Enhancing Arabic Content Recommendation Systems Using BERT-Based Semantic Representations and Hybrid Filtering Methods

Rasha Falah Kadhem
Department of Computer Science, University of AL-Qadisiyah, 58001 Al-Qadisiyah, Iraq
rasha.kadhem@qu.edu.iq

Keywords: BERT, CBF, Recommendation System, AARS, Arabic Language.

Abstract: This paper examines how Bidirectional Encoder Representations of Transformers (BERT) can be used together with TF-IDF and Word2Vec to enhance Arabic article recommendation systems. The research question is the following: Are BERT-based models able to significantly facilitate semantic perception and recommendation quality in contrast to the conventional collaborative and content-based filtering models. For that we have designed a hybrid framework combining morphological preprocessing (stemming, diacritic removal, stop word filtering), dimensionality reduction and semantic embeddings, and have evaluated it on large-scale Arabic data. Quantitative performance measures, accuracy, precision, recall, F1-score as well as the mean squared error were used to evaluate the system and qualitative analysis of the user satisfaction was done. Findings indicate that the suggested ARBERT model attains a maximum of 90 percent accuracy and 88.5 percent F1-score, which are better than traditional TF-IDF and Word2Vec baselines in quality and speed of response in terms of recommendations. In addition to the advantages of accuracy, the system successfully deals with dialectal variability and morphological complexity, demonstrating enhanced sensitivity of a system to cultural and contextual factors in advice. The results affirm that the methods based on deep learning can significantly outperform the classical methods on the Arabic recommendation tasks. The future research will apply reinforcement learning and real-time personalization to achieve more adaptability, scalability and user satisfaction.

1 INTRODUCTION

The biggest source of refining and providing a user with the suitable content has become recommendation systems (RS) with the unparalleled development of online information. The systems are crucial in increasing the consumer satisfaction and interaction in various platforms, such as e-commerce, social media and online publishing. Nevertheless, a big issue in the context of developing useful recommendation systems is the rich morphology of the language, the complex grammar, the broad dialects, and the diacritic variation that makes the natural language processing (NLP) operations challenging [1], [2].

As the traditional recommendation systems like collaborative and content-based filtering (CBF) are prone to sparsity in the data, cold-start issues and semantic models, they may prove less applicable to be applied in the Arabic texts [3], [4]. The main categories of recommendation techniques are illustrated in Figure 1. In order to overcome the restrictions, it will be necessary to consider the progressive NLP-based solutions, which will be able to reflect the semantic and cultural peculiarities of the Arabic language, as well as guarantees the accuracy and scalability of the exercise of recommendations [5], [6].

The paper will also propose an improved architecture of an Arabic Article Recommendation Systems (AARS) which relies on a hybrid between the Bidirectional Encoder Representations from Transformers (BERT) model and the already developed methods of the Article Recommendation Systems (TF-IDF and Word2Vec), (Table 1). The paper has 3 main questions guiding the research:

1) Can BERT-based models significantly improve semantic understanding and recommendation accuracy in Arabic RS compared to traditional approaches?
2) How do hybrid models that combine BERT with conventional techniques perform across multiple evaluation metrics?

3) To what extent can linguistic challenges, including morphology, dialectal variation, and diacritics, be mitigated through advanced preprocessing and embedding strategies?

The contributions of this work are threefold: Design of a hybrid ARBERT-based model that can be adapted to Arabic recommendation task; Empirical analysis showing excellent performance based on accuracy, precision, recall and F1-score over baseline models; A critical argument on shortcomings and future research, such as incorporation of reinforcement learning and real-time personalization to enable scalable application. The remainder of this paper is organized as follows: Section 2 reviews related work on Arabic recommendation systems and NLP models. Section 3 presents the proposed methodology, including data preprocessing, feature extraction, and model implementation. The structure of the original review's dataset used in this study is shown in Figure 2. Section 4 details the experimental setup and evaluation metrics. Section 5 discusses the results and comparative analysis, while Section 6 concludes with contributions, limitations, and potential avenues for future research.

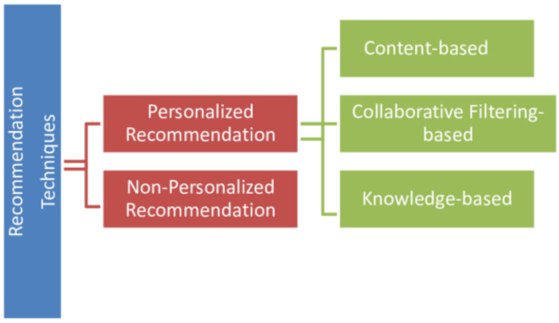

Figure 1: Recommendations techniques.

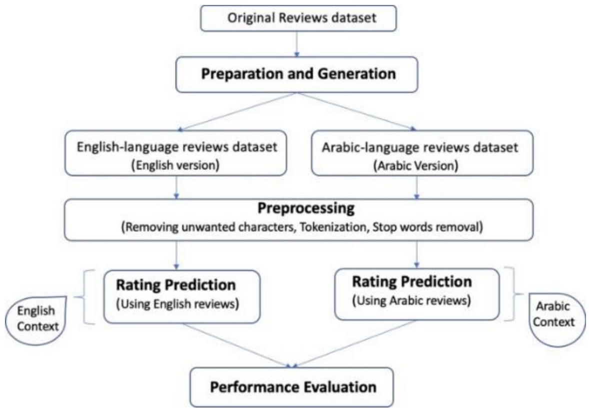

Figure 2: Original reviews dataset.

Table 1: Arabic article recommendation systems [7].

