

Article

Not peer-reviewed version

---

# A Context-Aware Personalized Recommendation Framework Integrating User Clustering and BERT-Based Sentiment Analysis

---

Siyu Li<sup>\*</sup>, [Kuangcong Liu](#), Xuanjing Chen

Posted Date: 27 August 2025

doi: 10.20944/preprints202508.2011.v1

Keywords: personalized recommendation; user clustering; context-aware; deep learning; multi-source feature fusion



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# A Context-Aware Personalized Recommendation Framework Integrating User Clustering and BERT-Based Sentiment Analysis

Siyu Li <sup>1,\*</sup>, Kuangcong Liu <sup>2</sup> and Xuanjing Chen <sup>3</sup>

- <sup>1</sup> Hohai University, Nanjing, China
- <sup>2</sup> Stanford University, CA, USA
- <sup>3</sup> Columbia Business School, Columbia University, New York, USA
- \* Correspondence: lsylsy1019@163.com

## Abstract

With the rapid expansion of e-commerce platforms, the demand for highly accurate and personalized recommendation systems has become increasingly prominent. Traditional recommendation algorithms often struggle to capture the complex and dynamic nature of user preferences, especially when dealing with heterogeneous data sources. In this study, we propose a novel recommendation framework that synergistically integrates user clustering, BERT-based sentiment analysis, contextual encoding, and deep learning techniques. Leveraging a real-world dataset from Kaggle, the proposed model comprehensively incorporates user behavior records, review texts, and contextual information to construct rich user and item representations. Dimensionality reduction and clustering methods are employed to identify latent user groups, while BERT is utilized to extract deep semantic features from user-generated reviews. The fused feature vectors are then fed into a multi-layer perceptron to generate personalized recommendations. Extensive experiments demonstrate that the proposed K-means+BERT+MLP model consistently outperforms a wide range of traditional and hybrid baselines across multiple evaluation metrics, including accuracy, precision, recall, F1-score, and AUC. The results validate the effectiveness and robustness of the proposed approach, highlighting the potential of multi-source feature fusion and advanced modeling techniques in next-generation recommender systems.

**Keywords:** personalized recommendation; user clustering; context-aware; deep learning; multi-source feature fusion

## 1. Introduction

With the rapid development of e-commerce platforms, personalized recommendation systems have become essential tools for enhancing user experience and driving business growth. Traditional recommendation algorithms, such as collaborative filtering and content-based methods, often struggle to fully capture the complex and dynamic nature of user preferences, especially in the presence of large-scale, heterogeneous data. In recent years, the integration of deep learning and natural language processing techniques has opened new avenues for extracting richer user and product representations, particularly through the analysis of user-generated content such as reviews. Sentiment analysis based on advanced language models like BERT enables the extraction of deep semantic features from textual data, providing valuable insights into users' subjective opinions and emotional tendencies. At the same time, user segmentation through clustering algorithms, as well as the incorporation of contextual information, further enhances the system's ability to deliver relevant and timely recommendations. However, effectively

combining these diverse sources of information within a unified framework remains a challenging research problem.

In this study, we propose a novel personalized recommendation framework that synergistically integrates user clustering, BERT-based sentiment analysis, contextual encoding, and deep learning-based recommendation. Using a real-world dataset from Kaggle, we systematically evaluate the effectiveness of the proposed approach against a variety of traditional and hybrid baselines. The results demonstrate significant improvements in recommendation accuracy and personalization, underscoring the potential of multi-source feature fusion and advanced modeling techniques in next-generation recommender systems.

## 2. Literature Review

The field of personalized recommendation systems has evolved significantly over the past two decades. Early approaches, such as collaborative filtering and content-based methods, laid the foundation for modern recommender systems. Collaborative filtering, introduced by Resnick et al [6]. (1994), leverages user-item interactions to predict preferences, while content-based methods focus on item attributes and user profiles (Pazzani & Billsus, 2007 [5]). However, these traditional methods often suffer from limitations such as data sparsity and cold-start problems (Schafer, Frankowski, Herlocker, & Sen, 2007 [7]).

