

XTRA: eXplainable Text-based Recommendation Algorithm

Hasan Serhat Gunduz¹, Aysun Bozanta^{2*}, Osman Yucel³

¹*Business Information Systems, Bogazici University, Istanbul, Turkey*

²*Management Information Systems Department, Bogazici University, Istanbul, Turkey*

³*Independent Researcher, New Jersey, United States of America*

Abstract – Recommendation systems built on ratings often fail to fully exploit the rich information contained in textual reviews and social-platform content. Text provides not only an additional signal for ranking, but also a natural basis for generating explanations. This study introduces XTRA, a rating-free, text-driven recommendation system that derives user relationships from shared review content and transforms sentiment expressed in text into personalized recommendation decisions. The proposed approach also incorporates explainability mechanisms to clarify both why users are considered similar and how textual sentiment contributes to the final recommendation. Experimental results on Amazon-Beauty, Yelp-Illinois, and IMDB-20K show that XTRA outperforms strong rating-based baselines on Yelp and IMDB, while remaining competitive on Amazon. These findings indicate that textual signals can serve as an effective alternative to ratings while also supporting transparent and interpretable recommendations. Implementation details are provided to support reproducibility and future extension.

Keywords – BERT, language models, recommendation systems, similarity, textual data.

I. INTRODUCTION

Recommendation systems aim to enhance user experience by predicting items of interest based on users' interactions with services or products [1]. They analyse behavioural patterns to deliver personalized and relevant suggestions [2].

Recommendation systems work by using explicit (ratings, reviews, etc.) or implicit (page views, purchases, etc.) feedback from the users [3]. Explicit feedback more accurately captures user satisfaction, as it reflects users' deliberate and intentional expressions. However, rating data is less representative because it relies solely on a numerical scale, which may not capture the full spectrum of user experiences [4]. To achieve more successful results and overcome data sparsity, different types of user and product data are used in recommendation systems [5]. Text data, especially user reviews, increases the success of recommendation systems by detailing users' perspectives [6].

Explainability in recommendation systems refers to the ability of these systems to clearly articulate the reasons behind their recommendations. Providing explainability is crucial because it enables transparency and accountability, allowing users to understand how their data is used and why specific

items are suggested [7]. This transparency fosters greater trust in the system, as users are more likely to accept recommendations when they can see the rationale behind them. Additionally, explainability can empower users to provide more targeted feedback and make informed decisions, further enhancing user satisfaction. However, many recommendation systems struggle to provide clear explanations because they rely solely on ratings or lack interpretable outputs.

The goal of this work is to show that text can stand in for ratings while preserving the collaborative benefits of user–user modelling and adding transparent, inspection-level explanations. XTRA is therefore introduced as a user-based collaborative filtering method that derives user similarity from language-model embeddings of reviews written for the same items and converts the sentiment of the most similar users into recommendation decisions through similarity-weighted aggregation. The objectives of this study are fourfold: (i) to demonstrate that a rating-free approach can match or surpass strong rating-based baselines across multiple domains; (ii) to study how the choice of text representation affects similarity estimation and downstream accuracy; (iii) to provide faithful, token-level explanations along two complementary axes – attention for why users are similar and Shapley Additive Explanations (SHAP) for how sentiment influences the decision; and (iv) to deliver a reproducible pipeline with clear preprocessing rules and reporting standards. Taken together, these aims position XTRA as a practical route to text-centric, explainable recommendation in settings where explicit ratings are scarce, unreliable, or strategically manipulated.

This study proposes XTRA, a text-driven recommendation framework that leverages user-generated reviews to model user preferences and generate personalized recommendations. To effectively capture semantic information from textual data, a wide range of language representation models are employed, including Bag of Words (BoW), Word to Vector (Word2Vec), Embeddings from Language Models (ELMo), Global Vectors for Word Representation (GloVe), Bidirectional Encoder Representations from Transformers (BERT), Distilled Bidirectional Encoder Representations from Transformers (Distil-BERT), Sentence BERT (SBERT), Robustly Optimized

* Corresponding author's e-mail: aysun.bozanta@bogazici.edu.tr
Article received 2026-03-09; accepted 2026-05-11

BERT Approach (RoBERTa), BERT Large Version (BERT Large), and SBERT Large Version (SBERT Large). The performance of the proposed approach is evaluated on three benchmark datasets – Amazon, Yelp, and IMDB – allowing a comprehensive comparison with state-of-the-art recommendation systems. In addition to its predictive capability, XTRA is designed with a strong focus on explainability. To this end, the attention mechanisms of transformer-based models are analysed using the Transformers-Interpret approach [8], while the SHAP method is utilised to identify the most influential words contributing to sentiment predictions [9]. This integrated explainability framework enables transparent recommendations by highlighting both the most relevant users and reviews, as well as the key textual components that drive the model's decisions.

This study aims to develop a novel and explainable recommendation system that can be an alternative to the existing rating-based recommendation systems and examines the following research questions:

- (1) How do recommendation systems that solely rely on user-generated textual data perform in comparison to rating-based recommendation systems?
- (2) What is the impact of employing different language models on the performance of similarity calculations in XTRA?
- (3) How can attention mechanisms and SHAP-based explanations improve the transparency of recommendation systems by providing interpretable insights into recommendation decisions?

The contributions of this study can be listed as follows:

- XTRA is introduced as a novel recommendation algorithm that eliminates the need for rating data while achieving performance comparable to, and in some cases surpassing, traditional rating-based systems.
- XTRA was rigorously evaluated on three diverse datasets, demonstrating its robustness as a viable alternative to conventional recommendation methods. This establishes its potential for broader application across various domains and problem settings.
- XTRA was tested with ten different language models to assess its modularity and flexibility. These evaluations identified the most effective models for generating high-quality recommendations.
- XTRA incorporates two explainability components: (1) The Transformers-Interpret approach to elucidate why two users are deemed similar, and (2) SHAP-based sentiment analysis to clarify how a reviewer's sentiment influences product recommendations. These components enhance transparency and trust in the recommendation process.

The rest of this study is organised as follows: Section 2 covers related work, and Section 3 details the methodology, including datasets, base algorithms, development of proposed algorithm, and the explainability aspect of the system. Section 4 presents the results, followed by a discussion in Section 5. Section 6 concludes with a summary, contributions, limitations, and future research areas.

II. LITERATURE REVIEW

The literature on recommendation systems has evolved significantly, exploring various approaches to enhance accuracy, personalization, and explainability. This section provides an overview of key methodologies, including Collaborative Filtering, Content-Based Filtering, and Hybrid models, along with recent advancements in leveraging textual data and explainability techniques to improve recommendation quality. Collaborative filtering generates recommendations based on the similarity between users' preferences [1], [10]. This method assumes that users with similar preferences in the past are likely to have similar preferences in the future. Content-based filtering focuses on the similarity between user profiles and item features [1]. Users have suggested products that are similar to the products they have preferred or rated highly in the past [1]. Hybrid systems aim to minimize the disadvantages of both methods by combining the strengths of collaborative filtering and content-based filtering methods [11].

