly adopted evaluation metrics: MAE in (19) and RMSE in (20) were employed for rating prediction. These measures are widely used in recommendation systems to quantify prediction accuracy. RMSE emphasizes larger deviations by squaring the errors, thereby penalizing significant mispredictions more heavily than MAE. For this reason, RMSE is frequently preferred when large errors are particularly undesirable. Nevertheless, because both metrics can be influenced by outliers, MAE may be more suitable for datasets with irregular or skewed rating distributions.

$$MAE = \frac{1}{N}\sum_{n=1}^{N}\left|r_{i,j} - \hat{r}_{i,j}\right| \tag{19}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(r_{i,j} - \hat{r}_{i,j}\right)^2} \tag{20}$$

Here, N denotes the total number of samples in the test set, while $r_{i,j}$ and $\hat{r}_{i,j}$ correspond to the actual and predicted ratings, respectively. Since both metrics measure prediction error, lower values indicate improved accuracy, and stronger predictive capability of the model.

4.1. Sentiment analysis performance evaluation

This section studies evaluated the sentiment classification performance using proposed WXGBoost based model with other existing machine learning-based classifiers [23]. The study showed in [23], that employing TF-IDF with machine learning classifier aided the performance in comparison with Countvectorizer-based machine learning model. The work introduces a novel XGBoost+optimized TF-IDF model ensuring an accuracy of 88%; then, the work introduced a novel WXGBoost combined with optimized TF-IDF, ensuring an accuracy of 94% as shown in Table 1. Thus, the result shows the proposed model attains significantly higher accuracy by learning through BERT-based scoring combined with XGBoost-based weighted classifier when compared with existing methods.

Table 1. Comparative study with machine learning-based sentiment analysis model

Model	Accuracy (%)
Logistic regression+TF-IDF [23], 2024	81.32
Naïve Bayes+TF-IDF [23], 2024	79.82
XGBoost+TF-IDF [23], 2024	81.67
Random forest+TF-IDF [23], 2024	83.01
CRF [22], 2024	93.4
XGBoost+O-TF-IDF [proposed]	88
WXGBoost+O-TF-IDF [proposed]	94.5

4.2. Recommendation performance evaluation

This section provides a comparative evaluation of the SAUIRS recommendation framework and several well-established baseline models. To verify the effectiveness of the SAUIRS model, the baselines were grouped into four main categories: i) methods that rely exclusively on numerical ratings (CF and probabilistic matrix factorization (PMF)), ii) approaches that integrate review text into the prediction process (hidden factors and hidden topics (HFT) and retrieval-augmented retrieval (RAR)V2), iii) attention-driven models that exploit neural attention mechanisms (neural attentional regression model with review-level explanations (NARRE) and review semantics based model (RSBM)), and iv) aspect-aware recommendation models (aspect attention-based neural collaborative filtering (A3NCF) and user and context aware model (UCAM)). A concise overview of the baseline techniques is summarized as follows:

- CF [26]: one of the earliest and most influential recommendation strategies, CF identifies similar users or items by analyzing historical rating patterns, and recommends items from the closest neighbors.
- PMF [27]: PMF extends the matrix factorization paradigm by introducing Gaussian priors, allowing it to respond effectively to highly sparse and imbalanced rating matrices.
- HFT [28]: the HFT model employs latent Dirichlet allocation (LDA) to derive underlying topics from user and item reviews and fuses these topics with latent factors obtained through matrix factorization.
- NARRE [29]: the NARRE mechanism that highlights impactful reviews. User and item embeddings are learned using matrix factorization, while a CNN identifies informative review segments for the final prediction.
- RARV2 [30]: this method enriches review-text representation by combining multiple embedding types. It utilizes BERT and robustly optimized BERT pretraining approach (RoBERTa) as auxiliary sources of semantic information within a deep matrix factorization framework.
- RSBM [31]: the RSBM extracts semantic features from reviews using a CNN and applies an attention network to capture aspect-specific evaluations for rating prediction.
- A3NCF [32]: an aspect-aware neural model that adapts to user-specific preferences across various item aspects. Topic modeling is used to identify user interests and item attributes, which are then integrated into the recommendation process.
- UCAM [33]: a deep-learning-based, context-aware model that merges user-item interactions with review representations. Review features are obtained using a BERT-based aspect-based sentiment analysis (ABSA) module to extract aspect-oriented sentiments.

The RMSE and MAE scores for the baseline models are reported in Table 2. The results demonstrate that the proposed SAUIRS framework, leveraging hybrid optimization, enhanced similarity computation, and a sentiment-aware classification module, achieves a substantial reduction in predictive error compared with all competing approaches. The MAE performance of different transformer and attention-based models is tabulated in Table 3. The results can be seen in the proposed SAUIRS model, employing hybrid optimization with optimized similarity metrics combined with sentiment classifier enabled the model to significantly reduce predictive error in comparison with different baseline models.