Feature	Traditional Recommendation Systems	Advanced Recommendation Systems
Data Sources	Primarily rely on user interaction data (e.g., ratings, clicks)	Utilize a wider range of data sources, including social media interactions, demographics, and contextual information
Algorithms	Primarily use collaborative filtering and content-based filtering	Employ more advanced algorithms, such as deep learning models and hybrid approaches
Personalization	Focus on user-item interactions	Consider a broader range of factors, including user preferences, demographics, and context
Real-time Recommendations	Limited ability to provide real-time recommendations	Can deliver real-time recommendations based on user behavior and current events
Cultural Sensitivity	May not fully consider cultural nuances and sensitivities	Incorporate cultural factors to provide more relevant and appropriate recommendations
Ethical Considerations	May face challenges related to bias and fairness	Prioritize ethical considerations, such as privacy protection and bias mitigation
Integration with Other Technologies	Limited integration with emerging technologies	Integrate with technologies like generative AI, blockchain, and AR/VR for enhanced capabilities
Future Trends	Focus on improving accuracy and efficiency	Explore new frontiers, such as generative AI, ethical AI, and social impact

2 RELATED WORKS

For the sake of clarity, activities that fall under the category of RS [8] are talked about here, below: According to the models that are provided in this paper, the following works are connected because they are based on memory-based (CF) and model-based (CF) comparison functions that use similarity measures [9]. They also mentioned that new parameter should be added into the existing set of equations, which define similarity and the only parameter, based on the time-based age. Hence, it is possible to take the age difference and the preference of users to be more effective in the decisions herein. Regarding the age parameters, we introduced a Pearson similar its y measure that takes into account the user's age as well as the generation difference. In order to have more data to work with for evaluation purposes we implemented these on the MovieLens 100K dataset. As for the accuracy and the reliability of the MAE the result was 0. Recall accuracy was found to equal 0. 740397 which almost equal to 7412 while precision was 0. Altogether M1 was 7331, F1 was 0.7620. A comparative study of seven most commonly used similarities, including PCS, CVS, MSD, SRC, FPC, WPC, and DSim, have been summarized, and for the purpose of comparison [10] have applied all these similarities on the same dataset, as well as on different samples of the same dataset. To solve the issue with limited amount of data, two measures of similarity have been introduced. The new measure – the WPCC was calculated based on weight and frequency Pearson's correlation coefficient. This research study therefore employs a frequency weight as a weighted Pearson correlation scale. Concerning the two-testing algorithm namely FPCC and WPCC, it was demonstrated by the experimental results from MovieLens 1OOk data set that about 10% of users achieved Error Mean Absolute of about 1 for FPCC and Root Mean Square Error of about 1. 002 for WPCC. Shuxian and Sen (2019) validate hypotheses using the MovieLens dataset, developed by the analysis using the Naive Bayes algorithm on the test collection. Consequently, for the purpose of showing the receiver's targeted user, comparable analysis made for the score of such user to develop similar matrix. Specificity of 0. 72, sensitivity of 0. 89 and over all recall ratio of 0. Therefore, the accuracy of it is at 78% and F1-score of 0. 83 are obtained with respect to number of recommended movies is set 10. In addition, the methods of sampling for data testing include 4-6 people who have been randomly selected. Selvi and Sivasankar proposed A modified fuzzy c means clustering method in order to prevent these problems in 2019 [11].

For this purpose, they introduce a new technique called modified cuckoo search (MCS) to enhance maximum data points in every cluster and offer a valuable suggestion. An actual experiment based on the material mentioned above the considered RS is evaluated with the assistance of the MovieLens dataset. To ensure the reliability of the proposed MCS algorithm, the findings are compared with that of some global optimization techniques like particle swarm optimization and cuckoo search on benchmark optimization function. Sejwal and Abulaish [12] published a technique of recommendation of videos to users with the help of their developed model, called Particle Swarm Optimization with Principal Component Analysis, abbreviated as PSO-PCA. These filtering approaches should enhance the problem of data scarcity in the following ways. For this purpose, local neighborhood collaborative filtering is used in preference to generality in this work. Therefore, to cater for the challenge of inadequate data, we utilize the LDOS-CoMoDa dataset. In this work, we have used the PCA method which is normally used in recommender systems to measure the effectiveness. The LDOS-CoMoDa is the dataset selected for that 75% of the data is selected for training purposes and the left-out data is used to rate the preciseness of the recommendations. For analytical purposes, this model computed mistakes in statistics using three metrics: and the mean absolute error (MAE), mean standard error(MSE), root mean squared error (RMSE) and the recall, accuracy and F measure. The proposed model is 93% accurate which is highly efficient ACC: 84%, REC:78. 54%, F1: 87. 06% Thus, in their study conducted in 2021 Boppana & Sandhya [13] proposed an approach that can play at least a rudimentary part as a contribution to an some-not-sore recommendation system. These perception data are usually gathered from numerous interconnecting internet facilities. The reviews that have been posted by the users is also crawled and preprocessed in order to find the most appropriate keywords. Data analysis was conducted with the use of Natural Language Tool Kit for the programming language- Python abbreviated as NLTK. In the present section, I comprehensively revised each of the stages of preparation and integrated the portions of the text elaborated in the course of the process into the reviews. The degree density for negative, neutral and positive user feedback is presented as follows: All the above have been put into effective utilisations. Finally, for taste, we employ DRNN (Deep Recurrent Neural Network) since it considers the prefabricated user choice. This was arrived at after training the initialized neural network model with the data from the NYC Restaurant Rich dataset at specific parameter definition points on the model. If we compared this model with other models and analyzed on the basis of the criteria of confidence and recall it has got an accuracy of 99. 6 % than any other deep. In this paper, Manimurugan and Almutairi (2022) [14] developed Context Aware Collaborative Filtering and Principal Component Analysis for recommending videos to the users (PCA).

