Expert Systems With Applications 261 (2025) 125533

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Integrated sentiment analysis with BERT for enhanced hybrid recommendation systems

Nossayba Darraz [a,*], Ikram Karabila [a], Anas El-Ansari [b], Nabil Alami [c], Mostafa El Mallahi [a]

[a] IPI Laboratory, Department of Mathematics and Computer Sciences, High Normal School, Sidi Mohamed Ben Abdellah University, Fez, Morocco
[b] MASI Laboratory, Computer Science Department, Polydisciplinary Faculty, Mohammed 1st University, Nador, Morocco
[c] MASI Laboratory, Mohammed First University, Higher School of Technology, Nador, Morocco

ARTICLE INFO

Dataset link: https://data.world/brianray/yelp-reviews/workspace/file?filename=yelp_training_set_review.csv

Keywords:
Recommendation system
Sentiment analysis integration
BERT model
Collaborative filtering
Content-based clustering
NMF and decisionTreeRegressor hybrid model

ABSTRACT

Recommendation systems play a crucial role in assisting users by providing personalized suggestions based on their preferences. However, these systems often face challenges such as data sparsity, cold start user issues, and overlooking the critical aspect of user sentiment. To address these limitations, this study integrates sentiment analysis into recommendation systems using a dataset from Yelp, focusing on two domains: restaurants and hotels. We conduct preprocessing tasks to facilitate sentiment analysis and leverage the power of BERT, achieving an accuracy rate of 90% for the restaurant domain and 87% for the hotel domain. We then incorporate the sentiment analysis component into collaborative filtering, utilizing model-based deep matrix factorization, and significantly reduce the root mean square error (RMSE) to 0.1 for the restaurant domain and 0.1186 for the hotel domain. Additionally, we integrate sentiment analysis into content-based recommendation using clustering, resulting in improved recommendations with a higher silhouette score of 0.533 for the restaurant domain and 0.42 for the hotel domain. To further enhance system performance, we propose a novel approach that combines these sentiment-aware components using NMF with DecisionTreeRegressor, achieving an even lower RMSE of 0.02 for the restaurant domain and 0.01 for the hotel domain. This integration of sentiment analysis into the recommendation system demonstrates its effectiveness in improving accuracy and personalization, providing users with more meaningful and relevant recommendations based on their sentiments. However, challenges such as data sparsity, cold start problems for new users, and other limitations remain, warranting further research to mitigate these challenges for a more robust recommendation system.

1. Introduction

The internet is a global network of interconnected computers that communicate using standardized protocols, allowing people to access and share information worldwide. Information can be in the form of text, images, audio, or video, and reviews are a specific type of information where people share their opinions about products, services, or experiences. These reviews can be positive, negative, or neutral and provide valuable insights into the quality of a product or service (Al-Ghuribi & Noah, 2019; Patel, Desai, & Panchal, 2017).

A recommendation system, also known as a recommender system, is a type of information filtering system that provides suggestions for items most relevant to a particular user. The primary goal of a recommendation system is to offer users meaningful recommendations for items or products that closely match their interests. Such systems are especially valuable when users face an overwhelming array of choices from a multitude of offerings provided by a service or platform (Lu, Wu, Mao, Wang, & Zhang, 2015; Zhang, Lu, & Jin, 2021). The approaches used in recommendation systems include collaborative filtering (CF), content-based (CB) filtering, and hybrid methods. CF relies on user–item interactions and identifies patterns by comparing a user's preferences to those of similar users. In contrast, CB filtering focuses on item attributes and recommends items that share characteristics with those the user has previously expressed interest in. Hybrid methods combine elements of both CF and CB filtering to improve recommendation accuracy (Al-Ghuribi & Noah, 2019; Wei, He, Chen, Zhou, & Tang, 2017).

Recommender systems, which originated in the mid-1990s and rely on users' ratings and preferences, have grown significantly over the past few decades. They are widely used in various online applications, including e-commerce platforms like Amazon, streaming services such

* Corresponding author.
E-mail addresses: nossayba.darraz@usmba.ac.ma (N. Darraz), ikram.karabila@usmba.ac.ma (I. Karabila), anas.elansari@gmail.com (A. El-Ansari), n.alami@ump.ac.ma (N. Alami), mostafa.elmallahi@usmba.ac.ma (M. El Mallahi).

https://doi.org/10.1016/j.eswa.2024.125533
Received 15 December 2023; Received in revised form 29 September 2024; Accepted 6 October 2024
Available online 11 October 2024
0957-4174/

as Netflix, and social media networks like Facebook. These systems enhance user experience through personalized suggestions based on individual preferences and behaviors. Beyond these popular domains, recommendation systems are also crucial in the restaurant and hotel industries, where platforms like Yelp and TripAdvisor use these algorithms to help users discover the best dining and lodging options (Lu et al., 2015). In the context of restaurant and hotel reviews and ratings, recommendation systems assist users in finding establishments that align with their preferences and expectations. They play a pivotal role in increasing user engagement, driving bookings, and aiding users in discovering new and relevant options, ultimately contributing to a more satisfying experience.

Recommendation systems employ a variety of technologies to provide personalized suggestions to users. Some of the main technologies used in recommendation systems include:

- **Text Mining**: Text mining techniques are used to extract and analyze textual information from various sources, such as product descriptions, reviews, and user-generated content. This helps in understanding user preferences and item characteristics (Raeesi Vanani, Mahmoudi, Jalali, & Pho, 2022).
- **K-Nearest Neighbors (KNN)**: KNN uses historical interactions to identify similar items or users, making it a collaborative filtering technique (Vinay, Kumaraswamy, & Basavaraddi, 2021).
- **Clustering**: Clustering algorithms group similar items or users together. This can be useful in content-based recommendation systems, where items are clustered based on their features, and users are recommended items from the same cluster (Isinkaye, Folajimi, & Ojokoh, 2015; Ko, Lee, Park, & Choi, 2022).
- **Matrix Factorization**: A method that decomposes a user–item matrix into two lower-dimensional matrices, which can then be used to recommend items based on the similarity of their corresponding columns or rows (Liu & Zhao, 2023; Mehta & Rana, 2017).
- **Neural Networks**: Neural networks, especially deep learning models, are increasingly utilized in recommendation systems to capture intricate patterns in user behavior and item characteristics (Ko et al., 2022).

The aim of this study is to address the gap in existing research regarding the integration of sentiment analysis into recommender systems. Specifically, we seek to critically assess the limitations of current approaches, including challenges such as cold start user issues and data sparsity, and propose a novel methodology that effectively leverages sentiment analysis to enhance recommendation accuracy and relevance. By conducting a comprehensive evaluation of existing techniques and proposing innovative solutions, we aim to contribute to the advancement of sentiment-aware recommender systems.

Sentiment analysis (SA), a domain in natural language processing, predicts the positive or negative polarity of an entity, providing insights into user attitudes, opinions, and emotions. It extracts valuable information from reviews, including overall sentiment towards a product or service, specific liked or disliked features, and sentiment trends for different aspects of an item (Darraz et al., 2023; Darraz, Karabila, El-Ansari, et al., 2024; Zhang, Wang, & Liu, 2018).

1. ***What role does Sentiment Analysis play in recommending restaurants and hotels?***
   Sentiment Analysis (SA) significantly enhances recommendation systems for both restaurants and hotels. By extracting and analyzing sentiments expressed in user reviews, sentiment analysis contributes in the following ways (Kalaivani, Kanimozhiselvi, Narmatha, & Bilal, 2023):

   - **Understanding User Preferences**: Sentiment analysis helps systems comprehend user sentiments towards specific cuisines, service quality, ambiance, or even specific dishes in both restaurants and hotels. This deep understanding enables recommendation systems to cater to individual preferences effectively.
   - **Improving Recommendations**: By considering both positive and negative sentiments, recommender systems can recommend restaurants and hotels that align with a user's emotional and experiential requirements, thereby enhancing the likelihood of user satisfaction in both domains (Darraz, Karabila, El-Ansari, Alami, 2024).
   - **Enhancing Diversity**: Sentiment analysis fosters diversity in recommendations by capturing the nuanced preferences of users for both restaurants and hotels, resulting in a more varied and personalized dining and accommodation experience.

2. ***Contributions:***
   The contributions of our work can be summarized as follows:

   - **Advanced Integration of Sentiment Analysis in Recommendation Systems:** Our approach uniquely incorporates sentiment analysis to refine recommendation systems, utilizing BERT-base-uncased with AdamW optimization. This method not only leverages cutting-edge language models but also introduces innovative enhancements tailored to improve sentiment detection in textual data, setting our work apart from existing popular tasks.
   - **Superior Sentiment Analysis Model Selection:** Through rigorous comparison of Vader, TextBlob, AFINN, BERT, and GRU, we identified BERT as the top performer for sentiment analysis, ensuring the highest accuracy and reliability in our recommendation system.
   - **Hybrid Recommendation System with Feature Combination:** We employ a hybrid recommendation approach that combines collaborative filtering, content-based filtering, and sentiment analysis using a hybrid algorithm based on feature combination. This method enhances the accuracy and coverage of our recommendations by integrating multiple strategies.
   - **Collaborative Filtering using Deep Matrix Factorization (DeepMF):** Our contribution employs Deep Matrix Factorization (DeepMF), a model-based approach leveraging neural networks to learn latent representations of users and items. We compare DeepMF with SVD and CoClustering, showcasing its superior ability to capture intricate user–item relationships for more accurate recommendations.
   - **Content-Based Recommendation with Clustering:** In our content-based recommendation system, we employ clustering techniques tailored to each domain. For the restaurant domain, we utilize KMeans clustering, while for the hotel domain, we employ Agglomerative Clustering. These techniques group items based on their textual attributes generated using Word2Vec (Skip-gram), and numerical attributes standardized using techniques like StandardScaler. To evaluate the effectiveness of our clustering algorithm, we employ the silhouette score, ensuring that our approach accurately captures item similarities and delivers personalized recommendations.
   - **Hybrid Recommendation System with NMF and DecisionTreeRegressor:** Our hybrid recommendation system combines collaborative filtering, content-based filtering, and sentiment analysis using NMF and DecisionTreeRegressor. NMF factors the matrix, while DecisionTreeRegressor predicts the recommendation score. We assess the performance of our approach by comparing it with other hybrid techniques, where NMF and DecisionTreeRegressor demonstrate superior results with a low RMSE.

These contributions collectively enhance our recommendation system by incorporating SA, utilizing advanced models like BERT, and employing hybrid approaches that combine CF, CB filtering, and SA techniques.

3. ***Paper Organization:***
The paper is organized into five sections: introduction, literature review, proposed system, results and discussion, and conclusion. The introduction sets the context and research problem, while the literature review explores existing research in the field. The proposed system section details the methodology and techniques used. The results and discussion section presents the experimental findings and their analysis. The conclusion summarizes the key findings and discusses potential future directions. Finally, the references section provides a list of cited sources.

2. Literature review

Adding sentiment analysis (SA) to recommendation systems (RS) is essential for several reasons:

- Sentiment analysis provides a deeper understanding of user preferences by capturing the emotional tone expressed in user-generated content. This allows for more accurate and personalized recommendations (Dang, Moreno-García, & Prieta, 2021a; Patel & Chhinkaniwala, 2022).
- Sentiment analysis helps address the problem of information overload by filtering and prioritizing items based on user sentiments, reducing decision fatigue (Dang et al., 2021a; Patel & Chhinkaniwala, 2022).
- Sentiment analysis contributes to building trust and credibility by identifying and filtering out biased or fake reviews, ensuring reliable recommendations (Dang et al., 2021a).
- Integrating sentiment analysis into recommendation systems enhances their performance, user experience, and the quality of recommendations (Dang et al., 2021a; Patel & Chhinkaniwala, 2022).