Table 2. Comparative study with baseline models

Model	RMSE	MAE
CF [26], 2001	1.291	1.602
PMF [27], 2007	1.003	1.008
HFT [28], 2013	0.901	1.297
RARV2 [29], 2018	0.655	0.809
NARRE [30], 2023	1.023	1.373
RSBM [31], 2020	0.578	0.799
A3NCF [32], 2018	0.992	1.307
UCAM [33] 2020	0.767	0.997
AARN [25], 2024	0.509	0.764
RTG [24], 2025	0.267	0.217
SAUIRS [proposed]	0.245	0.311

Table 3. Comparative study with transformers and attention-based models

Model	MAE
Doc2Vec	0.692
BERT	0.585
RoBERTa	0.576
AARN [25]	0.509
SAUIRS [proposed]	0.245

The results indicate significant improvements in recommendation accuracy. validating the effectiveness of sentiment-aware features in refining user-item recommendations. The proposed methodology offers a scalable and interpretable approach, bridging the gap between sentiment analysis and recommendation systems. This study contributes to advancing intelligent recommender systems by integrating deep learning-based sentiment classification with optimized similarity measures, paving the way for more accurate and user-centric recommendations.

5. CONCLUSION

This study introduced a novel SAUIRS framework integrating O-TF-IDF, P-BERT-based sentiment classification, WXGBoost, and an optimized similarity metrics model. The approach was evaluated using the Amazon multi-domain dataset comprising 30,000 reviews across diverse product categories. Performance was assessed using accuracy, MAE, and RMSE, ensuring a comprehensive evaluation of effectiveness. Experimental results demonstrated that the proposed framework outperforms existing recommendation models by achieving higher accuracy while reducing error metrics. The sentiment classification module, powered by P-BERT and WXGBoost, significantly improved prediction precision, leading to more relevant and user-centric recommendations. O-TF-IDF effectively captured essential features, mitigating sparsity issues and enhancing contextual understanding of reviews. Furthermore, sentiment-aware similarity metrics refined the recommendation process by aligning suggestions with users' emotional inclinations. Quantitative analysis confirmed accuracy improvements over traditional CF methods, while reduced MAE and RMSE validated robustness and reliability. Overall, the proposed SAUIRS model offers a scalable, efficient, and sentiment-aware solution. Future work will focus on real-time deployment, large-scale optimization, and reinforcement learning–based adaptive refinement.

FUNDING INFORMATION

This research received no external funding.

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
Snehal Bhogan	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
Vijay S. Rajpurohit	✓	✓			✓	✓				✓		✓	✓	
Sanjeev S. Sannakki	✓									✓		✓	✓	

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

CONFLICT OF INTEREST STATEMENT

The authors declare that no financial or commercial relationships exist that could be construed as a potential conflict of interest. The corresponding author confirms this declaration on behalf of all co-authors.

ETHICAL APPROVAL

Not applicable. This study did not involve human participants, human data, or any personally identifiable information. All data used were either publicly available, fully anonymized, or derived from non-human sources, and therefore no informed consent was required from individuals.

DATA AVAILABILITY

The data and tools used in this study are available from the corresponding author, [SB], upon reasonable request. Any use of the shared resources must be appropriately cited.

REFERENCES

[1] A. Daza, N. D. G. Rueda, M. S. A. Sánchez, W. F. R. Espíritu, and M. E. C. Quiñones, "Sentiment analysis on e-commerce product reviews using machine learning and deep learning algorithms: a bibliometric analysis, systematic literature review, challenges and future works," *International Journal of Information Management Data Insights*, vol. 4, no. 2, Nov. 2024, doi: 10.1016/j.jjimei.2024.100267.

[2] V. Gooljar, T. Issa, S. H.-Ramanan, and B. A.-Salih, "Sentiment-based predictive models for online purchases in the era of marketing 5.0: a systematic review," *Journal of Big Data*, vol. 11, no. 1, Aug. 2024, doi: 10.1186/s40537-024-00947-0.

[3] B. Sabiri, A. Khtira, B. El Asri, and M. Rhanoui, "Hybrid quality-based recommender systems: a systematic literature review," *Journal of Imaging*, vol. 11, no. 1, Jan. 2025, doi: 10.3390/jimaging11010012.

[4] P. Vijayaragavan *et al.*, "Publisher correction: sustainable sentiment analysis on e-commerce platforms using a weighted parallel hybrid deep learning approach for smart cities applications," *Scientific Reports*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-81475-y.

[5] M. Kumar, L. Khan, and H.-T. Chang, "Evolving techniques in sentiment analysis: a comprehensive review," *PeerJ Computer Science*, vol. 11, Jan. 2025, doi: 10.7717/peerj-cs.2592.