Deep learning algorithms are capable of processing diverse data types, making them highly effective for generating successful recommendations [12]. Language models, particularly transformer-based models have demonstrated remarkable success in the field of recommendation systems by having their advanced capabilities in understanding and processing textual data [13]. However, language models require large datasets and significant computational resources to deliver high-quality results in recommendation systems [13]. The integration of Large Language Models (LLMs) and prompt learning into recommendation systems has been another trending area of study in the field of recommendation systems in recent years. According to Zhao et al., LLMs have introduced capabilities in understanding and generating humanlike text, enabling systems to better capture nuanced user preferences and provide context-aware recommendations [14].

Explainability in recommendation systems is crucial for ensuring transparency, and user trust, particularly as these systems play an increasingly influential role in decision-making processes [15]. Traditional recommendation models, such as collaborative filtering, often operate as "black boxes," making it difficult for users to understand why certain recommendations are made [16]. Lack of explainability can lead to biases, reduced user satisfaction, and diminished accountability in critical domains like healthcare and finance [15]. Explainable recommendation systems leverage techniques such as natural language explanations and attention mechanisms to provide interpretable outputs, improving user engagement and trust [12]. Moreover, enhancing explainability aligns with ethical AI principles by mitigating potential biases embedded in data-driven recommendations [7].

A. Language Models

In recent years, many language models have been developed to process text-based data. The BoW model is one of the simplest and most widely used methods for text representation. It represents texts as words and their frequencies within a document [17]. However, BoW cannot capture word order, syntactic structure, or semantic relationships [17]. In contrast to

BoW, Word2Vec represents a significant improvement by using deep learning models to generate vector representations of words, and places semantically related words closer together in the vector space [18]. This enables Word2Vec to capture semantic relationships and word analogies [18].

GloVe uses a local context window to learn word representations and relies on global word co-occurrence statistics, unlike Word2Vec [19]. By leveraging both local and global information, GloVe provides a more robust representation of word meanings [19]. ELMo generates context-dependent word embeddings by training a bidirectional LSTM. Unlike traditional static embeddings, ELMo determines word meanings based on their context, making it effective for words with multiple meanings [20]. ELMo has outperformed its predecessors on many language processing tasks namely sentiment analysis, named entity recognition, and question answering [20].

BERT pre-trains deep language representations by utilising a bidirectional transformer architecture [21]. It predicts missing words in a sentence (masked language modeling) and sentence relationships (next sentence prediction), achieving greater effectiveness in tasks of text classification, sentiment analysis, and question answering [21]. SBERT is a model based on the BERT model that generates sentence embeddings instead of word embeddings [22]. This makes SBERT especially useful for semantic textual similarity, paraphrase identification, and sentence clustering [22]. RoBERTa improves BERT by optimizing its training process, achieving superior performance in many natural language processing tasks due to its different training processes [23]. Pre-trained language models, namely BERT, RoBERTa, and GPT, have been widely tested in recommendation systems. The results show that the models can be adapted for recommendation purposes and are promising, but that these models need to be customized and biases reduced [13].

B. Language Models in Recommendation Systems

Designing recommendation systems for specific domains using language models is an increasingly popular area of research. Numerous studies have leveraged language models – particularly BERT and its derivatives – to develop recommendation systems across diverse domains, such as location-based services [24], [25], academic collaboration networks [26], [27], and movie recommendations [28]. Moreover, several comprehensive reviews have examined the use of language models in recommendation systems, underscoring their growing importance and potential to enhance recommendation accuracy and user satisfaction [14], [29]–[32].

Movie recommendation algorithm based on sentiment analysis and Latent Dirichlet Allocation (LDA) presents the topics and emotional tendencies of movie reviews. It improves content-based recommendation algorithms by enriching the user interest model and product feature model based on emotional tendencies [33]. Unlike previous studies, it uses BERT to train sentiment classification models and LDA to extract topics [33]. The developed approach is tested on Douban

movie reviews, and it is observed that it improves the recommendation diversity [33].

The sentiments extracted from text data offer valuable insights for enhancing recommendation systems. Zhang and Zhang demonstrated this by using BERT to extract topics and emotional tendencies from movie reviews, which they then used to develop both a user interest model and a product feature model [33]. These models informed a content-based recommendation algorithm that combined sentiment analysis with the LDA topic model [33].

UNBERT is a user-news matching model that enhances text representation by using a pre-trained BERT model [34]. Besides improving text representation, UNBERT captures user-news matching signals at both the word and news levels. The model has been tested on the Microsoft News Dataset (MIND), and the results highlight a significant improvement in the accuracy of news recommendations using language models [34].

Another study explores the use of pre-trained language models (PLMs) such as BERT and GPT for recommendation systems [13]. Tests on a movie recommendation dataset show that PLMs outperform random base models in the zero-shot setting [13]. However, PLMs exhibit strong linguistic bias [13]. Fine-tuned PLMs reduce these biases but underperform compared to traditional recommendation models such as GRU4Rec [13]. The findings highlight both the potential and challenges of using PLMs in recommendation systems [13].

LLMs, with their strong language understanding, generation, and generalization capabilities, have shown considerable potential for enhancing recommender systems [14]. Although the performance of these systems has improved, challenges remain in their interpretability and explainability [14]. Explainable recommendation systems are designed to help users understand why the recommendations made by the algorithm are recommended [15]. In this way, it helps increase the transparency, persuasiveness, effectiveness, reliability, and satisfaction of user systems [15].

The study named MULTI-BERT presents a new document embedding method by using SBERT with NLP techniques [35]. It detects both intra-sentence and intersentence relations by processing sentences as tokens [35]. It has been tested on book recommendations using the MULTI-BERT model on the Goodreads dataset and compared with SBERT according to the precision metric [35]. The MULTI-BERT model outperforms SBERT in embedding quality [35].

The study integrates textual and visual information on the reviews, where the RoBERTa model is used for textual data and the VGG-16 model is used for visual data [36]. The generated approach has been tested on the Amazon dataset, and the recommendation accuracy has been significantly improved [36].

C. Explainability in Recommendation Systems

Despite being widely used, recommendation systems often operate as "black boxes" that do not explain the decision-making process behind the recommendation [37]. This lack of transparency can decrease trust and engagement in

recommendation systems when the recommendations are inaccurate [15], [37]. Therefore, there is increasing interest in the area of explainability studies that make recommendation systems transparent. SHAP explains the model's decisions by calculating the contribution of each feature to the model's prediction [38]. SHAP provides an explanation of particularly complex model outputs, and these explanations help users better comprehend the model's decision-making process [38]. When integrated with sentiment analysis of text representation models, SHAP demonstrates the impact of each word on the overall sentiment in texts [9]. Thus, explanations of the outputs of text-based systems are provided. This approach helps uncover hidden biases or inconsistencies [9].

In addition, explainability studies are also conducted in transformer-based language models. BERTViz is a tool that visualizes attention mechanisms in BERT models [39]. It provides a deeper understanding of the effects of different words or expressions on a given sentence [39]. This tool can be used to explain text-based recommendation systems built with BERT [39].