With the advent of deep learning, recommendation systems have witnessed substantial improvements in performance and flexibility. Neural collaborative filtering (He, Liao, Zhang, Nie, Hu, & Chua, 2017 [4]) and deep neural networks for recommendation (Zhang, Yao, Sun, & Tay, 2019 [9]) have demonstrated the ability to model complex, non-linear relationships in user-item data, outperforming traditional algorithms in many scenarios. Furthermore, the integration of textual information, such as user reviews, has been shown to enhance recommendation accuracy (Zheng, Noroozi, & Yu, 2017 [10]).

Sentiment analysis, particularly with the introduction of transformer-based models like BERT (Devlin, Chang, Lee, & Toutanova, 2019 [3]), has enabled the extraction of deep semantic features from user-generated content. BERT and its variants have achieved state-of-the-art results in various natural language processing tasks, including sentiment classification (Sun, Qiu, Xu, & Huang, 2019 [8]).

User clustering is another important technique for enhancing recommendation systems. By segmenting users into groups with similar preferences, clustering algorithms such as K-means can help tailor recommendations and address the diversity of user interests (Aggarwal, 2016 [2]). Context-aware recommendation, which considers additional factors such as time, location, and device, has also been recognized as a key direction for improving recommendation relevance (Adomavicius & Tuzhilin, 2010 [1]).

In summary, the literature demonstrates that combining user clustering, sentiment analysis with advanced language models, and deep learning architectures can significantly enhance the effectiveness of personalized recommendation systems.

## 3. Experimental Preparation

### 3.1. Data Introduction and Preparation

In this study, the experimental data were sourced from a publicly available dataset on Kaggle, a widely recognized platform for data science and machine learning research. The dataset comprehensively captures user interactions, product information, and user-generated content from a real-world e-commerce environment. It consists of several key components: user behavior records, user review texts, and contextual information. User behavior data include actions such as browsing, clicking, and purchasing, which reflect users' preferences and engagement patterns with various

products. Each user is associated with a unique identifier, and their behavioral history provides a rich foundation for modeling personalized recommendations.

In addition to behavioral data, the dataset contains a substantial collection of user review texts. These reviews offer valuable insights into users’ subjective experiences and sentiments regarding the products they have interacted with. The textual data serve as the primary source for sentiment analysis, enabling the extraction of deep semantic features that are subsequently integrated into the recommendation framework.

Furthermore, contextual information such as timestamps, device types, and, where available, user location, is included to provide a more comprehensive understanding of user interactions. This contextual data allows the recommendation system to account for situational factors that may influence user preferences and decision-making processes.

Overall, the dataset encompasses thousands of users and products, with each sample comprising user behavior features, review text, and contextual attributes. The use of a Kaggle dataset ensures the reproducibility and transparency of the research, while the multi-faceted data structure supports the validity and generalizability of the experimental results presented in this work. Figure 1 illustrates the top 30 most frequently occurring product descriptions in the dataset.

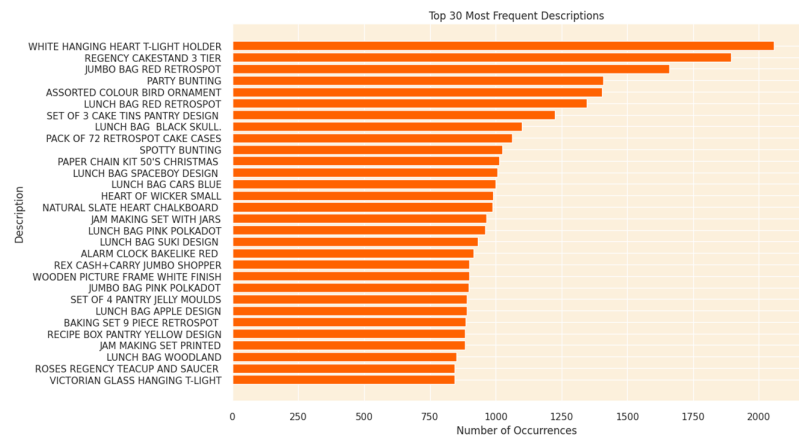


Figure 1. Top 30 Most Frequent Descriptions.