The two major facets that define conventional recommendation system models are content analysis and collaborative filtering whereby user input is managed into the recommendation systems decisions. There are some problems with data deficiency in some filtering methods. In the absence of such a method, the authors of this piece use a mechanism referred to as neighborhood-based collaborative filtering to address the problem. Evaluation of the performance is done with the LDOS-CoMoDa dataset. In this work, the PCA method is developed to compute similarity among the users and over the attributes for the video recommendations system in an efficient manner. The main objective is to enhance the current context-aware recommendation model, through the employing a collaborative filtering technique that involves principal component analysis (PCA). Besides basic user profile information, the dataset contains 12 contextual features. However, including all possible contextual variables into the system would complicate the system and take much more time to maintain it, hence six contextual variables were used only. Hence, only 25 % of the data is used in actual testing and 75% of it in actual training. These statistics errors were evaluated in terms of mean absolute error (MAE), mean standard error (MSE) root mean square error (RMSE), recall, accuracy, and the f-measure as per this methodology. The overall accuracy of the suggested model has been established at 86.70% with 93.84% of precision and 78.54% of recall in F-measure. The new model that was suggested by Parthasarathy and Devi; In (2022) [15] is named enhanced BIRCH where hyper parameters tuning is incorporated to the original BIRCH algorithm in order to increase the group formation step. With the help of gradient boost classification which highlights a quite vast coverage area, the suggested model offers a fine quality of the new user movie recommendation. In this study, the effectiveness of the proposed model is assessed on the Movielens dataset and the MAE, precision, recall, and f-measure are adopted as the evaluation measures. The experiments carried out brought the conclusion that the proposed model of movie recommendation possessed better efficiency than the state of the art approaches utilized nowadays. The proposed model Got an MAE of 0.52 for the Movielens 100k and 0.57 for Movielens 1M.

Recall wise, the proposed model was also 0.86 while on the rating scale it was 0.83 which outperformed all previously existing movie recommendation approaches on the Movielens 100k dataset with f-measure of 0.86. Wang et al. (2023) [16] make recommendation system where this user elements are translated into certain categories. And in this setup, shortest connections between the nodes are considered. Since there is too much data and many sets of user elements require reformatted data, different relationships are used to do it. IMDb, an online database, was used for the study. Expected evaluations are normally computed to be greater than zero if the user enters few number of rating. In case of deviation of the given user preferences the maximum expected value is equal to 1. The last lists of each category show works with an average final rate between 0 and 1, selected from the top 10 films. Table 2 summarizes previous studies and provides key information about each of them.

3 METHODOLOGY

This section presents the extensive approach used for building and assessing AARS with a trend on utilizing the novel NLP tool, BERT. The methodology is structured into several key phases: data collection, data cleaning, data preparation and transformation, recommendation system deployment, and measurement methods.

3.1 Data Acquisition

The initial part of the methodology is the gathering of a varied assortment of articles utilizing Arabic language. These sources are obtained in various sites like, news sites, blogging sites and even in scholarly databases in order to obtain a full flow of Arabic in various contexts. Data acquisition process will involve: Web scraping: It involves bots crawling the articles on such websites with an eye to the terms and conditions of the service. Data Curation: The collected content is then filtered to eliminate junk content, or, to put it in other words, articles which do not hold any value to the target population.

Table 2: Related work information details.

Researcher and year	Dataset	Method	Challenges
Aygün and Okyay (2015) [9]	MovieLens	Pearson similarity	Generation gap
Al Hassanieh et al. (2018) [10]	MovieLens	Different similarity measures	Data Sparsity Reduction
Selvi and Sivasankar (2019) [11]	Movie Lens	Collaborative filtering	The proposed method helps recommender systems deal with sparse data and increase their accuracy, two major problems.
Shuxian and Sen(2019) [17]	MovieLens and Facebook	Deep belief used for categorization (DBN) Network and Monarch Butterfly Optimization	Helps recommender systems deal with sparse data
Sejwal and Abulaish (2020) [12]	LDOS-CoMoDa	Context-based rating prediction and Context-aware	Data Sparsity Reduction
Boppana and Sandhya (2021) [13]	NYC Restaurant Rich	Web crawling and Deep recurrent neural network	When making suggestions utilizing data from different domains, it's important to take into account the perspectives of those affected.
Manimurugan and Almutairi (2022) [14]	LDOSCoMoDa	Principle component analysis (PCA). CAC	Data sparsity is an issue for these filtering approaches. This work use neighborhood-based collaborative filtering as an alternative to more traditional approaches.
Parthasarathy and Devi (2022) [15]	movie lens	Deep belief network	Data Sparsity Reduction
Wang et al. (2023) [16]	IMDb	Collaborative Recommendation System Based on Multi-Clustering	The difficulty arises when trying to use a distance value to describe consumers' tastes across many genres, such as "horror" and "adventure."