Another study proposes the Attentive Recurrent Neural Network (Ante-RNN) that integrates textual and visual information for explainability [40]. By utilising topical and contextual attention mechanisms, Ante-RNN detects users' interests, making the system transparent and explainable [40]. The proposed approach was tested on Amazon and Pinterest datasets, demonstrating strong performance [40].

Recently, explainability has been further advanced in this field with the integration of LLMs in recommendation systems. A study in this direction is the Chat-Rec system, which augments traditional recommendation systems and introduces explainability using ChatGPT [41]. According to the results of the study, Chat-Rec significantly improved the recommendation results and performed well in zero-shot rating prediction tasks [41].

TABLE I
LITERATURE REVIEW SUMMARY

AUTHORS	DATASETS	LANGUAGE MODELS	EXPLAINABILITY
Liu, et al. (2020) [40]	MovieLens, Pinterest	Ante-RNN	✓
Zhang Y., et al. (2021) [13]	MovieLens	BERT, GPT2	X
Zhang Q., et al. (2021) [34]	MIND	BERT	X
Li, et al. (2021) [42]	Amazon, Yelp, TripAdvisor	PETER	✓
Zhang, et al. (2022) [33]	Movie Data	BERT	X
Javaji, et al. (2023) [35]	Goodreads	RoBERTa, SBERT	X
Jeong, et al. (2024) [36]	Amazon	RoBERTa, VGG-16	X
XTRA	Amazon, Yelp, IMDB	Bag-of-Words, Word2Vec, GloVe, ELMo, BERT, DistilBERT, RoBERTa, SBERT	✓

Personalized Transformer for Explainable Recommendation (PETER) is a novel model that can generate both recommendations and explanations [42]. PETER is a small, untrained Transformer with two layers that produces personalized and explainable recommendations [42]. In addition, PETER presents recommendation explanations in textual form [42]. The model has achieved strong results in terms of both explainability and text quality metrics [42].

D. Most Related Work

Table I presents a comparative summary of this study alongside recent literature utilising textual data and language models. For comparison, relevant studies focusing on similar methodologies were selected. The comparison criteria include datasets, algorithms, and explainability. This study distinguishes itself by testing the developed recommendation system on a broader range of datasets and algorithms than previous works. Furthermore, the number of studies that integrate explainability into systems using textual data remains limited, highlighting the novelty and significance of proposed approach.

III. METHODOLOGY

The methodology of the study is discussed in four subsections: data collection and preprocessing, base recommendation systems, recommendation system development, and explainability of the recommendation system.

A. Data Collection and Preprocessing

In this section, the datasets are identified, acquired, and preprocessed to ensure their suitability for recommendation systems. Amazon [43], YELP [44], and IMDB [45] datasets were used for this study. These datasets were selected to enhance the generalizability of the findings by encompassing different types of businesses. The selection criteria required that each dataset includes user, item, rating, and review attributes to ensure consistency in analysis. To optimize processing efficiency and computational resources, the dataset sizes were reduced by filtering or eliminating specific subcategories from the raw data. The study utilised the Beauty category from the Amazon dataset, accommodation data from Illinois in the Yelp dataset, and a random sample of 20 000 users from the IMDB dataset (Table II).

TABLE II
SELECTED SUBCATEGORY RAW DATASET DIMENSIONS

	AMAZON BEAUTY	YELP-ILLINOIS	IMDB (20K USERS)
Reviews	701.528	51.832	45.121
Unique Users	631.986	22.248	20.000
Unique Items	115.709	2.145	1.150

Data elimination was performed on three datasets according to the following conditions: 1) Comments made in languages other than English were removed; 2) Data with empty text (review) columns were deleted; 3) Rows containing null values

were excluded from the dataset; 4) Since the similarity between users would be determined from the similarity between their reviews of the same product, users with more than five reviews and their comments were selected. Users with fewer reviews and their data were deleted. After the eliminations are made, the dimensions and information of the data are available in Table III.

TABLE III

DIMENSIONS OF DATASETS AFTER THE DATA ELIMINATION PROCESS

	AMAZON BEAUTY	YELP-ILLINOIS	IMDB (20K USERS)
Reviews	11.287	19.418	45.121
Unique Users	966	1.381	791
Unique Items	5.377	2.018	1.141

To create the training and test sets, each user's data was sorted chronologically, with the first 70 % retained in the training set and the last 30 % assigned to the test set.

B. Base Recommendation Systems

Matrix factorization (MF) algorithms, specifically Singular Value Decomposition (SVD) and Truncated Singular Value Decomposition (T-SVD), along with K-Nearest Neighbour (KNN), were selected and applied to all datasets. Each algorithm was used to generate recommendations based on user ratings. Recommendation systems using the T-SVD algorithm produce recommendations by selecting the 5 most similar users. In the recommendation system established with the KNN algorithm, similarity between users was calculated through cosine similarity, and the 5 nearest users were selected exhaustively searching the whole user base [46], [47].

C. XTRA

This study presents a review and sentiment-based recommendation system that identifies user similarity by analysing the content of their reviews. Given that user reviews are text-based, text representation models were employed to transform textual data into numerical vectors. The models utilised include BoW, Word2Vec, ELMo, GloVe, BERT, BERT (large version), DistilBERT, SBERT, SBERT (large version), and RoBERTa [18], [19]. Once the numerical vectors were generated, the similarity between users was computed using the Cosine similarity metric [48]. The similarity score between users is calculated by averaging the similarity scores of their reviews for the same products. This process is explained in Fig. 1.

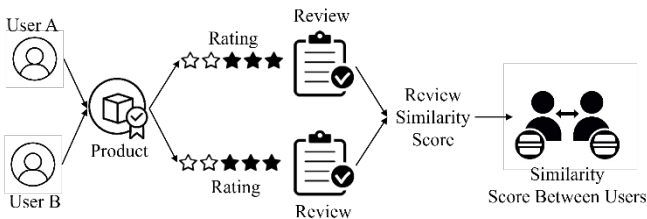


Fig. 1. Calculation of similarity.

Let R_A denote the set of items reviewed by user A , and $rev_{A,i}$ denote the review written by user A for item i . The similarity between users A and B is calculated as the average cosine similarity of their reviews for items they have both reviewed (see (1)):

$$\text{sim}_{A,B} = \frac{\sum_{i \in R_A \cap R_B} \text{cosineSim}(rev_{A,i}, rev_{B,i})}{|R_A \cap R_B|}. \quad (1)$$

This equation provides a normalized measure of similarity between users A and B by averaging the cosine similarity scores of their corresponding reviews on common items.

To determine the sentiment of each user review, sentiment analysis was conducted using a fine-tuned BERT model [49]. Each review was classified as either positive or negative, and a corresponding `bertsentiment` variable was created, where positive reviews were assigned a value of 1 and negative reviews were assigned -1. These sentiment scores, derived from the BERT model, were then utilised to enhance product recommendations among similar users.