[6] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: a systematic literature review," *Journal of King Saud University-Computer and Information Sciences*, vol. 36, no. 4, Apr. 2024, doi: 10.1016/j.jksuci.2024.102048.

[7] S. E. Sorour, A. Alojail, A. El-Shora, A. E. Amin, and A. A. Abohany, "A hybrid deep learning approach for enhanced sentiment classification and consistency analysis in customer reviews," *Mathematics*, vol. 12, no. 23, Dec. 2024, doi: 10.3390/math12233856.

[8] E. Hashmi and S. Y. Yayilgan, "A robust hybrid approach with product context-aware learning and explainable AI for sentiment analysis in Amazon user reviews," *Electronic Commerce Research*, vol. 25, no. 6, pp. 5139–5171, Dec. 2025, doi: 10.1007/s10660-024-09896-5.

[9] A. S. Sridhar and S. Nagasundaram, "SVM-RideNN: hybrid support vector machine-rider neural network based sentiment analysis using product review," *Australian Journal of Electrical and Electronics Engineering*, vol. 22, no. 4, pp. 729–741, Oct. 2025, doi: 10.1080/1448837X.2024.2440170.

[10] E. S. P. Krishna *et al.*, "Enhancing e-commerce recommendations with sentiment analysis using MLA-EDTCNet and CF," *Scientific Reports*, vol. 15, no. 1, Feb. 2025, doi: 10.1038/s41598-025-91275-7.

[11] O. Bellar, A. Baina, and M. Ballafkih, "Sentiment analysis: predicting product reviews for e-commerce recommendations using deep learning and transformers," *Mathematics*, vol. 12, no. 15, Aug. 2024, doi: 10.3390/math12152403.

[12] N. Darraz, I. Karabila, A. El-Ansari, N. Alami, and M. El Mallahi, "Integrated sentiment analysis with BERT for enhanced hybrid recommendation systems," *Expert Systems with Applications*, vol. 261, Feb. 2025, doi: 10.1016/j.eswa.2024.125533.

[13] I. Karabila, N. Darraz, A. EL-Ansari, N. Alami, and M. EL Mallahi, "BERT-enhanced sentiment analysis for personalized e-commerce recommendations," *Multimedia Tools and Applications*, vol. 83, no. 19, pp. 56463–56488, Dec. 2023, doi: 10.1007/s11042-023-17689-5.

[14] P. Rasappan, M. Premkumar, G. Sinha, and K. Chandrasekaran, "Transforming sentiment analysis for e-commerce product reviews: hybrid deep learning model with an innovative term weighting and feature selection," *Information Processing & Management*, vol. 61, no. 3, May 2024, doi: 10.1016/j.ipm.2024.103654.

[15] D. M. Alghazzawi, A. G. A. Alquraishee, S. K. Badri, and S. H. Hasan, "ERF-XGB: ensemble random forest-based XGBoost for accurate prediction and classification of e-commerce product review," *Sustainability*, vol. 15, no. 9, Apr. 2023, doi: 10.3390/su15097076.

[16] H.-S. Le *et al.*, "Predictive model for customer satisfaction analytics in e-commerce sector using machine learning and deep learning," *International Journal of Information Management Data Insights*, vol. 4, no. 2, Nov. 2024, doi: 10.1016/j.jjimei.2024.100295.

[17] R. Thomas and J. R. Jeba, "A novel framework for an intelligent deep learning based product recommendation system using sentiment analysis (SA)," *Automatika*, vol. 65, no. 2, pp. 410–424, Apr. 2024, doi: 10.1080/00051144.2023.2295148.

[18] H. B. Abdalla, M. Gheisari, and A. H. Awlla, "Hybrid self-attention Bi-LSTM and incentive learning-based CF for e-commerce recommendation systems," *Electronic Commerce Research*, vol. 25, no. 6, pp. 4947–4970, Dec. 2025, doi: 10.1007/s10660-024-09888-5.

[19] M. Ibrahim, I. S. Bajwa, N. Sarwar, F. Hajjej, and H. A. Sakr, "An intelligent hybrid neural CF approach for true recommendations," *IEEE Access*, vol. 11, pp. 64831–64849, 2023, doi: 10.1109/ACCESS.2023.3289751.

[20] A. Sami, W. El Adrousy, S. Sarhan, and S. Elmougy, "A deep learning based hybrid recommendation model for internet users," *Scientific Reports*, vol. 14, no. 1, Nov. 2024, doi: 10.1038/s41598-024-79011-z.

[21] C. Guo and X. Zhang, "Intelligent prediction of cross-border e-commerce customer satisfaction using deep learning embeddings," *IEEE Access*, vol. 12, pp. 173268–173278, 2024, doi: 10.1109/ACCESS.2024.3494776.