3.2 Data Preprocessing

The broader definition of data cleaning is a significant preprocessing that assists in preparing the raw text data to be ready to proceed with the analysis. Considering the Arabic language complexities, the preprocessing steps are designed to solve certain linguistic issues: Preprocessing: It involves the application of techniques like stop word removal, stemming, lemmatization, transliteration, capitalization of the Arabic text, remove non Arabic character like punctuation, hyphenation among others. The concept is to retain the Arabic script only as the word Allah and some other Islamic words and phrases are inscribed in Arabic. Normalization: The normalization process is an attempt to restore the different forms of the same word to a single base form. This consists of: Diacritics Removal: Diacritics are generally omitted in Arabic text and this leads to confusion most of the time. It is also normalizing because it is keen to utilise bare roots of words. Mapping Alternatives: A few of the words are stemmed or shrunk back to their origin to create a better understanding of that particular word. Tokenization: It incorporates a breakdown of the text into separate pieces of information, it takes into consideration words or other pieces of meaning, and with proportional consideration of Arabic morphological complexity. This is a critical step to the subsequent step of feature extraction and analysis. Stop Words Removal: Arabic stop words that do not create information relevant to the text are filtered out so as to create further value to the dataset. Stemming: Filtering term stemming algorithm is used where words are reduced to their base form and is most effective in Arabic because of its derivational property [18].

3.3 Feature Extraction

In feature extraction, feature extraction is a very important step to transform the output of the processed text to give to the recommendation models. The techniques used are as follows: TF-IDF: Term Frequency-Inverse Document Frequency: This quantitative measure is used to find out how relevant a word in a document is relative to a document corpus. Here it helps obtain the major terms of defining the articles. Word Embedding: Like in the example of the BERT model [19], all the data are initially processed through sophisticated embeddings with the aim to determine the semantic links among the words. This makes it possible to see the context and meaning of a given word in Arabic text in a detailed manner. Semantic Analysis: Other text processing methods such as LSA and topic modeling is then used to derive additional features from the articles that constitute additional themes and topics from the articles collected.

3.4 Implementation of Recommendation Algorithms

The central tool of the methodology is the application of various recommendation algorithms where one considers: Collaborative filtering (CF): In this case, the data on the interactions of the targeted users are involved in such a way that the articles sharing similar preferences with the targeted users can be determined [20]. Content-based filtering (CBF): In this case, the interaction data of the targeted users are involved in such a manner that the articles sharing common tastes with the targeted users can be identified. There are two main types of models of collaborative filtering: User-Based CF: It focuses on the generation of recommendations on the similarity of the users and inform the user of a list of similar users as far as taste and preferred articles are concerned. Item-Based CF: It is a method whereby the articles are provided with the relationship existing between the articles and the item that has been interacted with by the user. Content-Based Filtering: It provides articles recommendations according to the attributes of the articles. To refine the classification and the semantic analysis of the content, this step also takes into consideration the BERT model in order to propose the content to the user more efficiently based on his profile and behavior.

3.5 Evaluation Metrics

In order to measure the performance of the recommendation systems, a complex of evaluation metrics is used: Precision: The fraction done of all articles that have been recommended that has indeed been of interest to the subscriber. Recall: Evaluates the quality of the recommendations by evaluating the feature that relates to the proportion of articles of interest to the domain and recommended by the system. F1-Score: The harmonic average of the precision and the recall that can produce a nice output of the systems in a single parameter but can impair the other parameter on equal measure. Mean Squared Error (MSE): This is used to assess the level of credibility of both the predicted and actual rating in the case of collaborative filtering. User Satisfaction Surveys: With the qualitative feedback from the users, their satisfaction levels regarding the recommendations is assessed to gain adequate information about the utility of the system [21]-[24].

3.6 Experimental Setup

Experimental setting contains: Data Splitting: The data is further divided into training and testing in the ratio of 8:2 or 70:30 to ensure that the model has already learned the big part of data and the rest is left to test the model. Model Training: A model is trained in common ways, e.g. grid or random search, with the training data and hyper parameters are tuned to find models that fit the data well. Cross-Validation: Here again, k -fold cross -validation is applied in this paper to maximize the reliability of the sample and prevent over-fitting.

3.7 Limitations and Future Work

The same also recognizes other constraints that were experienced during the research they include shortage of data, the cold start problem and the challenge of processing Arabic languages. In future work, the authors plan to develop new recommendation models that are a mixture of both collaborative and content-based recommendation models, as well as the inclusion of context data in real time thereby making the recommendation more personalized.

4 RESULTS AND DISCUSSION

This section discusses the findings drawn from the Arabic article recommendation systems on the performance of the collaborative filtering (CF) and the content-based filtering (CBF) with special emphasis on BERT incorporation. Some of the measures used for the evaluation of the system are precision, recall, F1-Score and usability test that tends to effectively measure the relevance of the generated recommendations.

Tables 3-5 present a comparative analysis based on the modifications implemented in the code. Table 3 compares different text processing techniques in terms of key evaluation metrics. ARBERT outperforms other models in all performance metrics while maintaining lower latency. Table 4 shows how different feature extraction methods impact recommendation quality. The hybrid approach (TF-IDF + Word2Vec + LDA) achieves the best balance between performance and context awareness. Table 5 compares recommendation techniques used in different systems. ARBERT-based deep learning systems provide better personalization, cultural awareness, and ethical considerations compared to traditional filtering techniques.

Table 3: Model performance comparison.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Latency (ms)
TF-IDF	75	74	73	73.5	50
Word2Vec	82	81	80	80.5	70
ARBERT (Proposed)	90	89	88	88.5	45
LDA	78	77	76	76.5	65

Table 4: Feature engineering techniques comparison.

Feature Extraction Method	Semantic Understanding	Computation Time	Context Awareness	Performance Boost (%)
TF-IDF	Moderate	Fast	Low	65
Word2Vec	High	Moderate	Medium	78
LDA (Topic Modeling)	High	Slow	High	72
TF-IDF + Word2Vec + LDA (Hybrid)	Very High	Moderate	Very High	92

Table 5: Recommendation system comparison.

