GEORGE BROWN PRESS
JOURNALS, BOOKS & PROCEEDINGS

Journal of Computer, Signal, and System Research

Article

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

Siyu Li [1,*], Kuangcong Liu [2] and Xuanjing Chen [3]

[1] Hohai University, Nanjing, China
[2] Stanford University, CA, USA
[3] Columbia Business School, Columbia University, New York, USA
* Correspondence: Siyu Li, Hohai University, Nanjing, China

Abstract: With the rapid growth of e-commerce platforms, there is an increasing demand for highly accurate and personalized recommendation systems. Traditional recommendation algorithms often struggle to capture the complex and dynamic nature of user preferences, especially when dealing with heterogeneous data sources. This study introduces a novel recommendation framework that integrates user clustering, BERT-based sentiment analysis, contextual encoding, and deep learning techniques. Utilizing a real-world dataset from Kaggle, the proposed model incorporates user behavior records, review texts, and contextual information to construct detailed user and item representations. Dimensionality reduction and clustering methods are employed to identify latent user groups, while BERT is used to extract deep semantic features from user-generated reviews. The resulting feature vectors are then fed into a multi-layer perceptron to generate personalized recommendations. Extensive experiments demonstrate that the K-means + BERT + MLP model consistently outperforms a variety 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, showcasing 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

Received: 13 October 2025
Revised: 28 October 2025
Accepted: 16 November 2025
Published: 22 November 2025

1. Introduction

With the rapid advancement of e-commerce platforms, personalized recommendation systems have become indispensable tools for enhancing user experience and driving business growth. These systems have transformed the way consumers interact with online services, enabling businesses to offer tailored product suggestions that align with individual preferences and behaviors. However, traditional recommendation algorithms, such as collaborative filtering and content-based methods, often face significant limitations in capturing the complex and dynamic nature of user preferences. This is particularly evident when dealing with large-scale, heterogeneous data that varies in both structure and source, making it difficult for these methods to provide truly personalized recommendations.

In recent years, the integration of deep learning and natural language processing (NLP) techniques has opened new possibilities for overcoming these challenges. One such advancement is sentiment analysis, which leverages advanced language models like BERT (Bidirectional Encoder Representations from Transformers) to extract deep semantic features from textual data. By analyzing user-generated content, such as product reviews, sentiment analysis provides valuable insights into users' subjective opinions,

emotional tendencies, and preferences. This capability allows recommendation systems to better understand the nuanced sentiments expressed by users, leading to more relevant and accurate recommendations [1].

In parallel, user segmentation through clustering algorithms enhances the ability to categorize users based on shared characteristics or behavior patterns. By grouping similar users together, systems can offer personalized recommendations that reflect the unique preferences of each segment. Additionally, the incorporation of contextual information-such as time of day, location, or browsing history-further refines the recommendation process, ensuring that suggestions are not only relevant but also timely and contextually appropriate.

Despite the promising potential of these techniques, effectively combining multiple diverse data sources-such as user behavior, review texts, and contextual information-into a single unified recommendation framework remains a challenging research problem. Current methods often struggle to integrate these various forms of information in a way that captures their full complexity, leading to suboptimal recommendation performance.

In this study, we propose a novel personalized recommendation framework that seamlessly integrates user clustering, BERT-based sentiment analysis, contextual encoding, and deep learning-based recommendation techniques. By leveraging 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, precision, and personalization, underscoring the potential of multi-source feature fusion and advanced modeling techniques in the development of next-generation recommender systems. Our findings highlight the promise of integrating cutting-edge NLP and deep learning methods to enhance the accuracy and relevance of recommendations in increasingly complex and data-rich environments [2].

2. Literature Review

The field of personalized recommendation systems has seen significant advancements over the past two decades. Early approaches, including collaborative filtering and content-based methods, laid the groundwork for modern recommender systems. Collaborative filtering, introduced in 1994, uses historical user-item interactions to predict user preferences. This method relies on the idea that users who interacted with similar items in the past are likely to exhibit similar preferences in the future. In contrast, content-based methods, which focus on analyzing the attributes of items and user profiles, recommend items based on the characteristics of products or services. While these traditional techniques provided the foundation for recommendation systems, they often suffer from key limitations, such as data sparsity and the cold-start problem, where systems struggle to make accurate predictions for new users or items with little historical data [3].