Overall, sentiment analysis can help improve the accuracy and reliability of recommendation systems, leading to better customer satisfaction, decision-making, and processes.

Elahi, Kholgh, Kiarostami, Oussalah, and Saghari (2023) proposed an innovative hybrid RS that incorporates SA of user reviews to enhance the recommendation process. Advanced algorithms were employed to analyze sentiments and generate personalized recommendations for users. The study revealed that user review sentiments are not always strongly correlated with ratings, indicating that sentiment can serve as an alternative signal for user feedback, capturing different aspects of user preferences. The hybrid RS outperformed various baselines when evaluated using the Amazon Digital Music and Amazon Video Games datasets. The study employed a standard NLP pipeline for data cleaning and pre-processing, extracting sentence embeddings from review texts using the BERT and Python Transformers library. The BERT-base-uncased model was utilized to extract last layer representations, and PCA was applied to reduce the vector size to 10. The same BERT-base-uncased model was used for SA of the review texts. The recommendation system utilized item-based collaborative filtering, ALS, YoutubeRanker, and DeepFM techniques.

Avishek Garain's hotel recommendation system, as described in Ray, Garain, and Sarkar (2021), utilizes sentiment analysis and aspect-based review categorization of online hotel reviews from Tripadvisor.com. The system combines a BERT model for sentiment analysis with various textual features and a Random Forest classifier. Reviews are categorized using fuzzy logic and cosine similarity, and recommendations are made based on user queries, preferred locations, and review categories. The system achieved impressive classification accuracy and outperformed other models, forming compact review clusters using the K-RMS clustering algorithm.

Asani, Vahdat-Nejad, and Sadri (2021) introduced a context-aware recommender system designed to extract users' food preferences from their restaurant-related comments and provide tailored restaurant recommendations. Leveraging natural language processing techniques, the system identifies food names within user comments, employing a semantic similarity approach to conceptually cluster these items. The Wu-Palmer clustering method yields higher precision results. Sentiment analysis is applied to determine user sentiments towards these food items, distinguishing between positive and negative opinions. The recommender system takes into account not only user preferences but also factors like location, time, and other user feedback to recommend nearby restaurants that align with the user's tastes. Using data from TripAdvisor in 2018 for evaluation. Moreover, comparative analysis with previous research highlights the system's superior performance across various criteria, making it a valuable contribution to the field of restaurant recommendation systems.

The objective of the paper (Deac-Petruşel & Limboi, 2020) is to improve recommendation systems by integrating sentiment analysis techniques into the traditional k Nearest Neighbors collaborative filtering algorithm. The paper introduces two significant enhancements. Firstly, a sentiment rating approach is developed, which incorporates sentiment scores (sentiwordnet) for each item, replacing numerical ratings in the recommendation process. Secondly, a new sentiment-based user similarity measure called ARP is introduced, taking into account factors such as the attractiveness, relevance, and popularity of user reviews. The performance of ARP is evaluated using publicly available datasets from Yelp Restaurants Reviews and Datafiniti Hotel Reviews, comparing it to traditional similarity measures such as Pearson Correlation Coefficient, Cosine, or PIP. The results demonstrate that ARP effectively replaces traditional similarity measures in the context of collaborative filtering, illustrating its potential to enhance recommendation systems.

This research (Dang, Moreno-García, & De la Prieta, 2021b) presents a novel recommendation method that combines SA and genre-based similarity in collaborative filtering. It utilizes BERT for genre preprocessing, deep learning models for SA, and was evaluated on movie datasets, demonstrating substantial performance improvement. This user-based collaborative filtering approach aims to surpass traditional methods like SVD and NMF, enhancing recommendation reliability through SA and genre embedding.

This study (Ziani et al., 2017) introduces a multilingual RS with SA to assist Algerian users in making informed decisions based on online reviews. It combines RS and SA to provide highly accurate recommendations. To overcome data limitations, the paper utilizes semi-supervised Support Vector Machines (SVM) for sentiment polarity scoring. The system effectively analyzes Algerian reviews, detects sentiment polarity, and generates meaningful recommendations for users. It demonstrates its efficiency across various languages, employing user-based collaborative filtering with Spearman similarity as the optimal measure.

This article (Kumar, De, & Roy, 2020) introduces a hybrid movie RS that integrates collaborative filtering and content-based filtering with SA of tweets from microblogging sites. The system leverages MovieTweetings database, focusing on movies released after 2014 due to tweet availability. Sentiment analysis is performed on user tweets using VADER, enhancing the recommendation model. This hybrid approach combines content-based features and collaborative social filtering, using a weight matrix to calculate item similarities based on metadata and user preferences. The system provides robust movie recommendations by combining data from the recommendation system and sentiment analysis.

This research (Sahu, Kumar, MohdShafi, Shafi, Kim, & Ijaz, 2022) introduces a framework that combines SA and a hybrid RS to recommend unreleased movies. It begins by analyzing user comments

Table 1
Recommender systems with sentiment analysis: A comparative overview.

Paper	Collaborative filtering model	Content-based model	Sentiment analysis model	Proposed approach	Advantage	Limit
Elahi et al. (2023)	Item-based Collaborative Filtering, ALS, YoutubeRanker, DeepFM	N/A	BERT-base-uncased for last layer representations	Hybrid RS incorporating SA	Superior performance compared to different baselines	Sentiment-rating correlation limitations; PCA-based reduction may lose information.
Asani et al. (2021)	Item-to-item Collaborative Filtering with Cosine Similarity	Semantic similarity approach, Wu-Palmer clustering method	SentiWordNet-based	Hybrid RS extracting food preferences from user comments	Superior performance across various criteria, considers user preferences, location, time, and other factors	evaluation using TripAdvisor data from 2018 may not reflect current trends or user preferences
Deac-Petruşel and Limboi (2020)	k Nearest Neighbors Collaborative Filtering	N/A	SentiWordNet	Sentiment-based user similarity measure (ARP)	Enhanced effectiveness of RS	Sensitivity to inaccuracies in SA, e.g., SentiWordNet
Dang et al. (2021b)	User-based Collaborative Filtering with BERT-based Genre Preprocessing	N/A	Feature vector with BERT on reviews; LSTM+CNN models for SA	Sentiment analysis and genre-based similarity in hybrid collaborative filtering.	Substantial performance improvement surpassing SVD, NMF, and SVD++.	reliance on deep learning models may require significant computational resources
Ziani et al. (2017)	User-based Collaborative Filtering with Spearman Similarity	N/A	Semi-supervised Support Vector Machines (SVM)	Multilingual recommender with SA for Algerian users	Accurate multilingual recommendations for Algerian reviews.	Semi-supervised SVM reliance, generalization challenges.
Our proposed work	Model-based (Deep Matrix Factorization) with SA	Clustering using K-means for restaurant domain, Agglomerative Clustering for hotel domain. Utilizes Word2Vec, StandardScaler, and PCA.	BERT for SA (AdamW, LR 1e−5)	Hybrid approach integrating clustering, DeepMF and BERT with NMF and Decision-TreeRegressor for recommendation	Improved performance through advanced sentiment analysis and hybrid recommendation techniques.	Potential challenges with hyperparameter tuning; computational resources needed for BERT training.

on movie trailers from Netflix's official YouTube channel to predict movie ratings. The second module utilizes movie data from The Movie Database (TMDb) to create a hybrid recommender system for generating a list of preferred upcoming movies for individual users. Finally, the study merges sentiment-based movie ratings and user preferences to offer personalized recommendations. The research focuses on recommending unreleased movies based on sentiment analysis of user-generated social media comments, employing Vader and TextBlob sentiment analysis methods for rating predictions.

This study (Selmene & Kodia, 2020) combines sentiment analysis, specifically using TextBlob, of Twitter data with collaborative filtering to improve recommendation system accuracy, addressing sparsity and cold-start issues. It merges user ratings and sentiment analysis of tweets related to items, ultimately enhancing the recommendation process.

The Table 1 presents a comparative overview of RS integrated with SA. Each research paper proposes different models and approaches, highlighting their advantages and limitations. For example, some papers combine collaborative filtering with SA, achieving superior performance. Others consider factors like user preferences, location, and multilingual recommendations. The proposed work combines cosine similarity, clustering, and BERT for SA, offering improved performance through advanced techniques. However, challenges such as hyperparameter tuning and computational resource requirements for BERT training may need to be addressed. Overall, the table provides valuable insights into the integration of sentiment analysis in recommender systems, showcasing various approaches and their respective strengths and limitations.

3. Proposed system

Our proposed recommender system integrates collaborative filtering, content-based filtering, and sentiment analysis techniques to create a hybrid recommendation system. To achieve this, clustering was utilized for content-based filtering, while a model-based was employed for collaborative filtering. The primary objective is to enhance the reliability of user recommendations by incorporating sentiment analysis of user reviews or comments alongside traditional recommendation methods. Fig. 1 illustrates the architecture of our system.

3.1. Data set description

The Yelp dataset https://data.world/brianray/yelp-reviews/workspace/file?filename=yelp_training_set_review.csv, published in 2017, encompasses a wide range of characteristics that provide valuable insights into business reviews and user behavior. The dataset includes information about various domains such as restaurants, hotels, education, and fashion. For the purpose of this analysis, we will focus on the restaurant and hotel domains. The dataset includes details such as business name, location (city, latitude, longitude, full address, and state), business type, whether they are open, review count, star rating, and categories. It

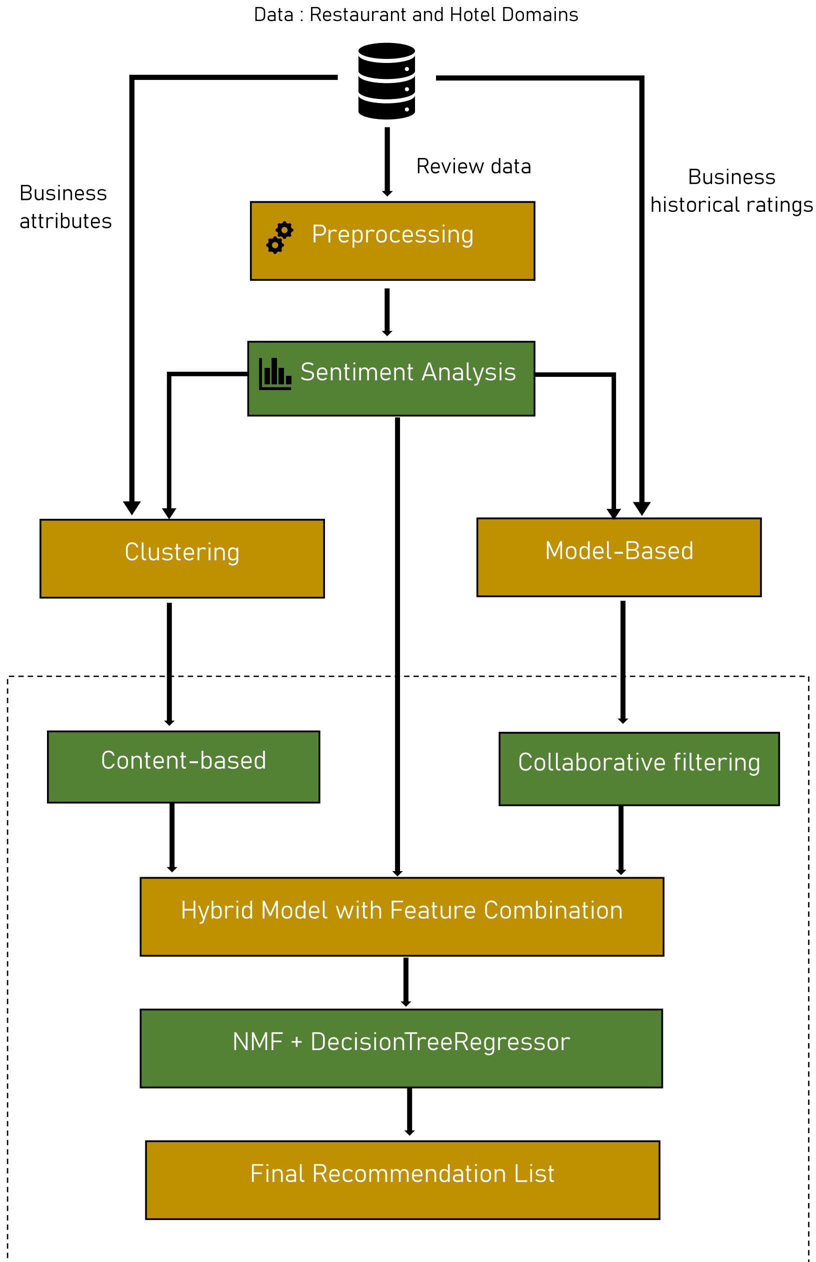

Fig. 1. A hybrid recommendation framework enhanced with sentiment analysis.