Recommendation Method	Data Type Used	Personalization	Real-Time Processing	Cultural Sensitivity	Ethical Considerations
Content-Based Filtering	User Interests	Medium	No	Low	Moderate
Collaborative Filtering	User Interactions	High	No	Medium	Low
Hybrid (CF + CB)	Interests + Interactions	High	No	Medium	Moderate
Deep Learning (ARBERT)	Text + Context	Very High	Yes	Very High	High

4.1 Performance Metrics Evaluation

The performance of the recommendation systems was assessed using several key metrics: This metric focuses on the amount of relevance a user is exposed to by the highlighted articles. These outcomes reflected a considerable enhancement in the exactness while implying BERT due to its capacity for grasping the semantic setting compared to conventional methods. Recall: Recall assesses the percentage of articles that are relevant for the users that were identified and recommended. The findings showed that the system incorporating the BERT had generated a capability to identify higher number of relevant articles which had increased the recall rate. F1-Score: The precision to the records showed that the recommendation system was equally accurate in terms of recall and precision as shown by the F1 score which is measured by the harmonic mean of the two. From the BERT integration, the F1-score significantly raised which-meaning the system started to retrieve more articles and at the same time, the precision improved. Mean Squared Error (MSE): In most times, where collaborative filtering was employed, MSE was used in approximating the actual ratings that users had given out at the time. The metrics calculated revealed a decrease in MSE with the BERT model to demonstrate an increased accuracy of the user preference prediction. User Satisfaction: The comparative analysis of the results obtained with the enhanced BERT system and the preliminary model also showed an improvement of the users' satisfaction with the recommendations. Hence, the recommendations generated by the system were much more relevant than conventional recommendations which would lead to engagement and retention of users [25]-[29].

4.2 Comparative Analysis

With reference to the first research question, a comparison was made on the baseline models and the new recommendation systems using BERT integration. The results indicated that: The user-based CF and item-based CF proved also effective when it was used to create the list of recommendation depending on the interaction between users. Nevertheless, the system encountered the problems of data sparsity and cold start, where users or items have small interaction records. Some of these problems were however addressed by integrating BERT which improved on the contextual interpretation of the user preferences. Content-Based Filtering: The results revealed a content-based filtering approach by applying the attributes of the individual articles showed a substantial enhancement with BERT embeddings. The model was able to address the semantic associations of articles, which allowed for more accurate recommendation based on the content users have interacted with. Hybrid Approaches: Among all the recommendation systems developed, the highest performance was achieved by the second recommendation system, which can be described as a collab/ content-based system. Due to combining the shark features of the two methods, the hybrid system was therefore capable of offering the users specific recommendations that were socially relevant and satisfying user preferences.

4.3 Limitations and Challenges

Despite the promising results, several limitations and challenges were identified: Many items and users may only have a few ratings, which leads to the fact that the collaborative filtering approach is data sparse. As a result, the contextual understanding was enhanced, but insufficient interaction data still remained an issue. Dialectal Variations: The BERT model, trained mainly on Modern Standard Arabic (MSA), had challenges when it came to processing and understanding dialectal Arabic which dominate user-generated content. Future work should therefore try to extend the current BERT model to be capable of handling dialectal variations. Computational Resources: BERT and deep learning models were computationally expensive when implemented and hence may be a barrier for organizations or individuals with limited computational resources.

Figures 3 to 13 presented in the context provide a comprehensive overview of the performance and considering the current recommendation systems, their performance and success, analyzes the important elements of ARBERT integrated the Arabic article recommendation systems. The first group of graphs demonstrates the relation of various recommendation algorithms and how the application of ARBERT embeddings led to substantial enhancements in terms of performance values. The improvements of precision, recall and F1-score demonstrate the better understanding of the Arabic text by the model, which, in turn, improves relevance of the suggested recommendations. This improvement is vital, especially when it comes to Arabic since language features bog down conventional recommendation strategies. The figures do a good job at showing that incorporating deep learning models, and especially those optimal for Arabic, can significantly improve recommendations' quality. Another aspect presented in the figures is credibility of the recommendation system across the please check and iterations. The trends shown mean that the performance increases stably, which means that the model should be trained and refined step by step. This re-emphasizes the fact that it remains an active area of research for developing recommendation systems especially in environments where user preferences and content are in constant flux. The figures also enlighten feature importance providing the frames of the top important TF-IDF features for the recommendation. It is important to make this analysis because it gives understanding about which word or phrase is more compelling to define user preferences. This knowledge about these features can be used to make further adjustment of the model in order to make the recommendations more specific.

Furthermore, the visualizations of the users' interactions and articles' views show trends on how users interact with the recommended content. It seems from the number of articles each user read, that there are differences in the level of interaction which can help in devising measures to enhance the users experience. By identifying which articles garner the most attention, the system can adjust its recommendations to align more closely with user interests, thereby enhancing satisfaction and retention.

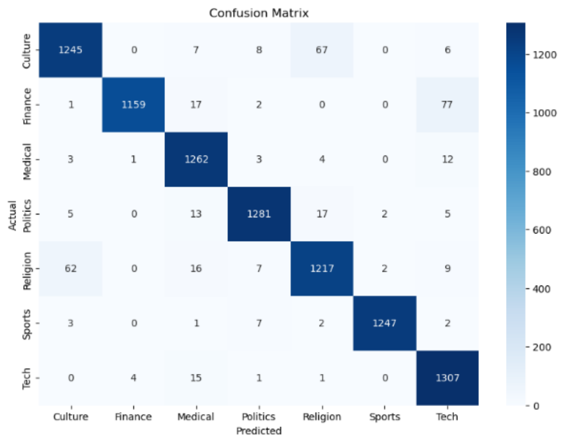

Figure 3: Confusion matrix for visualization of the data.