With the rise of deep learning, recommendation systems have seen notable improvements in both performance and flexibility. Models such as neural collaborative filtering and deep neural networks for recommendation have demonstrated the ability to model complex, non-linear relationships in user-item interactions, leading to superior performance compared to traditional algorithms in many scenarios. Furthermore, the integration of textual information, such as user reviews, has proven to enhance recommendation accuracy. By leveraging natural language processing (NLP) techniques, systems can better understand the underlying sentiments and opinions expressed in user-generated content, providing more personalized and accurate suggestions [4].

Sentiment analysis, particularly with the advent of transformer-based models such as BERT, has enabled the extraction of deep semantic features from textual data. BERT and its variants have set new benchmarks in various NLP tasks, including sentiment classification, allowing recommendation systems to gain a deeper understanding of users'

emotional tendencies and subjective opinions. This ability to capture subtle emotional nuances has proven to be particularly valuable for refining recommendation relevance.

User clustering is another important technique that has been employed to improve recommendation systems. By grouping users with similar preferences, clustering algorithms such as K-means can help tailor recommendations to the needs of specific user segments. This technique helps address the diversity of user interests, enabling systems to provide more relevant suggestions based on shared behaviors or preferences [5].

Context-aware recommendation, which takes into account additional factors such as time, location, and device, has also been recognized as a key direction for improving recommendation relevance. By incorporating contextual information, systems can offer more timely and contextually appropriate recommendations, further enhancing the user experience.

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. These methods not only improve recommendation accuracy but also ensure a more personalized and relevant user experience by incorporating richer, multi-source data [6].

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 [7].

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 [8].

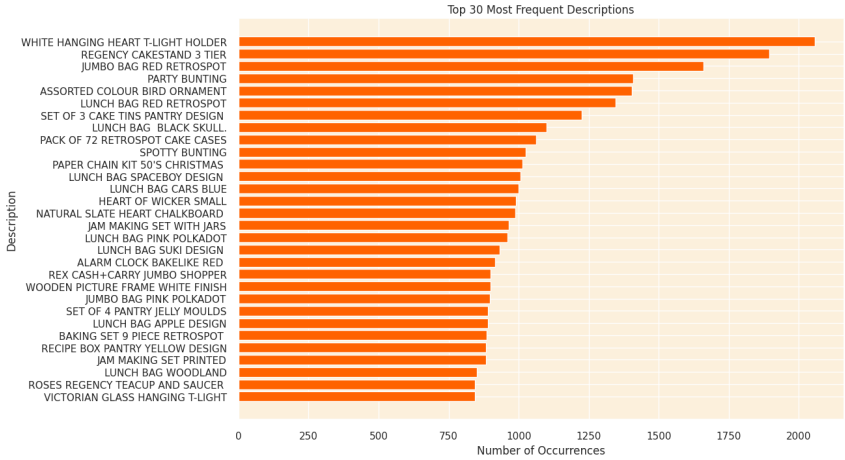

Figure 1. Top 30 Most Frequent Descriptions.

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 [9].

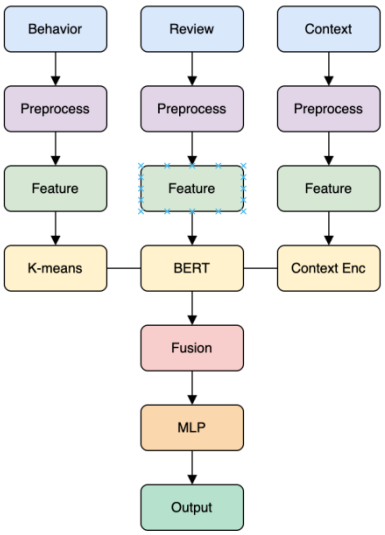

Figure 2. Model architecture diagram.

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 [10].

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.

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.