For the target user, those with similarity scores above a predefined threshold, α , were identified as similar users. This threshold value was selected based on preliminary experiments, which showed that $\alpha = 0.50$ provides a favourable balance between recommendation performance and user coverage. All products previously purchased by these similar users were added to the product recommendation pool.

Next, the predicted sentiment for the target user T is computed by aggregating the sentiment scores of similar users, weighted by the similarity scores between T and those users (see (2)):

$$\text{predictedSentiment}_T = \frac{\sum_{U \in \text{similarUsers}_T} \text{sim}(T,U) \cdot \text{sentiment}_U}{\sum_{U \in \text{similarUsers}_T} \text{sim}(T,U)}. \quad (2)$$

In this equation, similarUsers_T represents the set of users similar to T , $\text{sim}(T,U)$ is the similarity score between user T and user U , and sentiment_U is the sentiment score for user U . XTRA recommends an item to the user when the predicted sentiment score is positive, indicating a favourable evaluation (Fig. 2).

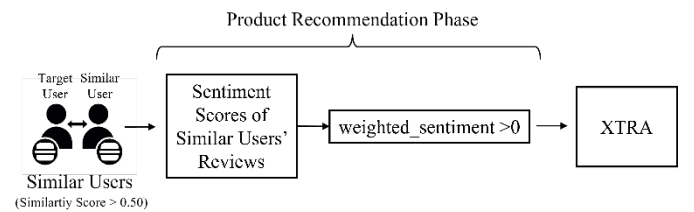


Fig. 2. Recommendation flow.

D. Evaluation Metrics

A confusion matrix was created to evaluate the performance of all recommendation systems. Two flags were used: `recommenderflag` and `likeflag`. The explanation of the confusion matrix (Table IV) is as follows: If the prediction is 1, it means that the item is recommended to the target user, while

a prediction of 0 means that the item is not recommended. Similarly, if the actual value is 1, it means that the item is liked by the target user, whereas an actual value of 0 means that the item is not liked by the target user. Accuracy, precision, recall, and F1 score were used together to evaluate the performance of the recommendation systems comprehensively.

TABLE IV
CONFUSION MATRIX

		PREDICTION	
		1	0
ACTUAL	1	TP (True Positive)	FN (False Negative)
	0	FP (False Positive)	TN (True Negative)

These metrics were selected because each captures a different aspect of recommendation performance, and relying on a single metric may lead to an incomplete or misleading interpretation, particularly when the distribution of relevant and non-relevant recommendations is imbalanced. Accuracy provides an overall measure of correct predictions; however, it may be insufficient on its own, since a model can achieve high accuracy while still failing to identify relevant items effectively. Therefore, precision and recall were also included to assess recommendation quality from different perspectives. Precision reflects the reliability of the generated recommendations by measuring how many predicted relevant items are relevant, while recall indicates the model's ability to capture user interests by measuring how many truly relevant items are successfully identified. Since precision is more sensitive to false positives and recall is more sensitive to false negatives, they complement each other in evaluating the trade-off between recommendation reliability and coverage. Finally, the F1 score was used as a balanced metric that combines precision and recall through their harmonic mean, preventing the evaluation from favouring a model that performs well on only one dimension. Thus, these metrics were interpreted jointly to provide a fairer, more robust, and multidimensional comparison with baseline recommendation methods.

E. Explainability Methods

This study enhances transparency using two explainability approaches. First, the determination of user similarity is explained; second, the sentiment analysis process is clarified.

Pairs of users who have reviewed the same product are identified, and an attention mechanism from BERT-based models [8] is then applied to compute attention scores for each word in their reviews. By highlighting the words with the highest attention, the textual features driving user similarity are revealed. As illustrated in Fig. 3, the main user's review of a product (Product A) is compared with reviews from potentially similar users. The attention mechanism extracts the most influential words (e.g., "Impressed," "Quality," "Buy," "Excellent"), which collectively contribute to the overall similarity score. This score is computed by averaging similarity across all shared products, offering a transparent view of why certain users are considered similar.

In addition, a pre-trained BERT model is integrated with SHAP to visually display the sentiment of user reviews [8]. SHAP identifies and quantifies the impact of individual words, allowing us to pinpoint which terms lead a review to be classified as positive or negative. By highlighting the influential words within each review, this method provides a clear rationale for sentiment outcomes, further enhancing transparency and user trust.

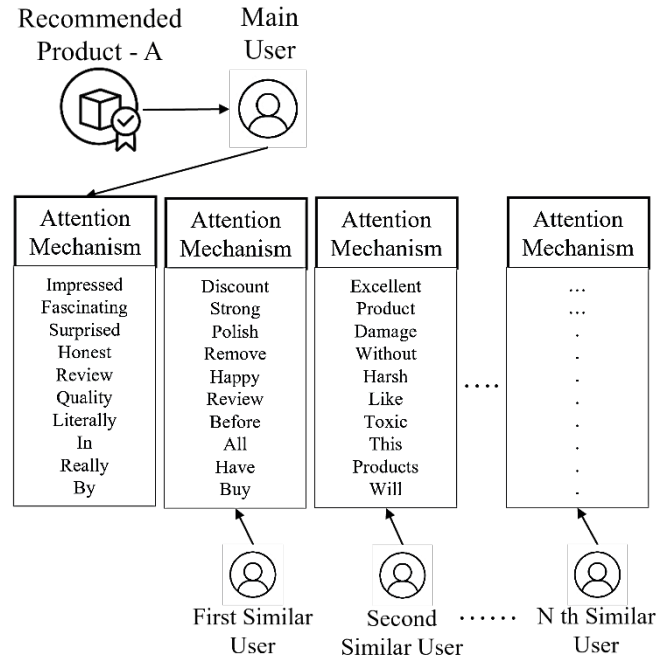


Fig. 3. Overview of the attention-based user similarity approach.

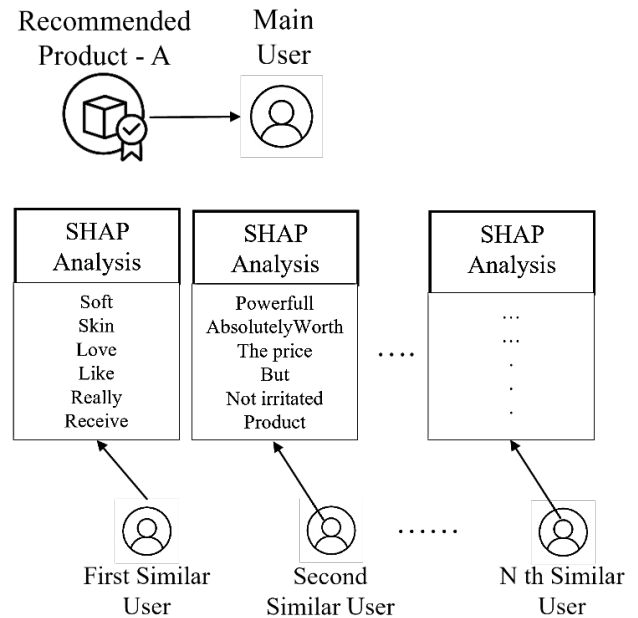


Fig. 4. Explanation of recommended product.