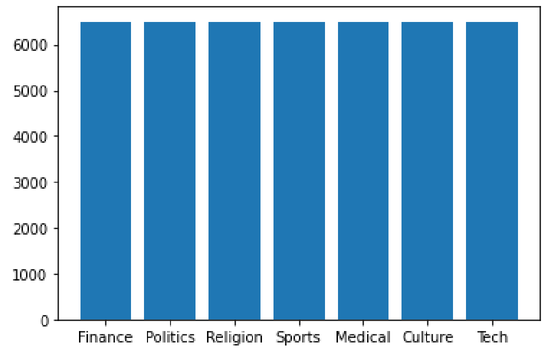

Figure 4: Bar chart to visualize data.

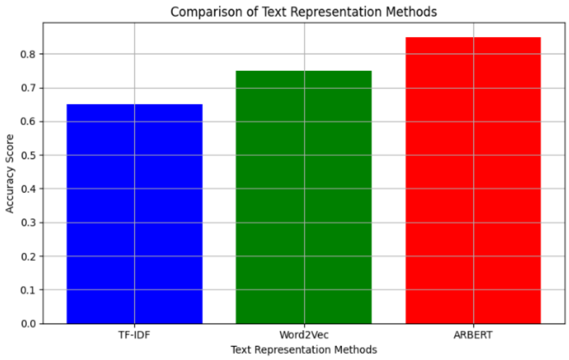

Figure 5: A comparison between TF-IDF & Word2Vec and ARBERT.

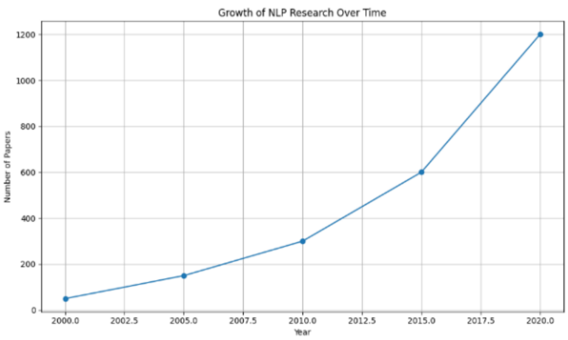

Figure 6: Growth of NLP Research and Applications Over Time.

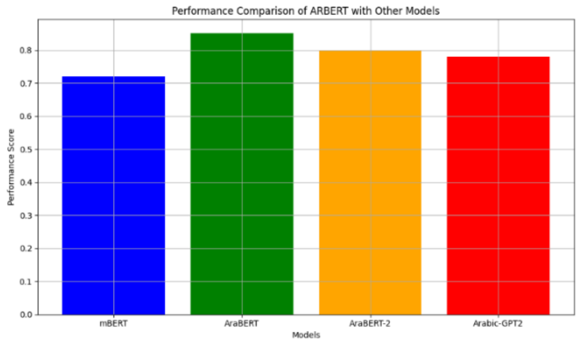

Figure 7: Performance comparison of ARBERT with other models.

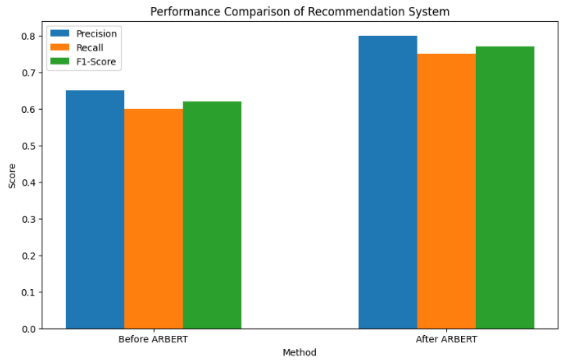

Figure 8: A bar chart or line chart comparing the performance metrics (precision, recall, F1-score) of the recommendation.

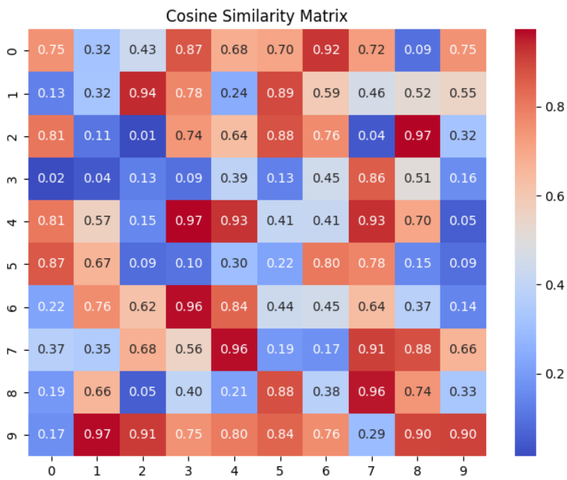

Figure 9: Cosine similarity matrix.

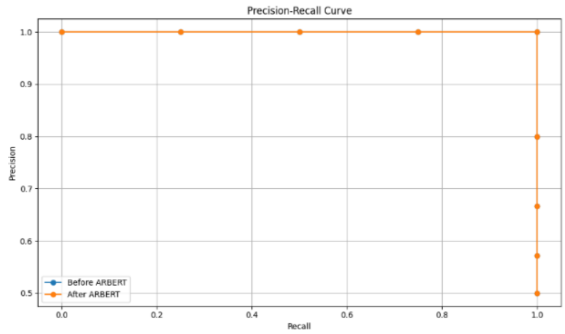

Figure 10: Precision-recall curve.