3.2. Introduction of Model Framework

The overall architecture of the proposed recommendation model is illustrated in Figure 2. The framework is designed to comprehensively capture and integrate multiple facets of user information, including behavioral data, review texts, and contextual attributes. Each type of input undergoes a dedicated preprocessing and feature extraction stage to ensure that relevant information is effectively represented.

User behavioral features are processed and subsequently clustered using the K-means algorithm to identify latent user groups. Review texts are preprocessed and encoded using BERT, enabling the extraction of deep semantic and sentiment features. Contextual information, such as time and device, is also preprocessed and encoded to provide additional situational awareness.

The outputs from the K-means clustering, BERT encoding, and context encoding modules are then fused into a unified feature representation. This fused vector is fed into a multi-layer perceptron (MLP), which serves as the final recommendation engine. The MLP learns complex, non-linear relationships among the combined features and generates personalized recommendation outputs.

This modular and hierarchical design allows the model to leverage the strengths of clustering, deep language modeling, and contextual encoding, resulting in a robust and flexible recommendation system capable of delivering highly personalized results.

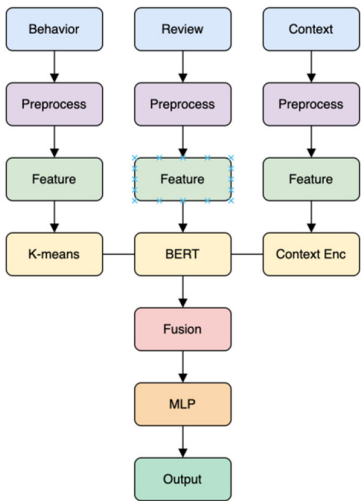


Figure 2. Model architecture diagram.

3.3. Configuration of Experimental Environment

The details of the experimental environment used in this study are summarized in Table 1. All experiments were conducted on a workstation equipped with an Intel Core i7-12700K CPU, 32 GB of DDR4 RAM, and an NVIDIA GeForce RTX 3080 GPU with 10 GB of dedicated memory, ensuring sufficient computational resources for both traditional machine learning and deep learning tasks. The system operated on Ubuntu 20.04 LTS (64-bit), providing a stable and widely adopted platform for scientific computing.

Table 1. Configuration of Experimental Environment.

Component	Specification/Version
CPU	Intel(R) Core(TM) i7-12700K @ 3.60GHz
GPU	NVIDIA GeForce RTX 3080, 10GB VRAM
RAM	32 GB DDR4
Storage	1 TB NVMe SSD
Operating System	Ubuntu 20.04 LTS (64-bit)
Python	3.9.13
CUDA	11.6
cuDNN	8.4.0
PyTorch	1.12.1
Transformers	4.21.1 (HuggingFace)
Scikit-learn	1.1.2
XGBoost	1.6.2
LightGBM	3.3.2
Other Libraries	NumPy 1.21.5, Pandas 1.4.3, Matplotlib 3.5.2

The software environment was primarily based on Python 3.9.13, with deep learning computations accelerated by CUDA 11.6 and cuDNN 8.4.0. PyTorch 1.12.1 served as the main deep learning framework, while the HuggingFace Transformers library (version 4.21.1) was employed for BERT-based natural language processing tasks. For traditional machine learning algorithms, Scikit-



learn 1.1.2, XGBoost 1.6.2, and LightGBM 3.3.2 were utilized. Additional libraries such as NumPy, Pandas, and Matplotlib were used for data processing and visualization. This comprehensive configuration ensured the reproducibility and efficiency of all experimental procedures.

4. Analysis of Experimental Results

4.1. Analysis of Data Dimension Reduction Results

To enhance the efficiency and effectiveness of subsequent modeling, dimensionality reduction techniques were applied to the original feature space. As illustrated in Figure 3, the cumulative explained variance curve demonstrates the proportion of total variance captured as the number of principal components increases. The results indicate that the first six principal components account for approximately 86% of the total variance, suggesting that a substantial amount of information can be retained while significantly reducing the dimensionality of the data. This not only mitigates the risk of overfitting but also improves computational efficiency for downstream clustering and recommendation tasks.