As illustrated in Fig. 4, similar users' reviews of the recommended product (Product A) are analysed, and SHAP highlights the keywords that drive each review's sentiment –

whether positive or negative – and ultimately influence the recommendation decision.

Together, these methods enhance the transparency of recommendation system by offering detailed, interpretable explanations for both the user similarity and sentiment evaluation processes.

IV. RESULTS

In this section, the performance of different recommendation algorithms is first compared, followed by an evaluation of the results obtained from the explainability approach.

A. Performance Comparison of the Algorithms

In this section, the performance of the algorithms was compared in terms of accuracy, F1 score, precision, and recall. The success of each model was tested using Amazon, Yelp, and IMDB datasets.

Ten different versions of XTRA were developed using various language models. The results obtained from both the baseline algorithms and the XTRA variants on the Amazon dataset are presented in Table V.

For baseline recommendation systems on the Amazon dataset, the system utilising the SVD model achieved the highest performance in three out of four metrics. Among XTRA variants, the model incorporating SBERT outperformed all other variants. Notably, the SBERT-based system demonstrated competitive performance, closely matching the effectiveness of the SVD-based system, which relied on rating data.

The results obtained from both the baseline algorithms and the XTRA variants on the Yelp dataset are presented in Table VI.

For recommendation systems built using rating data on the Yelp dataset, the baseline system utilising the KNN model achieved the highest performance across three out of four metrics. Among the XTRA versions, the model incorporating BoW emerged as the top performer, followed closely by SBERT and SBERT Large. Notably, these models outperformed traditional rating-based approaches.

TABLE V

THE PERFORMANCE OF THE ALGORITHMS ON THE AMAZON DATA

	ALGORITHM	ACCURACY	F1 SCORE	PRECISION	RECALL
BASE ALGORITHMS	SVD	0.767	0.855	0.851	0.858
	KNN	0.746	0.846	0.842	0.849
	T-SVD	0.746	0.847	0.834	0.861
VERSIONS OF XTRA	BoW	0.621	0.748	0.766	0.731
	Word2Vec	0.697	0.806	0.811	0.801
	ELMo	0.696	0.805	0.811	0.800
	GloVe	0.697	0.806	0.811	0.801
	BERT	0.697	0.806	0.811	0.801
	DistilBERT	0.697	0.806	0.811	0.801
	SBERT	0.724	0.824	0.824	0.824
	RoBERTa	0.695	0.805	0.809	0.802
	BERT Large	0.698	0.807	0.812	0.802
	SBERT Large	0.720	0.824	0.826	0.822

TABLE VI

THE PERFORMANCE OF THE ALGORITHMS ON THE YELP DATASET

	ALGORITHM	ACCURACY	F1 SCORE	PRECISION	RECALL
BASE ALGORITHMS	SVD	0.574	0.530	0.845	0.386
	KNN	0.599	0.687	0.664	0.711
	T-SVD	0.593	0.665	0.677	0.653
VERSIONS OF XTRA	BoW	0.656	0.765	0.726	0.808
	Word2Vec	0.641	0.749	0.704	0.799
	ELMo	0.641	0.749	0.704	0.799
	GloVe	0.641	0.749	0.705	0.799
	BERT	0.641	0.749	0.704	0.799
	DistilBERT	0.641	0.749	0.704	0.799
	SBERT	0.650	0.760	0.724	0.799
	RoBERTa	0.641	0.749	0.704	0.799
	BERT Large	0.642	0.749	0.705	0.799
	SBERT Large	0.648	0.756	0.710	0.808

TABLE VII

THE PERFORMANCE OF THE ALGORITHMS ON THE IMDB DATASET

	ALGORITHM	ACCURACY	F1 SCORE	PRECISION	RECALL
BASE ALGORITHMS	SVD	0.658	0.705	0.903	0.579
	KNN	0.637	0.739	0.751	0.727
	T-SVD	0.646	0.754	0.742	0.765
VERSIONS OF XTRA	BoW	0.659	0.769	0.716	0.832
	Word2Vec	0.666	0.778	0.716	0.852
	ELMo	0.666	0.778	0.716	0.852
	GloVe	0.666	0.778	0.716	0.852
	BERT	0.667	0.779	0.716	0.853
	DistilBERT	0.666	0.778	0.716	0.853
	SBERT	0.660	0.772	0.723	0.828
	RoBERTa	0.666	0.779	0.716	0.854
	BERT Large	0.666	0.778	0.716	0.853
	SBERT Large	0.670	0.781	0.720	0.854

On the IMDB dataset, the performance of the recommendation systems was analysed in detail, as presented in Table VII. The baseline recommendation system utilising the rating variable and the SVD model achieved the highest precision and accuracy, while the system using the rating variable and the T-SVD model performed best in terms of F1 score and recall among the three baseline models. SBERT Large shows marginally higher performance. Other BERT-based models were closely followed, with minimal performance differences among them.

B. Results of Explainability of Recommendation Systems

In addition to the recommendation system, this study introduces an explainability framework aimed at providing transparent explanations and clarifying the reasoning behind each recommendation.

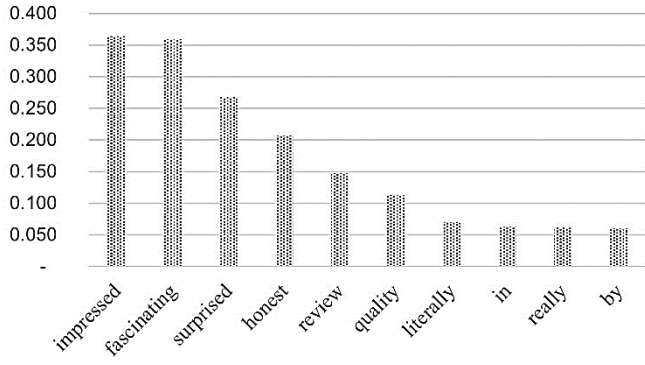


Fig. 5. Main User's important words for Product A.

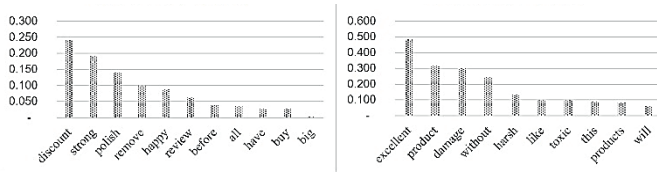


Fig. 6. Similar User 1 and 2's important words for Product A.

Using attention mechanisms, the system pinpoints the most influential words in user reviews to explain why certain users are deemed similar. As illustrated in Fig. 5, the main user's top attention-weighted words for Product A (Fig. 5) are compared with those of two similar users (Fig. 6). Each bar represents the attention score of a word, indicating how strongly it contributes to the user's overall review representation. The developed approach found User 1 to be 0.78 similar and User 2 to be 0.82 similar. By examining these high-impact words, the system provides a transparent rationale for the computed similarity between users. This approach helps end-users understand why they are deemed similar to other reviewers.