also contains reviewer-related attributes, such as reviewer name, type, average star rating, and review count. Additionally, the dataset features review-specific details like the review ID, date, text, and star rating, as well as metrics related to the usefulness, coolness, and funniness of the reviews. This rich collection of attributes enables comprehensive analysis of restaurant and hotel reviews, user preferences, sentiment analysis, and the development of recommendation systems, offering valuable insights for various research and analytical purposes. Table 2 showcases extracted features that provide valuable insights into its contents (see Figs. 2–6).

3.2. Pre-processing of review text data

In sentiment analysis, preprocessing tasks are essential for preparing text data by converting raw, unstructured text data into a structured and clean format that can be readily fed into sentiment classification models (Duong & Nguyen-Thi, 2021; Haddi, Liu, & Shi, 2013). These tasks include Fig. 2:

1. Stop Words Removal: Commonly used words such as “the”, “is”, and “and” were removed from the text as they typically do not carry significant meaning.
2. Tokenization: The text was divided into individual tokens or words to facilitate further analysis. This step helps in breaking down the text into meaningful units for processing.
3. Part-of-Speech Tagging: Each token was assigned a part of speech tag, such as noun, verb, adjective, or adverb. This information enables a better understanding of the grammatical structure and context of the text.
4. WordNet Part-of-Speech Information: is incorporated to align the POS tagging with WordNet’s lexical database, enhancing the system’s understanding of word relationships and meanings.
5. Lemmatization: is the process of reducing words to their base or root form. It helps in standardizing words and reducing the unique number of words used in the data

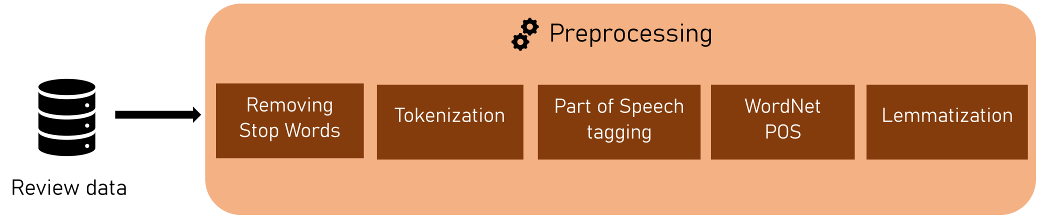

Fig. 2. Preprocessing operations.

Table 2
Example of yelp dataset features extracted.

Features	Value
business_blank	False
business_categories	Greek; Restaurants
business_city	Scottsdale
business_full_address	1336 N Scottsdale Rd Scottsdale, AZ 85257
business_id	–5jkZ3-nUPZxUvtcbr8Uw
business_latitude	33.463373
business_longitude	−111.926908
business_name	George's Gyros Greek Grill
business_open	True
business_review_count	11
business_stars	4.5
business_state	AZ
business_type	business
cool	0
date	2012–05–22
funny	0
review_id	rAS26HNMyqj_d6PBf4vuHw
reviewer_average_stars	3.83
reviewer_blank	False
reviewer_cool	1
reviewer_funny	1
reviewer_name	Natalie
reviewer_review_count	6
reviewer_type	user
reviewer_useful	9
stars	4
text	Good gyros, clean and friendly staff.
type	review
useful	1
user_id	S_NH8JMiL0v9mrNBcClp0Q

3.3. Sentiment analysis

The BERT-base-uncased model is a pretrained transformer-based language model specifically fine-tuned for sentiment analysis (SA) tasks. It has undergone training on a substantial amount of English data using a masked language modeling (MLM) objective. Being uncased, it does not consider the case sensitivity of characters in the input text. This versatile model has been applied to various natural language processing tasks, including sentiment analysis, where it can be fine-tuned to accurately identify emotions and sentiments within textual data (Geetha & Renuka, 2021; Xie, Cao, Wu, Liu, Tao, & Xie, 2020). In BERT-base-uncased, the "uncased" attribute signifies that the model is trained on lowercased text. This means that all input text is converted to lowercase during both pre-training and fine-tuning stages. Lowercasing helps in reducing the vocabulary size and improves generalization by treating the same word in different cases as identical (Durairaj & Chinnalagu, 2021).

The architecture of BERT-base-uncased Fig. 3 consists of multiple layers of transformer encoders. The transformer encoder is a fundamental building block that allows BERT to capture contextual information from both the left and right contexts of a given word. This bidirectional approach enables BERT to understand the meaning of a word in the context of the entire sentence, leading to better representation learning (Kula, Choraś, & Kozik, 2021; Sousa, Sakiyama, de Souza Rodrigues, Moraes, Fernandes, & Matsubara, 2019).

For Sentiment Analysis (SA) using BERT-base-uncased, a common approach is to fine-tune the pre-trained model on a specific sentiment classification task. During the fine-tuning process, the AdamW optimizer is often employed to update the BERT model's parameters. The AdamW optimizer combines the benefits of the Adam optimizer and weight decay regularization to prevent overfitting.

To prepare the training data for SA, the text is first tokenized using the BERT tokenizer. The tokenizer splits the input text into a sequence of subwords or tokens, taking into account word boundaries and character-level information. The resulting tokenized text is then fed into the BERT model, which is configured for sequence classification. In sentiment analysis, the BERT model is typically trained to classify text into two sentiment classes, namely 'Negative' and 'Positive'.

In the implementation at hand, sentiment analysis is conducted by fine-tuning the BERT-base-uncased model parameters. The AdamW optimizer, with a learning rate of 1e−5, is commonly utilized to update the model weights during training. The learning rate determines the step size at which the optimizer adjusts the model parameters based on the calculated gradients. By carefully selecting the learning rate and employing the AdamW optimizer, the training process can be optimized to achieve better performance in sentiment analysis tasks Fig. 4.

3.4. Hybrid recommendation system

Our proposed hybrid recommendation system combines collaborative filtering and content-based filtering techniques. Collaborative filtering considers user preferences and integrates the output of sentiment analysis derived from user ratings and reviews. This sentiment analysis helps to gauge the users' feelings and opinions about various businesses, providing a deeper understanding of their preferences. Content-based filtering, on the other hand, clusters businesses based on attributes such as category, name, address, city, and review count. By leveraging both methods, our system provides personalized recommendations tailored to user preferences, enhancing the ability to discover new businesses that align with their tastes. The integration of sentiment analysis ensures that the recommendations are not only based on numerical ratings but also on the emotional context of user feedback.

3.4.1. Collaborative filtering

We employ deep matrix factorization (DMF) as our chosen method for collaborative filtering tasks. Unlike conventional approaches, DMF integrates deep learning techniques to delve into the intricate relationships between users and items. Utilizing a deep neural network architecture, DMF directly learns from input matrices, facilitating the capture of nonlinear patterns present in collaborative filtering data. This departure from traditional matrix factorization methods enables DMF to offer more nuanced insights into user–item interactions, ultimately enhancing the accuracy of recommendation systems (Zhou, Wen, Li, & Zhou, 2019).

Moreover, DMF stands out for its robustness, trained across 8 folds to ensure adaptability and effectiveness across diverse datasets. This multi-fold training approach strengthens DMF's capability to generalize and accurately model various user–item interactions. By leveraging deep learning principles in collaborative filtering, DMF represents a

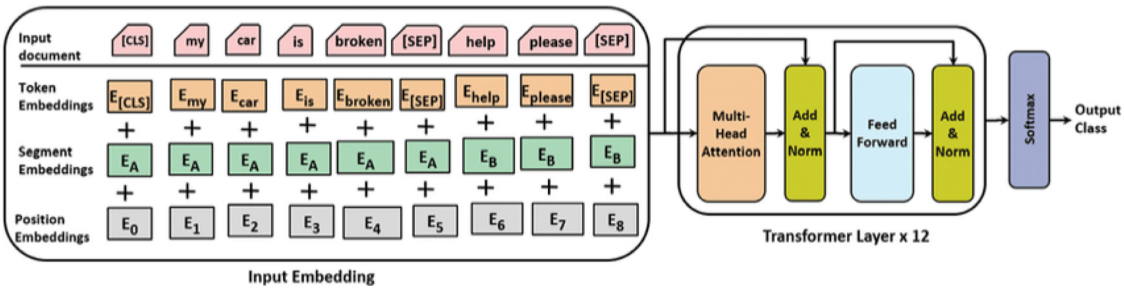

Fig. 3. BERT-base-uncased architecture.

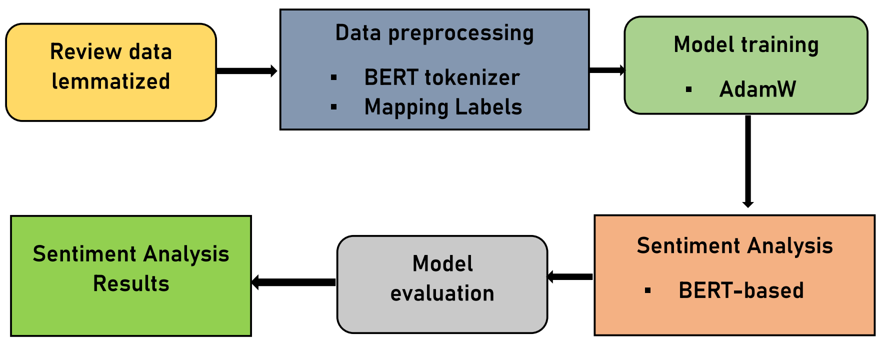

Fig. 4. BERT-based sentiment analysis pipeline.

potent tool in recommendation systems, capable of unraveling complex relationships between users and items for improved accuracy and performance.

3.4.2. Content-based

In our content-based recommendation system, we employ clustering techniques to group items based on their textual attributes and normalize numerical attributes, including sentiment scores predicted from the text. The process involves utilizing Word2Vec (Skip-gram) to generate vector representations for textual attributes, while numerical attributes are standardized using techniques like StandardScaler. The combined attribute vectors are then subjected to dimensionality reduction through PCA. Finally, the clustering algorithm is applied to group similar items together. By incorporating clustering, our system can offer personalized recommendations by considering both textual and numerical attributes, enabling a more comprehensive understanding of item characteristics and user preferences Fig. 5.