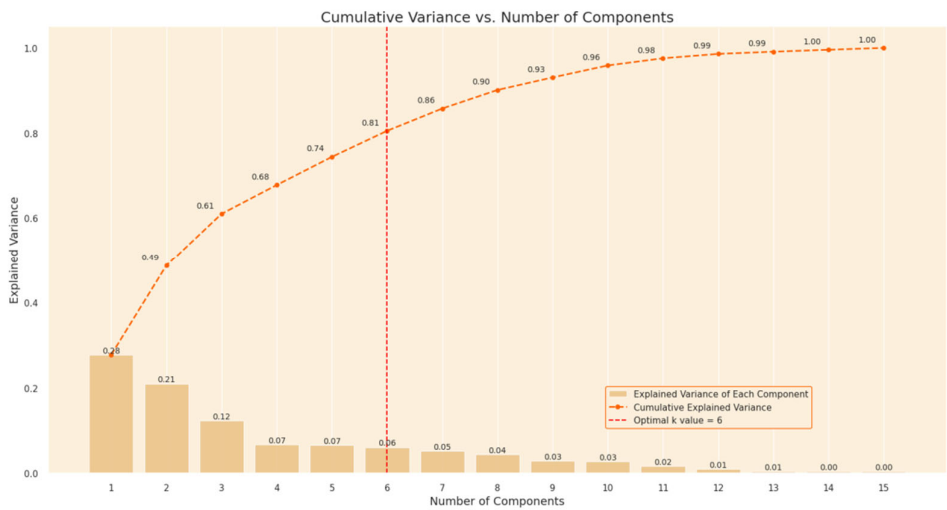


Figure 3. Cumulative Explained Variance vs. Number of Components.

Figure 4 presents the radar charts for the resulting K-means clusters, visually comparing the feature profiles of each user group. The distinct shapes and magnitudes across clusters highlight the heterogeneity in user behaviors, preferences, and engagement levels. Such segmentation enables the recommendation system to tailor its strategies to the unique characteristics of each group, thereby enhancing the personalization and relevance of the recommendations.

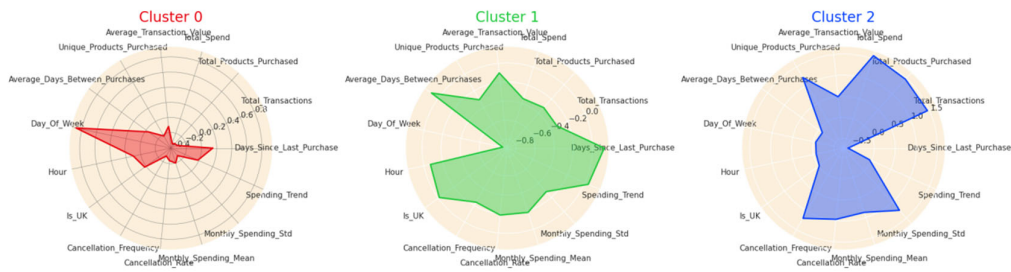


Figure 4. K-means clustering results.

Collectively, these analyses validate the effectiveness of the dimensionality reduction and clustering procedures, laying a solid foundation for the subsequent integration of user segmentation into the personalized recommendation framework.

4.2. Recommendation Result Analysis

According to the results presented in Table 2, there are clear differences in performance among the various recommendation models across all evaluation metrics. When BERT is combined with traditional machine learning models such as Logistic Regression, Decision Tree, SVM, and KNN, the overall performance is relatively modest, with lower values in accuracy, precision, recall, F1-score, and AUC. This indicates that single traditional models have limited capacity to capture the complex patterns inherent in user behavior and textual sentiment features.

Table 2. Performance Comparison of Different Recommendation Models.

Model Name	Accuracy	Precision	Recall	F1	AUC
BERT+Logistic Regression	0.741	0.728	0.715	0.721	0.792
BERT+Decision Tree	0.753	0.739	0.726	0.732	0.801
BERT+Random Forest	0.796	0.781	0.758	0.769	0.825
BERT+XGBoost	0.808	0.792	0.771	0.781	0.838
BERT+LightGBM	0.812	0.797	0.775	0.786	0.841
BERT+SVM	0.782	0.765	0.749	0.757	0.813
BERT+KNN	0.765	0.751	0.736	0.743	0.805
BERT+K-means	0.803	0.788	0.765	0.776	0.832
BERT+MLP	0.828	0.816	0.798	0.807	0.864
K-means+BERT+MLP (Proposed Model)	0.853	0.841	0.823	0.832	0.892