A pre-trained BERT model combined with SHAP is applied to highlight how specific words shape the sentiment of each review. By quantifying the contribution of individual words, SHAP reveals which terms have the most substantial positive or negative influence on the overall sentiment.

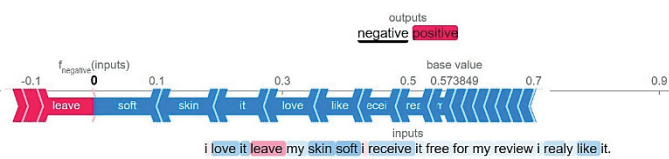


Fig. 7. Sentiment analysis of the first similar user's review.

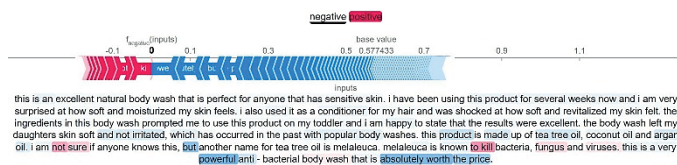


Fig. 8. Sentiment analysis of the second similar user's review.

As illustrated in Figs. 7 and 8, the reviews of two similar users are analysed to determine why they influenced the recommendation of the product to the target user. The words that contribute positively and negatively to the sentiment for

each similar user are shown. This granular breakdown allows end-users to understand precisely how similar users' sentiment was evaluated.

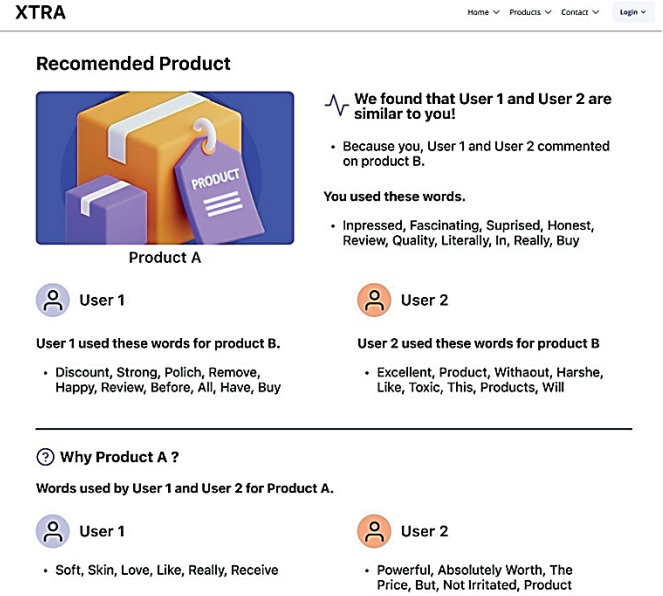


Fig. 9. Product recommendation user interface.

The product recommendation user interface, depicted in Fig. 9, demonstrates how the system conveys its explainability approach directly to the end user. First, the interface highlights the keywords that establish a similarity between the main user and the similar users. Next, it presents the most influential words from similar users' reviews, showing how these terms drive the sentiment analysis and ultimately contribute to the recommendation decision.

V. DISCUSSION

This study presents an explainable recommendation system that relies solely on user-generated reviews and is competitive with rating-based recommendation systems. Furthermore, the effects of different language models on performance, and the contributions of the attention mechanism and SHAP methods to explainability are examined in the light of research questions.

RQ 1: How do recommendation systems that solely rely on user-generated textual data perform in comparison to rating-based recommendation systems?

The review and sentiment-based recommendation system developed in this study outperformed traditional rating-based recommendation systems on the IMDB and Yelp datasets, while also achieving comparable results on the Amazon dataset. This is consistent with prior studies on the importance of the richness of textual data in capturing user preferences [13]. The performance of the models varied due to the differing characteristics of user reviews in each dataset. Factors such as the length, grammar structure, semantic depth, and emotional richness of user comments differed across platforms, influencing the performance of the models [50].

RQ 2: What is the impact of employing different language models on the performance of similarity calculations in a sentiment-based recommendation system?

The language models employed in the study produced varying results in detecting text similarities, which, in turn, influenced the performance of the recommendation systems. The SBERT model generally performed better than other language models, owing to its ability to capture semantic meaning at the sentence level [22]. SBERT delivered the best results in datasets with well-structured reviews, such as IMDB. However, datasets where reviews contained shorter and less carefully constructed sentences, such as the Yelp, Bag of Words, a word-level language model [17], performed best, with SBERT ranking second.

RQ 3: How can attention mechanisms and SHAP-based explanations improve the transparency of recommendation systems by providing interpretable insights into recommendation decisions?

The SHAP approach revealed the influence of each word on the sentiment score, showing whether words positively or negatively impacted user satisfaction [38]. This made the sentiment score, which reflects user satisfaction, more transparent. Additionally, the attention mechanism, analysed using the Transformer-Interpret approach [8], identified the words that most influenced the similarity calculations between users. This allowed system developers to determine the reasons behind user similarities. The integration of these two methods provided step-by-step explanations for each recommendation, thereby enhancing confidence in the system.

In summary, the results show that a rating-free, review and sentiment-driven collaborative filter (XTRA) can rival strong rating-based baselines across three domains – surpassing them on Yelp and IMDB and remaining competitive on Amazon. By pairing attention-based similarity rationales with SHAP attributions for sentiment, XTRA surfaces token-level evidence for both neighbourhood formation and the final decision, improving transparency and perceived trustworthiness [15]. Beyond accuracy, the text-centric design reduces reliance on explicit ratings and is resilient under sparse or strategically distorted feedback. XTRA is viewed as a practical step toward accountable recommendation systems and as a foundation for future work on explanation calibration, user studies, and cross-domain transfer.

VI. CONCLUSION

This study aimed to develop a novel explainable recommendation system that used only text data and could compete with the base recommendation systems created with rating data. The created recommendation systems were tested on Amazon, Yelp, and IMDB datasets. Among the XTRA versions, the one with SBERT outperformed all other models on the IMDB dataset. In the Yelp dataset, SBERT-based XTRA performed as the best model after the BoW-based model. On the Amazon dataset, SBERT-based XTRA demonstrated competitive performance.

Another focus of this study was the explainability of the recommendations. An explainability approach was developed

using the SHAP and Transformers-Interpret packages, allowing the reasons for each recommendation to be explained. As a result, this study showed that XTRA, which utilised only text data, could compete with base recommendation systems using rating data and produce explainable results.

Besides the scientific contributions of this study, there were also limitations it encountered. The first limitation was the use of data filtering and elimination steps to reduce processing time, due to limited computational power. If computation constraints were overcome with more advanced hardware, the developed recommendation systems could be tested on larger datasets without filtering. Additionally, the recommendation systems developed in this study could be tested on new datasets, making the results even more generalizable. Another limitation was that only user reviews were utilised, which led to a cold start problem in the recommendation systems. To overcome this issue, future work could focus on incorporating text-based product data, such as product descriptions, and developing a hybrid or content-based recommendation system.