By leveraging clustering in our content-based recommendation system, we can provide users with more relevant and personalized recommendations. The grouping of items based on their attributes allows for the identification of similarities and patterns, enhancing the system's ability to understand user preferences and make accurate recommendations. Incorporating both textual and numerical attributes further enriches the recommendation process, providing a holistic view of item characteristics. Overall, the integration of clustering techniques enhances the effectiveness and personalization of our content-based recommendation system, improving the overall user experience.

3.4.3. Hybrid recommender system: NMF and DecisionTreeRegressor fusion

Some common hybrid algorithms include:

- **Weighted Hybrid**: This type of hybrid recommendation system combines the scores or predictions from multiple recommendation methods using a weighted average. The score for each recommended item is calculated as the weighted sum of the individual recommendation scores from the different sources. The weights assigned to each data source or recommendation method are user-configurable, typically through interactive sliders or controls. This allows the end-user to adjust the relative importance of the different recommendation signals based on their preferences or the context of the recommendation task. Automatically optimizing the set of weights for each data source is a desirable but nontrivial challenge, as it requires balancing the contributions from the various recommendation components to achieve the best overall performance (Bhatt, Patel, & Gaudani, 2014).
- **Switching Hybrid:**The Switching Hybrid approach selects a single recommender from the available options for each user or situation. This allows leveraging the strengths of different techniques, like switching to a collaborative method if content-based lacks confidence. The choice of recommender can vary, but this does not fully solve issues like the ramp-up problem. Crucially, the system requires a reliable criterion to determine which recommender to use, which can be difficult to define effectively (Fayyaz, Ebrahimian, Nawara, Ibrahim, & Kashef, 2020).
- **Mixed Hybrid**: In this method, recommendations from each source are ranked, and the top items are picked by alternating between sources. This approach focuses on the position in the ranked list rather than individual scores. If a recommendation appears in multiple sources, the next item from the ranked list is chosen (Bhatt et al., 2014).
- **Feature Combination:** This method merges complementary features from different recommendation techniques. For example, collaborative-based features can be integrated into a content-based recommendation algorithm. Collaborative information is treated as additional feature data and combined with content-based features. This allows the system to use collaborative data without depending entirely on it, reducing sensitivity to the number of users who have rated an item (Fayyaz et al., 2020).
- **Cascade Hybrid:** The Cascade hybrid uses a strict hierarchical process, where a series of recommendation techniques are applied in order of priority. This ensures a weaker, lower priority method cannot override the decisions of a stronger, higher priority recommender. Instead, the lower priority approach is used to refine

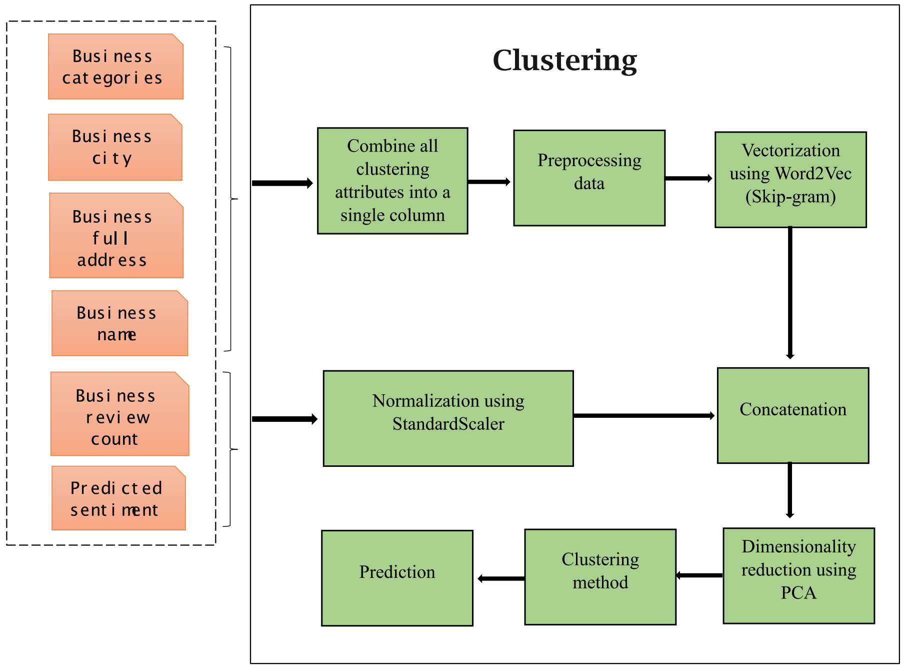

Fig. 5. Clustering-based item analysis.

and break ties in the scoring by the primary algorithm, but is not applied to items that are already well-differentiated or poorly rated by the first technique. This cascading structure makes the hybrid resilient to noise in the lower priority outputs, as it can only enhance, not reverse, the recommendations of the stronger method (Fayyaz et al., 2020).

- **Meta-Level Hybrid:** The Meta-Level hybrid uses an output model from one recommender as the input to another, allowing the final algorithm to operate on a compressed representation of user preferences rather than raw data. Transitioning a pair of recommenders into an effective Meta-Level hybrid can be challenging, as the contributing model must generate a usable high-level representation to serve as meaningful input, but this approach can enable techniques like collaborative filtering to function more effectively, especially with sparse data (Fayyaz et al., 2020).

Our hybrid recommendation system combines the output of sentiment analysis, content-based filtering, and collaborative filtering to provide users with personalized and accurate recommendations.

The core features used in this system include:

1. **Deep matrix factorization predictions (DMF_prediction):** This method employs deep matrix factorization to compute the similarity between items or users, taking into account their feature profiles. By utilizing the learned representations from the DMF model, it provides a more sophisticated measure of similarity, capturing complex relationships and patterns in the data for improved recommendation accuracy.
2. **Business cluster information (cluster):** By grouping similar businesses into clusters, the system can leverage the collective preferences and characteristics of these groups to make better recommendations.
3. **BERT-based sentiment predictions (BERT_Predictions):** Sentiment analysis using the BERT language model provides valuable insights into how users feel about the recommended items, which can be used to refine the recommendations.

To effectively integrate diverse data sources, we employ a hybrid "Feature Combination" algorithm that merges different feature types, aiming to harness the strengths of each while ensuring comprehensive data utilization. This approach utilizes non-negative matrix factorization (NMF) to condense feature space dimensionality while preserving critical information, facilitating a more compact representation. The NMF-transformed features are then utilized to train a Decision Tree Regressor, chosen for its ability to capture nonlinear relationships within complex data. By leveraging the enhanced feature representation provided by NMF, the Decision Tree Regressor can effectively model underlying patterns, resulting in more accurate predictions and insights.

The Decision Tree Regressor is well-suited for this task as it can capture complex non-linear relationships between the features and the target variable (in this case, the user ratings or preferences). By leveraging the tree-based structure of the Decision Tree, the model is able to make robust and accurate predictions, even in the presence of noisy or incomplete data.

The performance of this hybrid recommendation system is evaluated using the root mean squared error (RMSE) metric. RMSE provides a measure of the average deviation between the predicted ratings and the actual ratings, with lower values indicating better predictive accuracy. The goal is to minimize the RMSE, ensuring that the recommended items closely match the user's preferences.

This combination of sentiment analysis, content-based filtering, collaborative filtering, and advanced machine learning techniques, along with the "Feature Combination" algorithm and the Decision Tree Regressor, demonstrates the power of integrating multiple data sources and analytical methods to create a sophisticated recommendation system. By capitalizing on the complementary strengths of these approaches, the system is able to deliver superior performance and provide users with highly personalized and relevant recommendations.

ALGORITHM: Hybrid Recommender System using NMF and DecisionTreeRegressor

Step-by-Step Explanation

1. **Separate Features and Labels** : The algorithm first divides the dataset into two parts: features (inputs) and labels (outputs). The features consist of predictions from DeepMF, cluster assignments from K-means, and sentiment scores from BERT. The labels (stars) represent the actual ratings given by users and will be used as ground truth for training the model.

Algorithm 1 Hybrid Recommender System using NMF and DecisionTreeRegressor.

Require: DataFrame with features (DeepMF_predictions, cluster, BERT_Predictions) and labels (stars)
Ensure: Predicted ratings for each user
Separate features and labels:
    Extract features (DeepMF_predictions, cluster, BERT_Predictions) from the DataFrame
    Extract target labels (stars)
Apply Non-Negative Matrix Factorization (NMF) to the features:
    Initialize NMF with a specific number of components (e.g., 3)
    Fit and transform the features to generate latent features
Combine original features with NMF features:
    Concatenate the original features and NMF-generated latent features
Train a DecisionTreeRegressor on the combined features:
    Initialize DecisionTreeRegressor
    Train the regressor using the hybrid features and the target labels
Make predictions on the entire dataset:
    Predict user ratings using the trained DecisionTreeRegressor
Evaluate model using RMSE:
    Calculate RMSE to compare predicted ratings with actual ratings
return Final predicted ratings for each user

2. **Apply NMF to the Features** : Non-Negative Matrix Factorization (NMF) is applied to the feature set to reduce its dimensionality. NMF transforms the feature space into a smaller set of latent factors, capturing underlying patterns. The number of components for NMF is specified (e.g., 3), which controls how many latent features are extracted.
3. **Combine Original Features with NMF Features** : After generating the latent features using NMF, these are combined with the original feature set. The goal is to create a more comprehensive feature representation for the next step, leveraging both the original features and their lower-dimensional latent counterparts.
4. **Train a DecisionTreeRegressor** : A DecisionTreeRegressor is used to model the relationship between the hybrid features (original + NMF features) and the target labels (stars). The regressor is trained to learn how the hybrid feature set maps to user ratings.
5. **Make Predictions** : After training, the DecisionTreeRegressor makes predictions on the same dataset. These predictions represent the estimated ratings the system would recommend to each user.
6. **Evaluate the Model with RMSE** : Root Mean Square Error (RMSE) is calculated to assess how close the predicted ratings are to the actual ratings. This metric provides a measure of the prediction accuracy, with lower values indicating better performance.
7. **Return Final Recommendations** : The predicted ratings are returned as the final output, representing personalized recommendations for each user.

In summary, the proposed hybrid recommendation system utilizes feature combination, a method specifically designed for hybrid recommendation systems, to integrate diverse information sources. The system begins by applying Non-Negative Matrix Factorization (NMF) to reduce the dimensionality of features, which include DeepMF predictions, K-means cluster assignments, and BERT sentiment scores. These NMF-generated latent features are then combined with the original features to create a comprehensive representation. This feature combination method enhances the model's ability to capture both original and latent patterns in the data. A DecisionTreeRegressor is then trained on these hybrid features to predict user ratings, providing personalized recommendations. The effectiveness of this approach is measured using RMSE, ensuring that the model delivers accurate predictions by leveraging both raw and transformed data Fig. 6.