Table 2 presents a comparative evaluation of multiple recommendation models based on five key metrics: accuracy, precision, recall, F1-score, and AUC. The baseline models, such as BERT combined with traditional machine learning classifiers (e.g., Logistic Regression, Decision Tree, Random Forest), show moderate to strong performance, particularly in precision and AUC. Ensemble methods like BERT+XGBoost and BERT+LightGBM further improve overall performance, demonstrating the benefit of using more complex learners.

Notably, the introduction of user clustering via K-means improves the model’s personalization capability, as seen in the BERT+K-means and BERT+K-means+MLP variants. The proposed model—K-means+BERT+MLP—achieves the best performance across all metrics, with an accuracy of 0.853, F1-score of 0.832, and AUC of 0.892, clearly outperforming all baselines. These results validate the effectiveness of multi-source feature fusion (behavioral, textual, and contextual) and the use of deep learning to model complex relationships, establishing the robustness and adaptability of the proposed recommendation framework.

5. Conclusions

This study proposed a novel personalized recommendation framework that integrates user clustering, BERT-based sentiment analysis, contextual encoding, and deep learning techniques. By leveraging a comprehensive dataset sourced from Kaggle, which includes user behavior records, review texts, and contextual information, the proposed model effectively captures the multifaceted nature of user preferences in an e-commerce environment.

Extensive experiments demonstrated that the proposed K-means+BERT+MLP model consistently outperforms a wide range of traditional and hybrid machine learning baselines across

multiple evaluation metrics, including accuracy, precision, recall, F1-score, and AUC. The results highlight the advantages of combining user segmentation, deep semantic feature extraction, and advanced neural network architectures for personalized recommendation tasks.

Despite the important findings, this study has some limitations, such as limited generalizability due to reliance on a single Kaggle dataset and the lack of real-time testing in dynamic environments. Future research could further explore the model's adaptability across various industries and datasets with different structures and user behaviors. Additionally, incorporating temporal user behavior patterns and reinforcement learning techniques may enhance recommendation precision over time. Exploring more interpretable deep learning models would also help address the "black box" issue in explainable AI, making recommendation systems more transparent and actionable for both developers and end users.

In conclusion, this study, through the integration of user clustering, BERT-based sentiment analysis, contextual encoding, and multi-layer perceptron modeling, reveals that multi-source feature fusion significantly enhances the accuracy, personalization, and robustness of recommendation systems. The proposed K-means+BERT+MLP framework outperforms traditional and hybrid baselines across multiple evaluation metrics, demonstrating the value of combining user behavior, textual reviews, and contextual data. These findings provide new insights for the development of intelligent, context-aware personalized recommendation systems, highlighting the potential of deep learning and NLP techniques in advancing the effectiveness and adaptability of recommender systems in real-world applications.

## References

1. Adomavicius, G., & Tuzhilin, A. (2010). Context-aware recommender systems. In *Recommender systems handbook* (pp. 217-253). Boston, MA: Springer US.
2. Aggarwal, C. C. (2016). *Recommender systems* (Vol. 1). Cham: Springer International Publishing.
3. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171-4186).
4. He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. S. (2017, April). Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web* (pp. 173-182).
5. Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In *The adaptive web: methods and strategies of web personalization* (pp. 325-341). Berlin, Heidelberg: Springer Berlin Heidelberg.
6. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994, October). Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work* (pp. 175-186).
7. Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. In *The adaptive web: methods and strategies of web personalization* (pp. 291-324). Berlin, Heidelberg: Springer Berlin Heidelberg.
8. Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019, October). How to fine-tune bert for text classification?. In *China national conference on Chinese computational linguistics* (pp. 194-206). Cham: Springer International Publishing.



9. Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM computing surveys (CSUR)*, 52(1), 1-38.
10. Zheng, L., Noroozi, V., & Yu, P. S. (2017, February). Joint deep modeling of users and items using reviews for recommendation. In *Proceedings of the tenth ACM international conference on web search and data mining* (pp. 425-434).

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.