Furthermore, the sentiment analysis in this study categorized user comments as either positive or negative, without accounting for more nuanced emotions such as anger, sadness, or happiness. Incorporating a broader range of emotional states could potentially enhance the system's understanding of user sentiment and lead to more refined recommendations.

REFERENCES

- [1] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, June 2005. <https://doi.org/10.1109/TKDE.2005.99>
- [2] J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, "Recommender system application developments: A survey," *Decis. Support Syst.*, vol. 74, pp. 12–32, June 2015. <https://doi.org/10.1016/j.dss.2015.03.008>
- [3] A. Ansari, S. Essegiaier, and R. Kohli, "Internet recommendation systems," *Journal of Marketing Research*, vol. 37, no. 3, Aug. 2000. <https://doi.org/10.1509/jmkr.37.3.363.18779>
- [4] A. Ghose and P. G. Ipeirotis, "Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 10, pp. 1498–1512, Sep. 2010. <https://doi.org/10.1109/TKDE.2010.188>
- [5] G. Adomavicius and J. Zhang, "Impact of data characteristics on recommender systems performance," *ACM Trans. Manag. Inf. Syst.*, vol. 3, no. 1, pp. 1–17, 2012. <https://doi.org/10.1145/2151163.2151166>
- [6] S. Kanwal, S. Nawaz, M. K. Malik, and Z. Nawaz, "A review of text-based recommendation systems," *IEEE Access*, vol. 9, pp. 31638–31661, Feb. 2021. <https://doi.org/10.1109/ACCESS.2021.3059312>
- [7] B. Abdollahi and O. Nasraoui, "Transparency in fair machine learning: The case of explainable recommender systems," in *Human and Machine Learning: Visible, Explainable, Trustworthy and Transparent*, June 2018, pp. 21–35. https://doi.org/10.1007/978-3-319-90403-0_2
- [8] C. D. Pierse, "Transformers-Interpret," *GitHub repository*. [Online]. Available: <https://github.com/cdpierse/transformers-interpret>
- [9] A. Diwali, K. Saeedi, K. Dashtipour, M. Gogate, E. Cambria, and A. Hussain, "Sentiment analysis meets explainable artificial intelligence: A survey on explainable sentiment analysis," *IEEE Trans. Affect. Comput.*, vol. 15, no. 3, July 2023. <https://doi.org/10.1109/TAFFC.2023.3296373>
- [10] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 5–53, Jan. 2004. <https://doi.org/10.1145/963770.963772>
- [11] R. Burke, "Hybrid recommender systems: Survey and experiments," *User Model. User-Adapt. Interact.*, vol. 12, pp. 331–370, Nov. 2002. <https://doi.org/10.1023/A:1021240730564>