3.4.4. Metrics

The metrics employed to assess the accuracy and reliability of rating predictions include:

1. RMSE short for Root Mean Square Error, quantifies the average discrepancy between predicted and actual values, serving as a metric to evaluate a model's performance in regression analysis. It is calculated by taking the square root of the mean of squared prediction errors. A lower RMSE suggests higher accuracy and a better fit between predicted and actual values (Katarya & Verma, 2017).
$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (1)$$
In this equation: n is the total number of observations, y_i represents the actual or observed values, $\hat{y}_i$ represents the predicted values.
2. MAE or Mean Absolute Error, measures the average absolute difference between predicted and actual values, providing an assessment of a model's accuracy, with lower values indicating better performance and a closer match between predictions and observed data (Nawara & Kashef, 2020).
$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (2)$$
In this equation: n is the total number of observations, y_i represents the actual or observed values, $\hat{y}_i$ represents the predicted values.
3. NDCG short for Normalized Discounted Cumulative Gain, is a ranking evaluation metric that assesses the quality of ranked lists or recommendations. It takes into account both the relevance and the position of items in the list. NDCG calculates the cumulative gain of items, discounting the relevance based on their position, and normalizes the result to provide a value between 0 and 1. Higher NDCG scores indicate better ranking quality and more relevant recommendations (Giabelli, Malandri, Mercorio, Mezzanzanica, & Seveso, 2021).
$$\mathrm{NDCG} = \frac{DCG}{IDCG} \quad (3)$$
DCG (Discounted Cumulative Gain) is calculated as $\sum_{i=1}^{n} \frac{2^{rel_i}-1}{\log_2(i+1)}$, where rel_i is the relevance of the result at position i. IDCG (Ideal Discounted Cumulative Gain) represents the ideal scenario where the results are perfectly ranked by relevance.

4. Results and discussion

4.1. Sentiment analysis results

Sentiment analysis is a crucial task in understanding customer opinions and feedback, especially in the service industry. In this analysis, we explore the performance of various sentiment analysis models across two important domains: restaurants and hotels. By examining the accuracy of these models, we can gain insights into the unique challenges and nuances associated with sentiment analysis in these specific contexts.

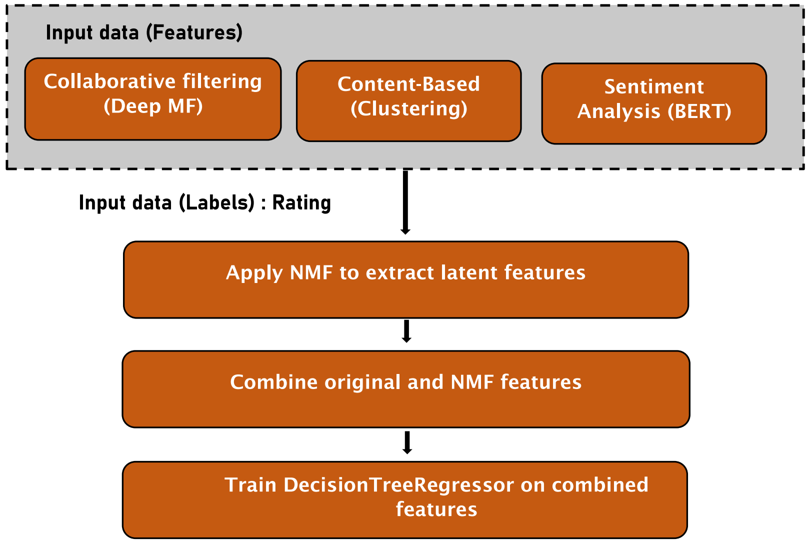

Fig. 6. Hybrid recommendation system with feature combination using NMF and DecisionTreeRegressor.

Table 3
Sentiment analysis results for the Restaurant Domain.

Sentiment models	Accuracy
AFINN	86%
Vader	85%
TextBlob	87%
BERT	**90%**
GRU	84%

Table 4
Sentiment analysis results for the Hotel Domain.

Sentiment models	Accuracy
AFINN	82%
Vader	82%
TextBlob	81%
BERT	**87%**
GRU	79%

4.1.1. Domain: Restaurant

The sentiment analysis results for the restaurant domain show that the BERT model outperforms the other models with an impressive accuracy of 90%. This suggests that the BERT model is particularly well-suited for capturing the nuanced and contextual nature of restaurant-related sentiment. (Table 3)

The AFINN, VADER, and TextBlob models also perform reasonably well, with accuracies ranging from 85% to 87%. These more traditional sentiment analysis approaches, which rely on lexicons and rule-based methods, are still able to provide reliable results in the restaurant domain.

In contrast, the GRU (Gated Recurrent Unit) model, a type of recurrent neural network, has the lowest accuracy at 84%. This could be due to the inherent complexity of the restaurant domain, where sentiment is influenced by a multitude of factors, such as food quality, service, ambiance, and personal preferences. The GRU model may struggle to capture all of these nuances effectively.

The superior performance of the BERT model suggests that advanced language models that can leverage contextual information are particularly well-suited for sentiment analysis in the restaurant domain. However, the strong results of the other models indicate that a combination of approaches, or an ensemble model, may provide the most accurate and comprehensive sentiment analysis for restaurant-related data.

4.1.2. Domain: Hotel

Similar to the restaurant domain, the BERT model performs the best in sentiment analysis for the hotel domain, with an accuracy of 87%. This suggests that the BERT model is well-suited for capturing the nuances of hotel-related sentiment, which can be influenced by factors such as room quality, staff service, amenities, and overall customer experience. (Table 4)

The AFINN and VADER models also perform well, with an accuracy of 82% each. The TextBlob model has the lowest accuracy at 81%, which could be due to its reliance on a more simplistic lexicon-based approach that may not be as effective in the hotel domain.

Table 5
Hyperparameter configuration for BERT sentiment model training.

Hyperparameter	Value
Test size	30%
Pretrained model	'bert-base-uncased'
Num labels	2
Batch size	8
Learning rate	1×10^{-5}
Optimizer	ADAM
Max length	128

The GRU model performs the worst in the hotel domain, with an accuracy of 79%. This may be due to the complexity of hotel-related sentiment, which can be influenced by a wide range of factors, and the GRU model's potential difficulty in capturing all of these nuances.

Overall, the sentiment analysis results for the hotel domain suggest that the BERT model is the most effective, but a combination of different models may provide the most accurate and comprehensive results.

The Table 5 presents a summary of key hyperparameters used in training a sentiment analysis model based on BERT.

In our pursuit of creating a comprehensive and powerful RS, we seamlessly integrated the BERT predictive model into our content-based (CB) and collaborative filtering (CF) approaches, culminating in a robust hybrid RS that incorporates SA. By combining CB filtering, CF, and SA, we leverage the strengths of each approach to provide personalized and sentiment-aware recommendations. This integration allows us to consider various factors, such as item characteristics, collaborative patterns, and sentiment expressions, resulting in a more holistic understanding of user preferences. By leveraging BERT for sentiment analysis within the hybrid RS, we enhance the accuracy,

Table 6
Clustering algorithms and performance metrics in the Restaurant Domain.

Clustering algorithm	Elbow method for cluster count	Silhouette score
AgglomerativeClustering	10	0.500
BIRCH	11	0.520
KMeans	**11**	**0.533**

Table 7
Clustering algorithms and performance metrics in the Hotel Domain.

Clustering algorithm	Elbow method for cluster count	Silhouette score
AgglomerativeClustering	**9**	**0.45**
BIRCH	7	0.36
KMeans	12	0.42

relevance, and user satisfaction of our recommendations, ultimately delivering an exceptional recommendation experience.

4.2. Content-based : Clustering result

4.2.1. Domain: Restaurant

The Table 6 presents a comparative analysis of three different clustering algorithms – AgglomerativeClustering, BIRCH, and KMeans – and their performance on the restaurant domain data. The key metrics reported in the table are the Elbow method for determining the optimal cluster count and the Silhouette score, which measures the quality of the clusters.

The results show that the KMeans algorithm achieves the highest Silhouette score of 0.533, indicating that it has produced the most well-defined clusters among the three algorithms. A Silhouette score closer to 1 suggests that the data points within each cluster are highly similar to one another and distinctly different from data points in other clusters. This implies that the KMeans algorithm was able to identify clear and cohesive groupings of restaurants based on their underlying characteristics.

Furthermore, the Elbow method also suggests an optimal cluster count of 11 for the KMeans algorithm. This means that the KMeans clustering was able to uncover 11 distinct segments or groups within the restaurant data.

In contrast, the AgglomerativeClustering and BIRCH algorithms did not perform as well, with lower Silhouette scores indicating less well-defined clusters. The Elbow method also suggested different optimal cluster counts for these algorithms, highlighting their relatively weaker performance compared to the KMeans approach.

The figure below (Fig. 7) depicts the clustering results obtained using the KMeans algorithm for the restaurant domain. The clusters demonstrate clear groupings of data points, indicating the algorithm's ability to effectively partition the data into meaningful segments. Each color represents a different cluster, showing the distinct separation and distribution of data points.

The provided Fig. 8 presents a series of word clouds, each representing a distinct cluster of restaurants identified through the clustering analysis. These word clouds offer a compelling visual representation of the key terms and their relative frequencies within each cluster, shedding light on the unique characteristics and specialties of the restaurant segments. The clusters reveal intriguing insights, such as a focus on sushi and bars in Cluster 0, a broader range of establishments including American, pizza, and breweries in Cluster 1, a concentration on vegetarian and traditional American cuisine in Cluster 2, a strong presence of Japanese and coffee/tea-focused restaurants in Clusters 4 and 5, and a focus on Indian and Middle Eastern cuisine in Clusters 7 and 8. These distinct word cloud patterns provide valuable insights into the diverse restaurant landscape, allowing stakeholders to identify and understand the unique characteristics and offerings of different restaurant clusters, which can inform strategic decision-making, targeted marketing, and the development of tailored products and services.

4.2.2. Domain: Hotel

In the hotel domain, the performance of different clustering algorithms was evaluated using the Elbow method and Silhouette scores. The table below (Table 7) summarizes the results, highlighting the cluster counts and the quality of the clusters formed by each algorithm.

The KMeans algorithm was identified as the most effective for the hotel domain, achieving a Silhouette score of 0.42 with 12 clusters. Although AgglomerativeClustering had a higher score of 0.45, it identified fewer clusters, suggesting that KMeans provided a more detailed segmentation of the data.

The following figure (Fig. 9) illustrates the clustering results obtained using the AgglomerativeClustering algorithm for the hotel domain. The figure showcases the hierarchical nature of this algorithm, with data points clustered into well-defined groups.

The provided figure (Fig. 10) presents a series of word clouds, each representing a distinct cluster of travel and hotel-related entities identified through the clustering analysis. These word clouds offer a compelling visual representation of the key terms and their relative frequencies, shedding light on the unique characteristics and specialties of different segments in the travel and hotel industry. The clusters reveal intriguing insights, such as a focus on hotels, event planning, and travel-related services in Clusters 0 and 1, a concentration on hotels, travel, and public transportation in Cluster 2, a strong presence of hotels and travel in Cluster 3, a mix of hotels, travel services, event planning, and related offerings in Cluster 4, a focus on hotels, travel, resorts, and event planning in Cluster 5, an emphasis on hotels, travel, historical landmarks, and government-related services in Cluster 6, a combination of hotels, travel, event planning, and associated services in Cluster 7, and a highlight of hotels, event planning, travel services, and venue/event spaces in Cluster 8. These distinct word cloud patterns provide valuable insights into the diverse travel and hotel ecosystem, allowing stakeholders to identify and understand the unique characteristics and offerings of different segments, which can inform strategic decision-making, targeted marketing, and the development of tailored products and services.