[22] L. Yao and N. Zheng, "Sentiment analysis based on improved transformer model and conditional random fields," *IEEE Access*, vol. 12, pp. 90145–90157, 2024, doi: 10.1109/ACCESS.2024.3418847.

[23] S. S. Roy, A. Kumar, and R. S. Kumar, "Metadata and review-based hybrid apparel recommendation system using cascaded large language models," *IEEE Access*, vol. 12, pp. 140053–140071, 2024, doi: 10.1109/ACCESS.2024.3462793.

[24] R. Garapati and M. Chakraborty, "Enhancing sentiment analysis and rating prediction using the review text granularity (RTG) model," *IEEE Access*, vol. 13, pp. 20071–20100, 2025, doi: 10.1109/ACCESS.2025.3534261.

[25] S. Yang, Q. Li, H. Lim, and J. Kim, "An attentive aspect-based recommendation model with deep neural network," *IEEE Access*, vol. 12, pp. 5781–5791, 2024, doi: 10.1109/ACCESS.2023.3349291.

[26] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based CF recommendation algorithms," in *Proceedings of the 10th international conference on World Wide Web*, New York, NY, USA: ACM, Apr. 2001, pp. 285–295, doi: 10.1145/371920.372071.

[27] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in *Advances in Neural Information Processing Systems 20 - Proceedings of the 2007 Conference*, 2008.

[28] J. McAuley and J. Leskovec, "Hidden factors and hidden topics," in *Proceedings of the 7th ACM conference on Recommender systems*, Oct. 2013, pp. 165–172, doi: 10.1145/2507157.2507163.

[29] C. Chen, M. Zhang, Y. Liu, and S. Ma, "Neural attentional rating regression with review-level explanations," in *Proceedings of the 2018 World Wide Web Conference on World Wide Web-WWW '18*, 2018, pp. 1583–1592. doi: 10.1145/3178876.3186070.

[30] Y.-H. Liu, Y.-L. Chen, and P.-Y. Chang, "A deep multi-embedding model for mobile application recommendation," *Decision Support Systems*, vol. 173, Oct. 2023, doi: 10.1016/j.dss.2023.114011.

[31] R. Cao, X. Zhang, and H. Wang, "A review semantics-based model for rating prediction," *IEEE Access*, vol. 8, pp. 4714–4723, 2020, doi: 10.1109/ACCESS.2019.2962075.
[32] Z. Cheng, Y. Ding, X. He, L. Zhu, X. Song, and M. Kankanhalli, "A3NCF: an adaptive aspect attention model for rating prediction," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, Jul. 2018, pp. 3748–3754, doi: 10.24963/ijcai.2018/521.
[33] M. Unger, A. Tuzhilin, and A. Livne, "Context-aware recommendations based on deep learning frameworks," *ACM Transactions on Management Information Systems*, vol. 11, no. 2, pp. 1–15, Jun. 2020, doi: 10.1145/3386243.

BIOGRAPHIES OF AUTHORS

Snehal Bhogan is a research scholar in the Karnatak Law Society Gogte Institute of Technology College in Belgaum, Karnataka. Currently working in Department of Computer Engineering at Agnel Institute of Technology and Design, Goa university. She has a Bachelor's degree in Computer Engineering from Goa University, Master's degree in Information Technology from Goa University. She has served in the education sector for more than 17 years at college level. Her area of interests in research are recommendation systems, deep learning and machine learning technologies. She can be contacted at email: snehalrbhogan@gmail.com, bhogansnehal@gmail.com.

Vijay S. Rajpurohit is a professor in the Department of Computer Science and Engineering at Karnatak Law Society Gogte Institute of Technology, Belgaum, Karnataka, India. He has got his B. E. in Computer Science and Engineering from Karnataka University Dharwad, M. Tech. from N.I.T.K Surathkal, and completed his Ph.D. from Manipal University in 2009. With an extensive background in image processing, cloud computing, and machine learning. Over the course of his illustrious career, he has authored more than 50 peer-reviewed journal publications, contributed to over 100 conferences, and penned few book chapters. He can be contacted at email: vijaysr2k@yahoo.com.

Sanjeev S. Sannakki is working as a professor in the Department of Computer Science and Engineering at Karnatak Law Society Gogte Institute of Technology, Belgaum, Karnataka, India. He pursued his B.E. in Electronics and Communication from Karnataka University Dharwad in 2004, M. Tech., and Ph.D. from VTU, Belagavi in 2009, and 2016 respectively. His research areas include image processing, cloud computing, computer networks, and data analytics. He has published a good number of papers in journals, International, and National conferences. He is also a life member of CSI and ISTE associations. He can be contacted at email: sannakkisanjeev@yahoo.co.in.