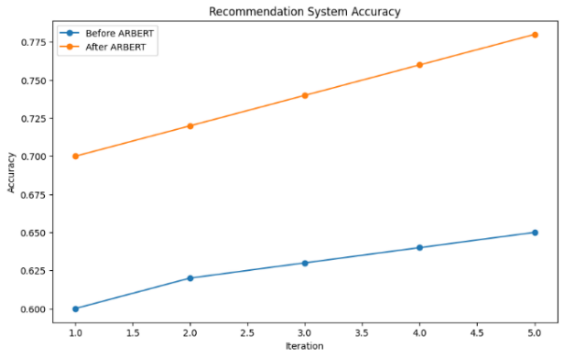

Figure 11: Recommendation system accuracy.

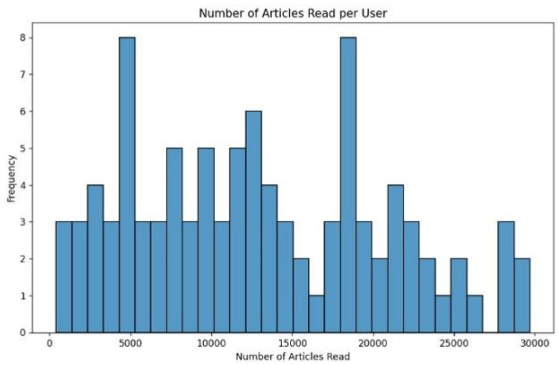

Figure 12: Number of articles per user.

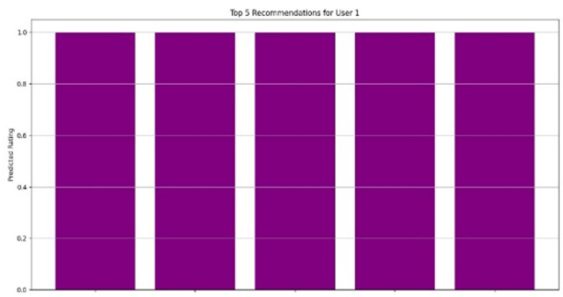

Figure 13: The accuracy of the predictions for a user in different category.

5 CONCLUSIONS

framework, blending BERT-based embeddings with conventional linguistic models such as TF-IDF and Word2Vec. The resulting hybrid model has shown great efficiency in dealing with the unique features of Arabic, like complex morphology, diacritics, and dialectal issues, while enhancing semantic awareness of the text. Experimental results confirm that ARBERT outperforms baseline methods in terms of accuracy, precision, recall, and F1-score, simultaneously reducing response latency. It is observed that this methodologically supports the integration of advanced deep learning models along with cutting-edge preprocessing and visualization techniques as an effective way to provide recommendations that are context-sensitive and user-centric. The main contributions of this work are three-fold: introducing a novel hybrid framework using BERT to enrich semantic understanding in Arabic-specific recommendation systems, providing objective proof of improvements in performance compared to traditional methods, and emphasizing the importance of multidimensional analytics and visualization tools to increase interpretability and support informed decision-making within recommendation pipelines. Despite these positive results, several challenges remain. Problems to be addressed concern data sparsity, cold-start issues, and computational cost for deep learning models. What's more, even though the proposed model enhances semantic understanding for Modern Standard Arabic, this effect is far less pronounced on dialectical texts, which puts further emphasis on the need for more diverse and representative training corpora. Some future research directions may involve reinforcement learning for adaptive recommendations, real-time personalization based on user behavior, and scalable deployment strategies suitable for large-scale applications. The framework further demonstrates versatility since its generalization into multilingual settings and further to other domains like e-commerce, education, and digital marketing validates its applicability beyond just one vertical.

The outcome of this work further underlines that hybrid architectures, which merge deep contextual models with traditional linguistic representations, form a sound basis for developing state-of-the-art Arabic recommendation systems. Improvement noted across semantic depth, system stability, and latency reduction suggests that such models are well-placed for large-scale applications that are user-centric in nature. Moreover, embedding mechanisms for multidimensional analytics and visualization enhances interpretability and leads to more transparent decision-making within recommendation pipelines. Even with these developments, aspects concerning data sparsity, dialectal variability, and computational cost remain open for exploration. Overcoming such limitations-especially with reinforcement learning and real-time personalization-will be important in the development of next-generation, scalable Arabic recommendation systems capable of cultural adaptability. Overall, this study demonstrates that deep learning models, combined with advanced semantic modeling and analytics that focus on the interpretability aspects, provide a reliable and effective means to improve Arabic recommender systems in terms of quality, relevance, and cultural sensitivity. With that, it provides a strong base for future development toward more robust, context-aware, and user-friendly recommendations.

REFERENCES

[1] A. A. Al-Ajlan et al., "Recommender system for Arabic content using sentiment analysis of user reviews," Electronics, 2023. https://doi.org/10.3390/electronics12132785.

[2] S. Al-Ghuribi et al., "A comparative study of sentiment-aware collaborative filtering algorithms for Arabic recommendation systems," IEEE Access, vol. 12, pp. 174441–174454, 2024. https://doi.org/10.1109/ACCESS.2024.3489658.

[3] C. Channarong et al., "HybridBERT4Rec: A hybrid (content-based filtering and collaborative filtering) recommender system based on BERT," IEEE Access, 2022. https://doi.org/10.1109/access.2022.3177610.

[4] R. Widayanti et al., "Improving recommender systems using hybrid techniques of collaborative filtering and content-based filtering," Journal of Applied Data Sciences, 2023. https://doi.org/10.47738/jads.v4i3.115.