4.3. Collaborative filtering result

4.3.1. Domain: Restaurant

The Table 8 presents a comparative analysis of different collaborative filtering models applied to the restaurant domain, focusing on three performance metrics: RMSE (Root Mean Square Error), MAE (Mean Absolute Error), and NDCG (Normalized Discounted Cumulative Gain). Each model's performance is evaluated both with and without sentiment analysis (SA).

The results indicate that incorporating sentiment analysis significantly enhances the performance of all models. The DeepMF model exhibits the lowest RMSE (0.1) and MAE (0.054) with sentiment analysis, suggesting superior accuracy in predicting user ratings. Although its NDCG score (0.26) is lower with SA compared to other metrics, it remains highly competitive.

The SVD and CoClustering models also show marked improvements with sentiment analysis. SVD's RMSE and MAE improve notably from 1.09 to 0.5496 and from 0.8722 to 0.2110, respectively. Similarly, CoClustering's RMSE decreases from 1.2243 to 0.5453, and its MAE drops from 0.961 to 0.1989. Both models achieve excellent NDCG scores of approximately 0.999 with sentiment analysis, highlighting their effectiveness in ranking recommendations accurately.

Overall, the integration of sentiment analysis enhances the predictive accuracy and ranking quality of collaborative filtering models in the restaurant domain, with DeepMF emerging as the top performer in terms of RMSE and MAE.

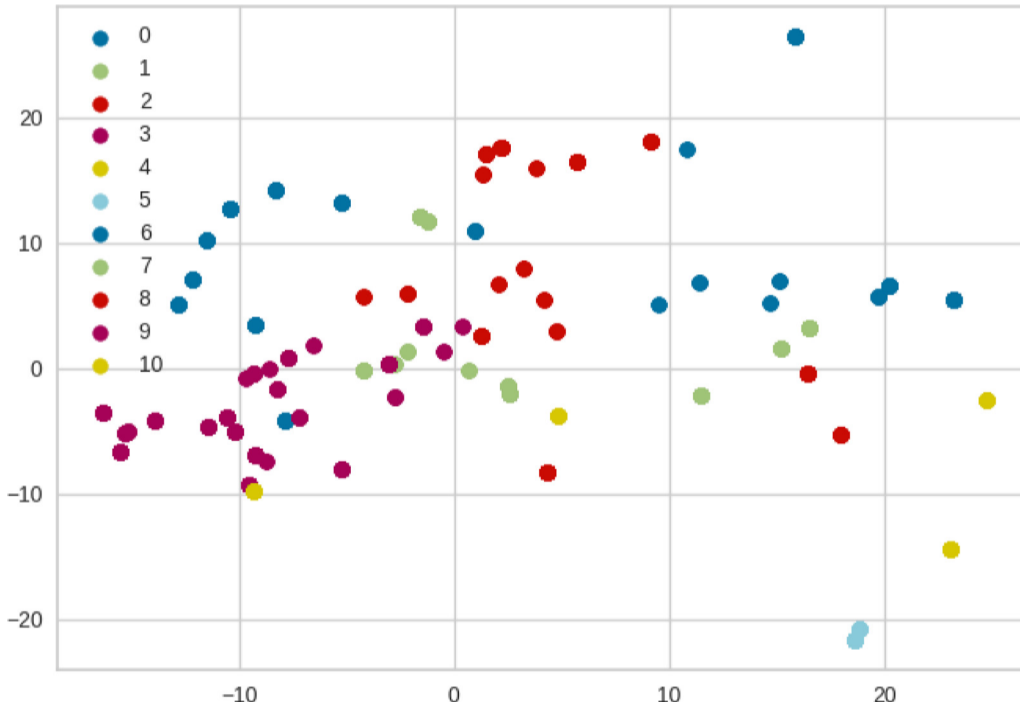

Fig. 7. Clustering results using KMeans.

Table 8
Performance of sentiment-based recommendation models in the Restaurant Domain.

Model	RMSE		MAE		NDCG	
	With SA	Without SA	With SA	Without SA	With SA	Without SA
DeepMF	**0.1000**	0.2300	**0.0540**	0.1800	0.2600	**0.9982**
SVD	0.5496	1.0900	0.2110	0.8722	0.2708	**0.9989**
CoClustering	0.5453	1.2243	0.1989	0.9610	0.2713	**0.9984**

4.3.2. Domain: Hotel

The Table 9 presents a comparative analysis of three collaborative filtering models—DeepMF, SVD, and CoClustering—in the hotel domain, evaluated with and without sentiment analysis (SA). The performance metrics used are RMSE (Root Mean Square Error), MAE (Mean Absolute Error), and NDCG (Normalized Discounted Cumulative Gain).

DeepMF shows the most significant improvement with the inclusion of sentiment analysis, achieving an RMSE of 0.1186 and an MAE of 0.07, indicating higher accuracy in predicting user preferences. However, its NDCG score is 0.24, slightly lower than when sentiment analysis is excluded, which scores 0.9978. This suggests that while DeepMF with SA excels in prediction accuracy, it performs slightly less in ranking quality.

The SVD and CoClustering models also benefit from sentiment analysis, with notable reductions in RMSE and MAE. SVD's RMSE improves from 1.1455 to 0.5504, and its MAE decreases from 0.923 to 0.2522. Similarly, CoClustering's RMSE drops from 1.3233 to 0.5478, and its MAE from 1.09 to 0.2456. Both models achieve nearly perfect NDCG scores when sentiment analysis is excluded, with SVD scoring 0.9987 and CoClustering scoring 0.9983.

Overall, incorporating sentiment analysis enhances the predictive accuracy of all models, with DeepMF showing the greatest improvement in RMSE and MAE, while SVD and CoClustering demonstrate superior ranking quality as reflected in their NDCG scores.

The Table 10 outlines the key hyperparameter settings used in the DeepMF (Deep Matrix Factorization) recommendation model, which are crucial for its performance and effectiveness. The configuration specifies an 8-fold cross-validation process, 20 training epochs, a batch size of 50, and a learning rate of 0.002. These hyperparameter choices reflect the need to balance model complexity, training time, and generalization performance, as the number of folds, epochs, and the batch size control the trade-offs between capturing data patterns and ensuring model robustness. By documenting the hyperparameter configuration, researchers and practitioners can better understand the experimental setup, reproduce the results, and evaluate the impact of various hyperparameter choices on the final outcomes, which is essential for the assessment and comparison of different sentiment-based recommendation models in the hotel domain.

4.4. Performance comparison of hybrid models

In our study, we compared the performance of two different hybrid recommendation models: the weighted hybrid model and the feature combination hybrid model. These models incorporate DeepMF for collaborative filtering, clustering for content-based filtering, and BERT prediction for sentiment analysis. By conducting this comparison, we aimed to understand which model performs better in generating personalized recommendations. We evaluated their performance using RMSE (Root Mean Square Error) as a measure of accuracy, allowing us to determine which model provides more accurate predictions and higher-quality recommendations. This comparison provides valuable insights into the effectiveness of different hybrid recommendation approaches and informs the design and optimization of recommendation systems across various domains.

4.4.1. Domain: Restaurant

The Table 11 includes several hybrid models, including a weighted hybrid approach and feature combination hybrid models. The weighted

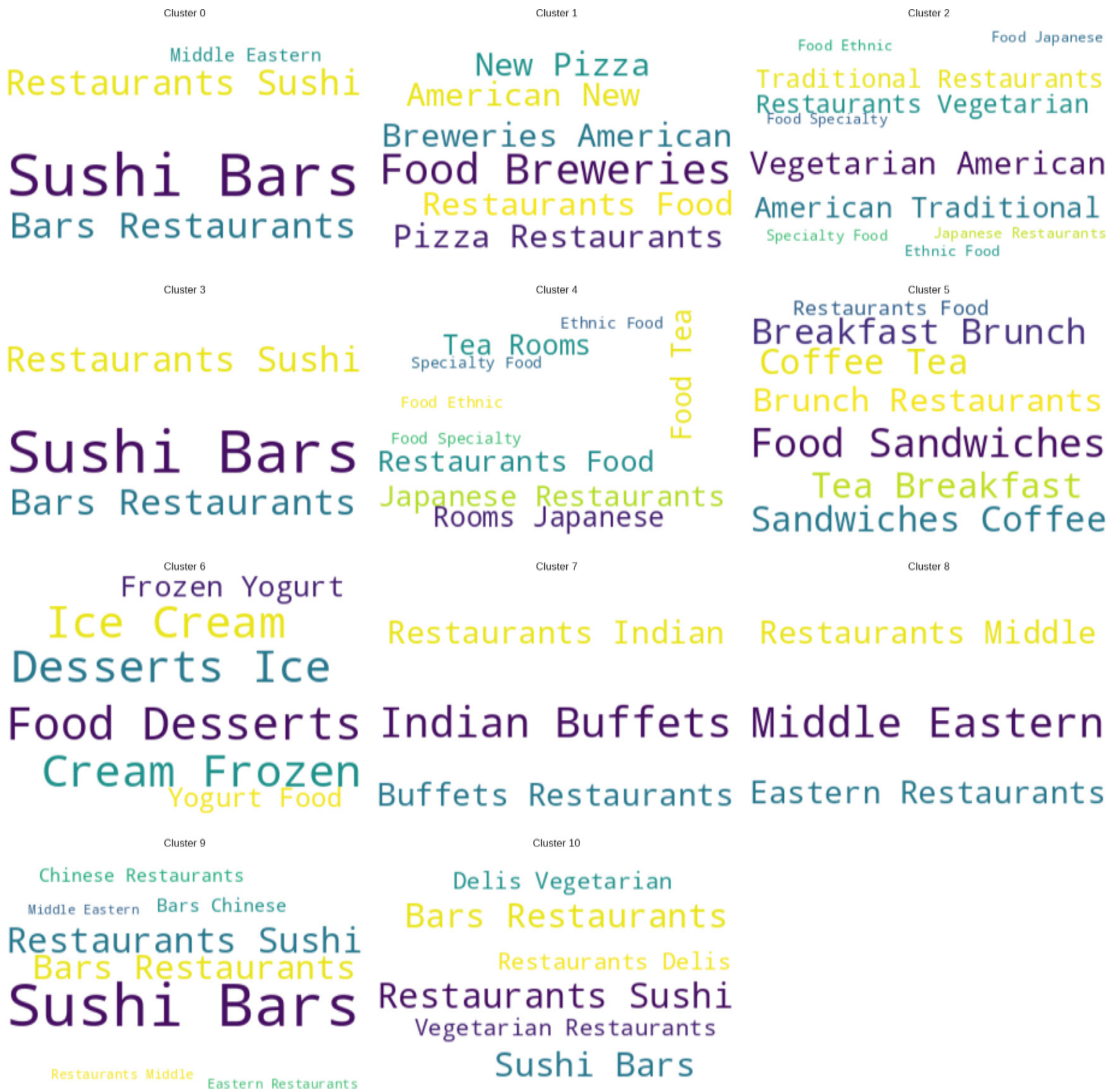

Fig. 8. WordCloud representation of restaurant clusters.

Table 9
Performance of sentiment-based recommendation models in the Hotel Domain.

Model	RMSE		MAE		NDCG	
	With SA	Without SA	With SA	Without SA	With SA	Without SA
DeepMF	**0.1186**	0.2542	**0.0700**	0.200	0.2400	**0.9978**
SVD	0.5504	1.1455	0.2522	0.923	0.2522	**0.9987**
CoClustering	0.5478	1.3233	0.2456	1.090	0.2513	**0.9983**

Table 10
DeepMF hyperparameter configuration.