- [12] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *ACM Comput. Surv.*, vol. 52, no. 1, pp. 1–38, Feb. 2019. <https://doi.org/10.1145/3285029>
- [13] Y. Zhang, H. Ding, Z. Shui, Y. Ma, J. Zou, A. Deoras, and H. Wang, "Language models as recommender systems: Evaluations and limitations," in *NeurIPS 2021 ICBINB Workshop*, 2021.
- [14] Z. Zhao *et al.*, "Recommender systems in the era of large language models (LLMs)," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 11, Nov. 2024. <https://doi.org/10.1109/TKDE.2024.3392335>
- [15] Y. Zhang and X. Chen, "Explainable recommendation: A survey and new perspectives," *Found. Trends Inf. Retr.*, vol. 14, no. 1, pp. 1–101, Mar. 2020. <https://doi.org/10.1561/15000000066>
- [16] N. Tintarev and J. Masthoff, "Explaining recommendations: Design and evaluation," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, and B. Shapira, Eds. Springer, 2015, pp. 353–382. https://doi.org/10.1007/978-1-4899-7637-6_10
- [17] E. Gabrilovich and S. Markovitch, "Overcoming the brittleness bottleneck using Wikipedia: Enhancing text categorization with encyclopedic knowledge," in *Proc. AAAI*, 2006, pp. 1301–1306.
- [18] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *arXiv:1310.4546*, Oct. 2013. <https://doi.org/10.48550/arXiv.1310.4546>
- [19] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," in *Proc. EMNLP*, Doha, Qatar, Oct. 2014, pp. 1532–1543. <https://doi.org/10.3115/v1/D14-1162>
- [20] J. Sarzynska-Wawer, A. Wawer, A. Pawlak, J. Szymanowska, I. Stefaniak, M. Jarkiewicz, and L. Okruszek, "Detecting formal thought disorder by deep contextualized word representations," *Psychiatry Res.*, vol. 304, Oct. 2021, Art. no. 114135. <https://doi.org/10.1016/j.psychres.2021.114135>
- [21] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, Minneapolis, USA, June 2019, pp. 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- [22] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," *arXiv preprint arXiv:1908.10084*, Aug. 2019. <https://doi.org/10.48550/arXiv.1908.10084>
- [23] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," *arXiv preprint arXiv:1907.11692*, July 2019. <https://doi.org/10.48550/arXiv.1907.11692>
- [24] A. Noorian, "A BERT-based sequential POI recommender system in social media," *Comput. Stand. Interfaces*, vol. 87, Jan. 2024, Art. no. 103766. <https://doi.org/10.1016/j.csi.2023.103766>
- [25] N. Darraz, I. Karabila, A. El-Ansari, N. Alami, and M. ElMallahi, "Integrated sentiment analysis with BERT for enhanced hybrid recommendation systems," *Expert Syst. Appl.*, vol. 261, Feb. 2025, Art. no. 125533. <https://doi.org/10.1016/j.eswa.2024.125533>
- [26] N. Yang, J. Jo, M. Jeon, W. Kim, and J. Kang, "Semantic and explainable research-related recommendation system based on semi-supervised methodology using BERT and LDA models," *Expert Syst. Appl.*, vol. 190, Mar. 2022, Art. no. 116209. <https://doi.org/10.1016/j.eswa.2021.116209>
- [27] H. Thangaraj, V. Vashisht, E. Joshi, K. Verma, and D. Katariya, "A BERT based hybrid recommendation system for academic collaboration," *arXiv preprint arXiv:2502.15223*, Feb. 2025. <https://doi.org/10.48550/arXiv.2502.15223>
- [28] S. Patil, R. R. Bijapur, A. P. Bidargaddi, S. D. Kothari, and S. H. Dabade, "Enhancing movie recommendation systems with BERT: A deep learning approach," in *Proc. INCET*, Belgaum, India, May 2024, pp. 1–8. <https://doi.org/10.1109/INCET61516.2024.10593284>
- [29] A. Vats, V. Jain, R. Raja, and A. Chadha, "Exploring the impact of large language models on recommender systems: An extensive review," *arXiv preprint arXiv:2402.18590*, Feb. 2024. <https://doi.org/10.48550/arXiv.2402.18590>
- [30] L. Wu *et al.*, "A survey on large language models for recommendation," *World Wide Web*, vol. 27, no. 5, 2024, Art. no. 60. <https://doi.org/10.1007/s11280-024-01291-2>
- [31] Y. Zhu, L. Wu, Q. Guo, L. Hong, and J. Li, "Collaborative large language model for recommender systems," in *Proc. ACM Web Conf.*, May 2024, pp. 3162–3172. <https://doi.org/10.1145/3589334.3645347>
- [32] J. Lin *et al.*, "How can recommender systems benefit from large language models: A survey," *ACM Trans. Inf. Syst.*, vol. 43, no. 2, pp. 1–47, 2025. <https://doi.org/10.1145/3678004>
- [33] Y. Zhang and L. Zhang, "Movie recommendation algorithm based on sentiment analysis and LDA," *Procedia Comput. Sci.*, vol. 199, pp. 871–878, 2022. <https://doi.org/10.1016/j.procs.2022.01.109>
- [34] Q. Zhang, J. Li, Q. Jia, C. Wang, J. Zhu, Z. Wang, and X. He, "UNBERT: User-news matching BERT for news recommendation," in *Proc. IJCAI*, 2021, pp. 3356–3362. <https://doi.org/10.24963/ijcai.2021/462>
- [35] S. R. Javaji and K. Sarode, "Multi-BERT for embeddings for recommendation system," *arXiv preprint arXiv:2308.13050*, Aug. 2023. <https://doi.org/10.48550/arXiv.2308.13050>
- [36] E. Jeong, X. Li, A. Kwon, S. Park, Q. Li, and J. Kim, "A multimodal recommender system using deep learning techniques combining review texts and images," *Appl. Sci.*, vol. 14, no. 20, Oct. 2024, Art. no. 9206. <https://doi.org/10.3390/app14209206>
- [37] J. L. Herlocker, J. A. Konstan, and J. Riedl, "Explaining collaborative filtering recommendations," in *Proc. ACM Conf. Comput. Supported Cooperative Work*, Dec. 2000, pp. 241–250. <https://doi.org/10.1145/358916.358995>
- [38] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
- [39] J. Vig, "BERTViz: A tool for visualizing multihead self-attention in the BERT model," in *ICLR Workshop: Debugging Machine Learning Models*, 2019.
- [40] P. Liu, L. Zhang, and J. A. Gulla, "Dynamic attention-based explainable recommendation with textual and visual fusion," *Inf. Process. Manag.*, vol. 57, no. 6, Nov. 2020, Art. no. 102099. <https://doi.org/10.1016/j.ipm.2019.102099>
- [41] Y. Gao, T. Sheng, Y. Xiang, Y. Xiong, H. Wang, and J. Zhang, "Chat-REC: Towards interactive and explainable LLMs-augmented recommender system," *arXiv preprint arXiv:2303.14524*, Mar. 2023. <https://doi.org/10.48550/arXiv.2303.14524>
- [42] L. Li, Y. Zhang, and L. Chen, "Personalized transformer for explainable recommendation," *arXiv preprint arXiv:2105.11601*, May 2021. <https://doi.org/10.48550/arXiv.2105.11601>
- [43] Y. Hou, J. Li, Z. He, A. Yan, X. Chen, and J. McAuley, "Bridging language and items for retrieval and recommendation," *arXiv preprint arXiv:2403.03952*, Mar. 2024. <https://doi.org/10.48550/arXiv.2403.03952>
- [44] Yelp, "Yelp Open Dataset." [Online]. Available: <https://business.yelp.com/data/resources/open-dataset/>
- [45] A. Pal, A. Barigidad, and A. Mustafi, "Identifying movie genre compositions using neural networks and introducing GENREC – A recommender system based on audience genre perception," in *Proc. ICCCS*, Patna, India, Oct. 2020, pp. 1–7. <https://doi.org/10.1109/ICCCS49678.2020.9276893>
- [46] D. A. Adeniyi, Z. Wei, and Y. Yongquan, "Automated web usage data mining and recommendation system using k-nearest neighbor (KNN) classification method," *Appl. Comput. Inform.*, vol. 12, no. 1, pp. 90–108, Jan. 2016. <https://doi.org/10.1016/j.aci.2014.10.001>
- [47] P. Velentzas, M. Vassilakopoulos, A. Corral, and C. Antonopoulos, "GPU-based algorithms for processing the k nearest-neighbor query on spatial data using partitioning and concurrent kernel execution," *Int. J. Parallel Program.*, vol. 51, no. 6, pp. 275–308, July 2023. <https://doi.org/10.1007/s10766-023-00755-8>
- [48] F. Rahutomo, T. Kitasuka, and M. Aritsugi, "Semantic cosine similarity," in *Proc. Int. Student Conf. Adv. Sci. Technol. (ICAST)*, 2012.
- [49] Fine-tuned BERT sentiment model (implementation). [Online]. Available: <https://huggingface.co/sadhakl/bert-base-uncased-finetuned-sst2-v2>
- [50] L. Li, T.-T. Goh, and D. Jin, "How textual quality of online reviews affect classification performance: A case of deep learning sentiment analysis," *Neural Comput. Appl.*, vol. 32, pp. 4387–4415, 2020. <https://doi.org/10.1007/s00521-018-3865-7>

Hasan Serhat Gunduz received his Master's degree in Business Information Systems from Bogazici University in 2025. He has been working as a Data Scientist at private companies in Türkiye since 2020. His research interests include machine learning, explainable artificial intelligence, and recommendation systems. He currently works as an Advanced Analytics Manager at a pharmaceutical company.

E-mail: hasanserhatgunduz@gmail.com

ORCID iD: <https://orcid.org/0009-0000-7786-7719>

Aysun Bozanta is an Assistant Professor of Management Information Systems at Boğaziçi University, specializing in artificial intelligence, machine learning, and data science. Her research focuses on large language models, explainability, recommender systems, and AI applications in education and business. She has led interdisciplinary projects on fraud detection, digital technologies, and sustainable value chains, collaborating with international partners across sectors including finance, manufacturing, and retail. Her research has been published in reputable international journals. In addition to her research, she teaches courses on systems analysis and design, information technologies, and recommender systems, and actively mentors students in technology and innovation. She also supports women's participation in technology through the MIS Sisterhood initiative.

E-mail: aysun.bozanta@bogazici.edu.tr

ORCID iD: <https://orcid.org/0000-0002-1768-6278>

Osman Yucel received his PhD degree in Computer Science from the University of Tulsa in 2017, with a specialization in recommender systems. He has held applied research and engineering roles at leading technology and e-commerce companies, including Groupon, Amazon, and Meta, and subsequently moved into the financial services industry. His experience spans large-scale personalization, recommendation engines, and intelligent systems across domains including e-commerce, social media, and financial services. His research interests include recommender systems and agent-based and multi-agent systems.

E-mail: mail@osmanyucel.com

ORCID iD: <https://orcid.org/0009-0006-4754-8307>