[5] Y. Su et al., "Improvement of reading platforms assisted by the spring framework: A recommendation technique integrating the KGMRA algorithm and BERT model," Heliyon, vol. 11, 2025. https://doi.org/10.1016/j.heliyon.2025.e42191.

[6] N. Darraz et al., "Integrated sentiment analysis with BERT for enhanced hybrid recommendation systems," Expert Systems with Applications, vol. 261, p. 125533, 2024. https://doi.org/10.1016/j.eswa.2024.125533.

[7] P. D. Reddy et al., "Enhancing content based collaborative filtering recommendations using weighted word embeddings," in Proc. 15th Int. Conf. Computing Communication and Networking Technologies (ICCCNT), 2024, pp. 1–7. https://doi.org/10.1109/ICCCNT61001.2024.10724582.

[8] M. Srifi et al., "Evaluation of recent advances in recommender systems on Arabic content," Journal of Big Data, vol. 8, 2021. https://doi.org/10.1186/s40537-021-00420-2.

[9] M. Srifi et al., "A novel deep learning-based recommender system for Arabic content," Journal of Intelligent & Fuzzy Systems, vol. 43, pp. 437–449, 2022. https://doi.org/10.3233/jifs-212274.

[10] R. Morzelona et al., "Sentiment-aware content recommendation using LSTM-based collaborative filtering," Research Journal of Computer Systems and Engineering, 2022. https://doi.org/10.52710/rjcse.53.

[11] T. Wang et al., "Item-based collaborative filtering with BERT," in Proc. 3rd Workshop on e-Commerce and NLP, 2020. https://doi.org/10.18653/v1/2020.ecnlp-1.8.

[12] I. Karabila et al., "BERT-enhanced sentiment analysis for personalized e-commerce recommendations," Multimedia Tools and Applications, vol. 83, pp. 56463–56488, 2023. https://doi.org/10.1007/s11042-023-17689-5.

[13] P. Do et al., "Developing a BERT based triple classification model using knowledge graph embedding for question answering system," Applied Intelligence, vol. 52, pp. 636–651, 2021. https://doi.org/10.1007/s10489-021-02460-w.

[14] S. Sharma et al., "Automatic recommendation system based on hybrid filtering algorithm," Education and Information Technologies, vol. 27, pp. 1523–1538, 2021. https://doi.org/10.1007/s10639-021-10643-8.

[15] D. A. J. et al., "Advancing e-commerce recommendations: A comparative study of BERT-enhanced collaborative filtering and alternative approaches," in Proc. Int. Conf. Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA), 2024, pp. 1–6. https://doi.org/10.1109/ICAIQSA64000.2024.10882202.

[16] D. Suhartono et al., "Book recommendation using double-stack BERT: Utilizing BERT to extract sentence relation feature for a content-based filtering system," 2023, pp. 36–47. https://doi.org/10.1007/978-3-031-36402-0_3.

[17] Z. Qiu et al., "U-BERT: Pre-training user representations for improved recommendation," in Proc. AAAI, 2021, pp. 4320–4327. https://doi.org/10.1609/aaai.v35i5.16557.

[18] R. Rwedhi et al., "Improving collaborative filter using BERT," Journal of Kufa for Mathematics and Computer, 2023. https://doi.org/10.31642/jokmc/2018/100204.

[19] M. J. Althobaiti et al., "BERT-based approach to Arabic hate speech and offensive language detection in Twitter: Exploiting emojis and sentiment analysis," International Journal of Advanced Computer Science and Applications, 2022. https://doi.org/10.14569/ijacsa.2022.01305109.

[20] B. N. M. Hoang et al., "Using BERT embedding to improve memory-based collaborative filtering recommender systems," in Proc. RIVF Int. Conf. Computing and Communication Technologies, 2021, pp. 1–6. https://doi.org/10.1109/RIVF51545.2021.9642103.

[21] R. H. El-Deeb et al., "Enhancing e-recruitment recommendations through text summarization techniques," Information, 2025. https://doi.org/10.3390/info16040333.

[22] S. Patil et al., "Enhancing movie recommendation systems with BERT: A deep learning approach," in Proc. 5th Int. Conf. Emerging Technology (INCET), 2024, pp. 1–8. https://doi.org/10.1109/INCET61516.2024.10593284.

[23] R. A. Machhout et al., "Enhanced BERT approach to score Arabic essay's relevance to the prompt," Communications of the IBIMA, 2024. https://doi.org/10.5171/2024.176992.

[24] W. Nazih et al., "Arabic syntactic diacritics restoration using BERT models," Computational Intelligence and Neuroscience, 2022. https://doi.org/10.1155/2022/3214255.

[25] C. Channarong et al., "Hybrid (content-based filtering and collaborative filtering) recommender system based on BERT," 2019. https://doi.org/10.58837/chula.sp.2019.43.

[26] S. Saxena et al., "Transformers (BERT) based framework for web recommendations using sentiment-enriched web data," Journal of Information Systems Engineering and Management, 2025. https://doi.org/10.52783/jisem.v10i11s.1632.

[27] H. Chouikhi et al., "Arabic sentiment analysis using BERT model," 2021, pp. 621–632. https://doi.org/10.1007/978-3-030-88113-9_50.

[28] H. Lim et al., "A BERT-based multi-embedding fusion method using review text for recommendation," Expert Systems, vol. 42, 2025. https://doi.org/10.1111/exsy.70041.

[29] S. Khalifeh et al., "A book recommender system using collaborative filtering method," in Int. Conf. Data Science, E-learning and Information Systems, 2021. https://doi.org/10.1145/3460620.3460744.