Hyperparameter	Value
num folds	8
num epochs	20
Batch size	50
Learning rate	0.002

hybrid model combines the individual predictions from different techniques, while the feature combination hybrid models integrate the features from different approaches, such as Non-Negative Matrix Factorization (NMF) and various regression algorithms like RandomForestRegressor, GradientBoostingRegressor, Support Vector Regressor (SVR), and DecisionTreeRegressor.

The results show that the feature combination hybrid model, which combines NMF and DecisionTreeRegressor, achieved the best performance with an RMSE of 0.02. This suggests that the integration of the collaborative filtering approach (NMF) and the decision tree-based regression model was highly effective in accurately predicting restaurant ratings.

In contrast, the weighted hybrid model performed the worst with an RMSE of 1.28. This indicates that the weighted hybrid approach was not as successful in capturing the complex relationships between the different recommendation signals (collaborative, content-based, and sentiment-based) in the restaurant domain.

The table provides valuable insights for researchers and practitioners working on restaurant recommendation systems. By understanding the comparative performance of these hybrid models, they can make informed decisions about the most appropriate modeling approaches to

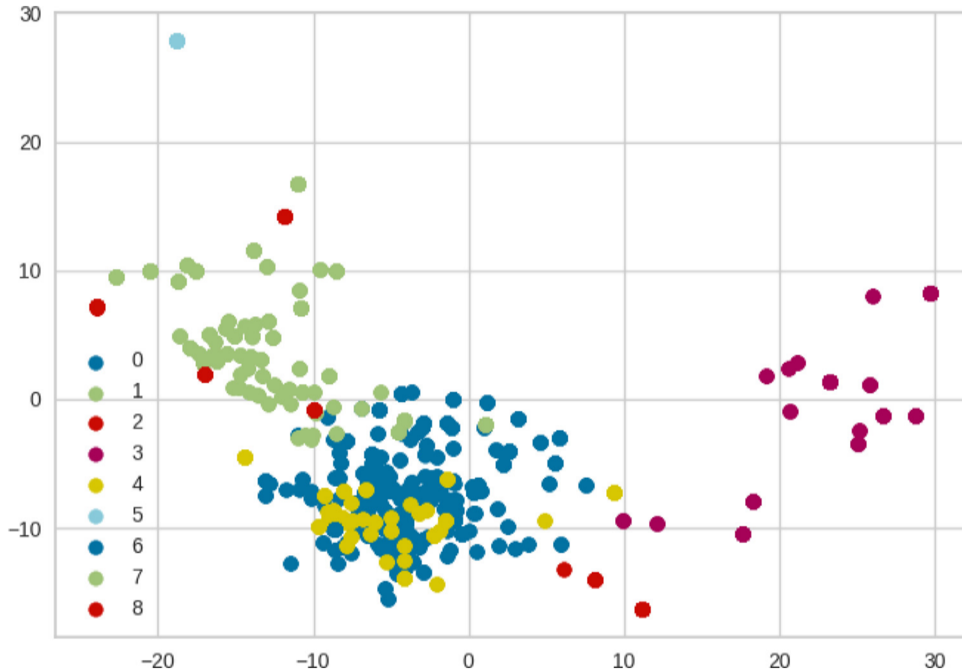

Fig. 9. Clustering results using AgglomerativeClustering.

Fig. 10. WordCloud representation of hotel clusters.

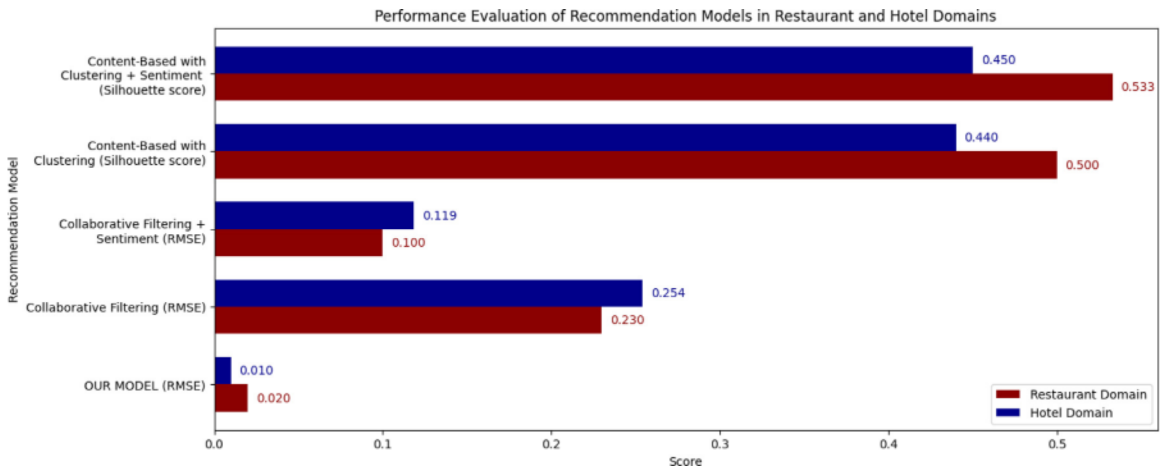

Fig. 11. Performance evaluation of collaborative filtering, content-based, and our model with sentiment analysis.

Table 11
Comparative performance of hybrid restaurant recommendation models.

Model	Performance (RMSE)
Weighted Hybrid	1.28
Feature Combination: NMF + RandomForestRegressor	0.51
Feature Combination: NMF + GradientBoostingRegressor	1.08
Feature Combination: NMF + SVR	1.17
Feature Combination: **NMF + DecisionTreeRegressor**	**0.02**

Table 12
Comparative performance of hybrid hotel recommendation models.

Model	Performance (RMSE)
Weighted Hybrid	1.43
Feature Combination: NMF + RandomForestRegressor	0.55
Feature Combination: NMF + GradientBoostingRegressor	1.18
Feature Combination: NMF + SVR	1.33
Feature Combination: **NMF + DecisionTreeRegressor**	**0.01**

adopt, depending on the specific requirements and constraints of their application.

4.4.2. Domain: Hotel

The Table 12 includes a weighted hybrid model as well as feature combination hybrid models, where the features from different techniques are integrated to leverage the strengths of diverse recommendation approaches. The models explored include combinations of Non-Negative Matrix Factorization (NMF) for collaborative filtering and regression algorithms such as RandomForestRegressor, GradientBoostingRegressor, Support Vector Regressor (SVR), and DecisionTreeRegressor. The results show that the feature combination hybrid model that integrates NMF and DecisionTreeRegressor achieved the best performance, with an RMSE of 0.01. This suggests that the integration of the collaborative filtering approach (NMF) and the decision tree-based regression model was highly effective in accurately predicting hotel ratings. In contrast, the weighted hybrid model performed the worst with an RMSE of 1.43. This indicates that the weighted hybrid approach was not as successful in capturing the complex relationships between the different recommendation signals (collaborative, content-based, and sentiment-based) in the hotel domain. These findings provide valuable insights for researchers and practitioners working on hotel recommendation systems. By understanding the comparative performance of these feature combination hybrid models, they can make informed decisions about the most appropriate modeling approaches to adopt, depending on the specific requirements and constraints of their application.

4.5. Final recommendation system

The Fig. 11 presents a comparative analysis of various recommendation models in the restaurant and hotel domains, evaluated using the RMSE (Root Mean Square Error) and Silhouette score metrics. In the restaurant domain, our model achieved the lowest RMSE of 0.02, significantly outperforming the other models. The Collaborative Filtering model had an RMSE of 1.2790, while the addition of sentiment analysis slightly improved its performance to an RMSE of 1.2603. Content-Based models with Clustering showed moderate Silhouette scores, with the sentiment-enhanced version scoring higher at 0.2270 compared to 0.1922 without sentiment analysis.

Similarly, in the hotel domain, our model also excelled with an RMSE of 0.01, demonstrating superior accuracy. The basic Collaborative Filtering model scored an RMSE of 2, which improved drastically to 1 with the inclusion of sentiment analysis. Content-Based models with Clustering exhibited improved Silhouette scores when augmented with sentiment analysis, increasing from 0.4 to 0.9. This analysis underscores the effectiveness of incorporating sentiment analysis into recommendation models, particularly highlighting the superior performance of our proposed model across both domains.

Figures below (12 and 13) showcases the results of our advanced recommendation model, featuring a comprehensive list of personalized recommendations specifically tailored for an individual user. Leveraging a combination of collaborative filtering, content-based clustering, and sentiment analysis techniques, our model has carefully analyzed the user's preferences, browsing history, and demographic information to curate this unique selection. Each recommendation has been meticulously chosen to align perfectly with the user's individual tastes and interests, guaranteeing an exceptional and enriching user experience.

5. Conclusion and future work

In conclusion, the work on integrating sentiment analysis (SA) into a recommendation system (RS) has yielded promising results. This study employed BERT for SA, DeepMF for collaborative filtering (CF) with SA, and CB clustering incorporating sentiment predicted by BERT. These techniques were combined using NMF with DecisionTreeRegressor, resulting in a recommendation model with an even lower RMSE of 0.02 for the restaurant domain and 0.01 for the hotel domain.

The individual components of the system demonstrated their effectiveness. The SA using BERT achieved an impressive accuracy rate of 90% for the restaurant domain and 87% for the hotel domain, enabling the system to accurately capture and incorporate user sentiments into the recommendation process. CF combined with SA yielded a significantly reduced RMSE of 0.1 for the restaurant domain and 0.1186 for

	user_id	business_id	business_categories	Hybrid_Recommendation
3379	0bNXP9quoJEgyVZu9ipGgQ	6oRAC4uyJCsJl1X0WZpVSA	Middle Eastern; Restaurants	5.0
2259	0bNXP9quoJEgyVZu9ipGgQ	tyETqrYijm3cY4noCwl9Ww	Middle Eastern; Restaurants	4.0
3420	0bNXP9quoJEgyVZu9ipGgQ	yktWUtKBja_Lzk3wwR6RFA	Middle Eastern; Restaurants	4.0
181	0bNXP9quoJEgyVZu9ipGgQ	6SMQl2vR37HvjYWwSl1V3w	Indian; Buffets; Restaurants	4.0
2637	0bNXP9quoJEgyVZu9ipGgQ	VYUfIfAlU1KuYGC6ITRY3A	Sushi Bars; Restaurants	2.0
3174	0bNXP9quoJEgyVZu9ipGgQ	T4ox3wUzFjWvyrjaTRoBbA	Sushi Bars; Restaurants	2.0
427	0bNXP9quoJEgyVZu9ipGgQ	JuBygU4XDjqSeW9okyvbsQ	Food; Sandwiches; Coffee & Tea; Breakfast & Br...	2.0
3454	0bNXP9quoJEgyVZu9ipGgQ	lVHtVCyFJOqAH0ltn7pLuw	Sushi Bars; Restaurants	2.0
87	0bNXP9quoJEgyVZu9ipGgQ	nMHhuYan8e3cONo3PornJA	Food; Tea Rooms; Japanese; Restaurants	1.0

Fig. 12. List of recommended restaurant domains.

	user_id	business_id	business_categories	Hybrid_Recommendation
3733	kGgAARL2UmvCcTRfiscjug	xm8F51qKjx0cfdOSukTDng	Hotels & Travel; Event Planning & Services; Ho...	5.0
1972	kGgAARL2UmvCcTRfiscjug	5AsVHsqADuurv7NqG-RFLA	Hotels & Travel; Event Planning & Services; Ho...	5.0
4145	kGgAARL2UmvCcTRfiscjug	SYqTY48DJa1cYhglvmgvsQ	Hotels & Travel; Event Planning & Services; Ho...	5.0
5072	kGgAARL2UmvCcTRfiscjug	Xe-zwcsX5EikIK6Ul5bGlw	Hotels & Travel; Event Planning & Services; Ho...	4.0
4742	kGgAARL2UmvCcTRfiscjug	xtX2qS64zP2NRPV_7NNqHw	Hotels & Travel; Resorts; Event Planning & Ser...	4.0
4529	kGgAARL2UmvCcTRfiscjug	ZRrIwqIUMIl1b2keMI8QRw	Hotels & Travel; Event Planning & Services; Ho...	4.0
756	kGgAARL2UmvCcTRfiscjug	koeabE4cC2YsJprWXUHdhA	Hotels & Travel; Event Planning & Services; Ho...	4.0
1077	kGgAARL2UmvCcTRfiscjug	KNIFSqzQADOZWDO_7T-KzA	Hotels & Travel; Event Planning & Services; Ve...	4.0
1564	kGgAARL2UmvCcTRfiscjug	CD-2yHTSObgvAhW-pYHfMw	Hotels & Travel; Event Planning & Services; Ho...	3.0
1147	kGgAARL2UmvCcTRfiscjug	-sAoGZTnFtDZUY9JYWHqlg	Hotels & Travel; Event Planning & Services; Ho...	3.0
1263	kGgAARL2UmvCcTRfiscjug	hW0Ne_HTHEAgGF1rAdmR-g	Hotels & Travel; Airports	2.0

Fig. 13. List of recommended hotel domains.

the hotel domain, indicating improved accuracy by considering user sentiments in the recommendation process.

Moreover, the CB approach with clustering, incorporating sentiment predicted by BERT, resulted in improved recommendations with a higher silhouette score of 0.533 for the restaurant domain and 0.42 for the hotel domain. This suggests that incorporating sentiment information enhances the clustering performance and improves the quality of CB recommendations.

To further enhance the system's performance, the proposed novel approach that combines these sentiment-aware components – BERT for SA, DeepMF for CF with SA, and CB clustering incorporating sentiment – using NMF with DecisionTreeRegressor has outperformed the individual techniques, achieving an even lower RMSE of 0.02 for the restaurant domain and 0.01 for the hotel domain. This indicates that the model effectively leverages the advantages of each approach to provide personalized and sentiment-aware recommendations.

Moreover, for mitigating the cold start problem, sentiment analysis plays a crucial role by providing insights into the preferences of new users, enabling the system to make informed recommendations from the start. Additionally, to address data sparsity, we employed techniques such as collaborative filtering and content-based clustering, which utilize user–item interactions and textual attributes, respectively, to generate recommendations even in sparse data scenarios.

In conclusion, the integration of SA into RS, as demonstrated in this work, shows promise for enhancing recommendation accuracy and personalization across the restaurant and hotel domains. Continued research and refinement of the proposed model can contribute to the advancement of sentiment-aware RS, benefiting users in these domains.

Future work can focus on several areas to further enhance the recommendation system. Firstly, exploring alternative sentiment analysis models or fine-tuning the existing BERT model could potentially improve the accuracy of sentiment classification. Additionally, investigating different fusion methods or incorporating advanced ensemble techniques may lead to further improvements in integrating the individual components.

CRediT authorship contribution statement

Nossayba Darraz: Conceptualization, Methodology, Validation, Writing – original draft, Visualization, Writing – review & editing. **Ikram Karabila:** Conceptualization, Validation, Methodology. **Anas El-Ansari:** Conceptualization, Validation, Methodology. **Nabil Alami:** Conceptualization, Validation, Methodology. **Mostafa El Mallahi:** Conceptualization, Validation, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We used the data for our research by downloading it from this link: https://data.world/brianray/yelp-reviews/workspace/file?filename=yelp_training_set_review.csv.

References

Al-Ghuribi, S. M., & Noah, S. A. M. (2019). Multi-criteria review-based recommender system–the state of the art. *IEEE Access*, *7*, Article 169446-169468.

Asani, E., Vahdat-Nejad, H., & Sadri, J. (2021). Restaurant recommender system based on sentiment analysis. *Machine Learning with Applications*, *6*, Article 100114.

Bhatt, B., Patel, P. J., & Gaudani, H. (2014). A review paper on machine learning based recommendation system. *International Journal of Engineering Development and Research*, *2*(4), 2321–9939.

Dang, C. N., Moreno-García, M. N., & Prieta, F. D. L. (2021a). An approach to integrating sentiment analysis into recommender systems. *Sensors*, *21*(16), 5666.

Dang, C. N., Moreno-García, M. N., & De la Prieta, F. (2021b). Using hybrid deep learning models of sentiment analysis and item genres in recommender systems for streaming services. *Electronics*, *10*(20), 2459.

Darraz, N., Karabila, I., El-Ansari, A., Alami, N., & El mallahi, M. (2024). Optimizing hybrid recommendations: VADER-enhanced sentiment analysis. In *Proceedings of the 7th international conference on networking, intelligent systems and security* (pp. 1–7). http://dx.doi.org/10.1145/3659677.3659748.

Darraz, N., Karabila, I., El-Ansari, A., Alami, N., Lazaar, M., & El Mallahi, M. (2023). Using sentiment analysis to spot trending products. In *2023 sixth international conference on vocational education and electrical engineering* (pp. 48–54). IEEE.

Darraz, N., Karabila, I., El-Ansari, A., et al. (2024). Detecting trending products through moving average and sentiment analysis. *Multimedia Tools and Applications*, http://dx.doi.org/10.1007/s11042-024-19430-2.

Deac-Petruşel, M., & Limboi, S. (2020). A sentiment-based similarity model for recommendation systems. In *2020 22nd international symposium on symbolic and numeric algorithms for scientific computing* (pp. 224–230). IEEE.

Duong, H. T., & Nguyen-Thi, T. A. (2021). A review: preprocessing techniques and data augmentation for sentiment analysis. *Computational Social Networks*, *8*(1), 1–16.

Durairaj, A. K., & Chinnalagu, A. (2021). Transformer based contextual model for sentiment analysis of customer reviews: A fine-tuned BERT. *International Journal of Advanced Computer Science and Applications*, *12*(11).

Elahi, M., Kholgh, D. K., Kiarostami, M. S., Oussalah, M., & Saghari, S. (2023). Hybrid recommendation by incorporating the sentiment of product reviews. *Information Sciences*, *625*, 738–756.

Fayyaz, Z., Ebrahimian, M., Nawara, D., Ibrahim, A., & Kashef, R. (2020). Recommendation systems: Algorithmsm, challenges, metrics, and business opportunities. *Applied Sciences*, *10*(21), 7748.

Geetha, M. P., & Renuka, D. K. (2021). Improving the performance of aspect based sentiment analysis using fine-tuned bert base uncased model. *International Journal of Intelligent Networks*, *2*, 64–69.

Giabelli, A., Malandri, L., Mercorio, F., Mezzanzanica, M., & Seveso, A. (2021). Skills2Job: A recommender system that encodes job offer embeddings on graph databases. *Applied Soft Computing*, *101*, Article 107049.

Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment analysis. *Procedia Computer Science*, *17*, 26–32.

Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, *16*(3), 261–273.

Kalaivani, K. S., Kanimozhiselvi, C. S., Narmatha, B. R., & Bilal, Z. M. (2023). Improving the restaurant recommendation performance using sentiment analysis. In *2023 5th international conference on inventive research in computing applications* (pp. 296–300). IEEE.

Katarya, R., & Verma, O. P. (2017). An effective collaborative movie recommender system with cuckoo search. *Egyptian Informatics Journal*, *18*(2), 105–112.

Ko, H., Lee, S., Park, Y., & Choi, A. (2022). A survey of recommendation systems: recommendation models, techniques, and application fields. *Electronics*, *11*(1), 141.

Kula, S., Choraś, M., & Kozik, R. (2021). Application of the BERT-based architecture in fake news detection. *vol. 12*, In *13th international conference on computational intelligence in security for information systems* (pp. 239–249). Springer International Publishing.

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

Liu, N., & Zhao, J. (2023). Recommendation system based on deep sentiment analysis and matrix factorization. *IEEE Access*, *11*, 16994–17001.

Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender system application developments: a survey. *Decision support systems*, *74*, 12–32.

Mehta, R., & Rana, K. (2017). A review on matrix factorization techniques in recommender systems. In *2017 2nd international conference on communication systems, computing and IT applications* (pp. 269–274). IEEE.

Nawara, D., & Kashef, R. (2020). Iot-based recommendation systems–an overview. In *2020 IEEE international IOT, electronics and mechatronics conference* (pp. 1–7). IEEE.

Patel, J., & Chhinkaniwala, H. (2022). The role of sentiment analysis in a recommender system: a systematic survey. *International Journal of Web Engineering and Technology*, *17*(1), 29–62.

Patel, B., Desai, P., & Panchal, U. (2017). Methods of recommender system: A review. In *2017 international conference on innovations in information, embedded and communication systems* (pp. 1–4). IEEE.

Raeesi Vanani, I., Mahmoudi, L., Jalali, S. M. J., & Pho, K. H. (2022). Using text mining algorithms in identifying emerging trends for recommender systems. *Quality & Quantity*, *56*(3), 1293–1326.

Ray, B., Garain, A., & Sarkar, R. (2021). An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews. *Applied Soft Computing*, *98*, Article 106935.

Sahu, S., Kumar, R., MohdShafi, P., Shafi, J., Kim, S., & Ijaz, M. F. (2022). A hybrid recommendation system of upcoming movies using sentiment analysis of youtube trailer reviews. *Mathematics*, *10*(9), 1568.

Selmene, S., & Kodia, Z. (2020). Recommender system based on user's tweets sentiment analysis. In *Proceedings of the 4th international conference on e-commerce, e-business and e-government* (pp. 96–102).

Sousa, M. G., Sakiyama, K., de Souza Rodrigues, L., Moraes, P. H., Fernandes, E. R., & Matsubara, E. T. (2019). BERT for stock market sentiment analysis. In *2019 IEEE 31st international conference on tools with artificial intelligence* (pp. 1597–1601). IEEE.

Vinay, D. R., Kumaraswamy, B. V., & Basavaraddi, C. C. S. (2021). Machine learning based recommendation system on movie reviews using KNN classifiers. *Journal of Physics: Conference Series*, *1964*(4), Article 042081.

Wei, J., He, J., Chen, K., Zhou, Y., & Tang, Z. (2017). Collaborative filtering and deep learning based recommendation system for cold start items. *Expert Systems with Applications*, *69*, 29–39.

Xie, S., Cao, J., Wu, Z., Liu, K., Tao, X., & Xie, H. (2020). Sentiment analysis of chinese e-commerce reviews based on bert. *vol. 1*, In *2020 IEEE 18th international conference on industrial informatics* (pp. 713–718). IEEE.

Zhang, Q., Lu, J., & Jin, Y. (2021). Artificial intelligence in recommender systems. *Complex & Intelligent Systems*, *7*, 439–457.

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), Article e1253.

Zhou, J., Wen, J., Li, S., & Zhou, W. (2019). From content text encoding perspective: A hybrid deep matrix factorization approach for recommender system. In *2019 international joint conference on neural networks* (pp. 1–8). IEEE.

Ziani, A., Azizi, N., Schwab, D., Aldwairi, M., Chekkai, N., Zenakhra, D., et al. (2017). Recommender system through sentiment analysis. In *2nd international conference on automatic control, telecommunications and signals*.