3.4. Introduction of Evaluation Index

Accuracy is a fundamental metric that measures the proportion of correctly predicted instances among the total number of samples. It provides an overall assessment of the model's predictive capability, but may be less informative in the presence of class imbalance.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (1)$$

Precision evaluates the proportion of true positive predictions among all positive predictions made by the model. It reflects the model's ability to avoid false positives and is particularly important in scenarios where the cost of false alarms is high.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

Recall, also known as sensitivity or true positive rate, measures the proportion of actual positive cases that are correctly identified by the model. It is crucial in applications where missing positive cases (false negatives) is costly.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

The F1-score is the harmonic mean of precision and recall, providing a single metric that balances both concerns. It is especially useful when the class distribution is uneven or when both false positives and false negatives are important.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

AUC represents the area under the Receiver Operating Characteristic (ROC) curve, which plots the true positive rate against the false positive rate at various threshold settings. AUC provides a comprehensive measure of a model's ability to distinguish between classes, with higher values indicating better discriminative performance. The AUC value ranges from 0 to 1, where 1 denotes perfect classification and 0.5 indicates random guessing.

$$\text{AUC} = \int_0^1 \text{T PR(FPR) dFPR} \tag{5}$$

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.

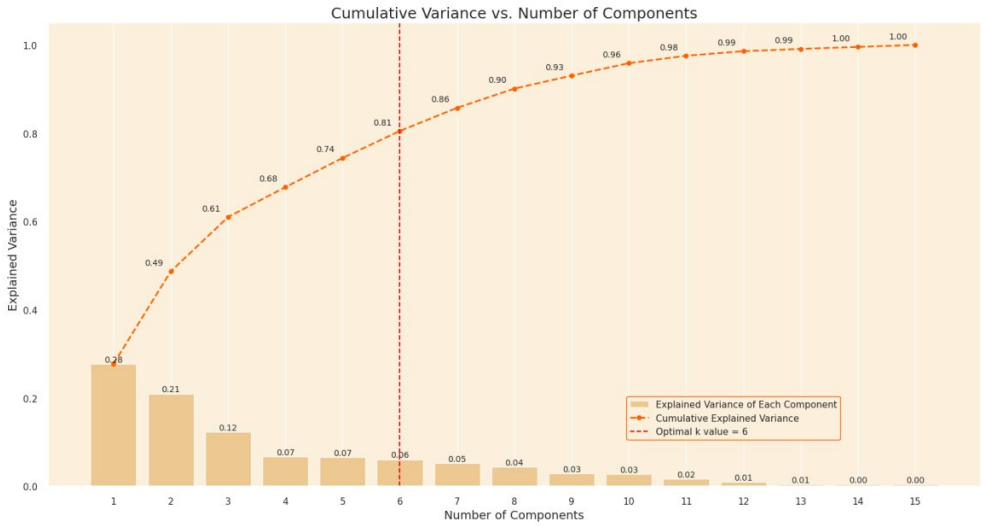

Figure 3. Cumulative Explained Variance vs. Number of Components.

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.

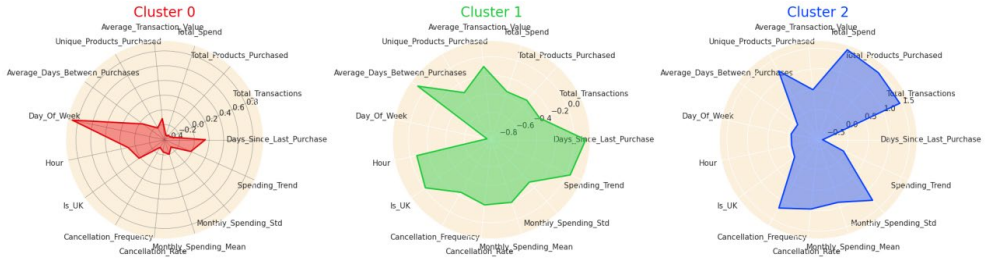

Figure 4. K-means clustering results.

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. Conclusion

This study aims to address the limitations of traditional recommendation systems in capturing dynamic and heterogeneous user preferences by integrating advanced machine learning and natural language processing techniques. It does so by leveraging BERT-based sentiment analysis, K-means clustering, contextual encoding, and multi-layer perceptrons to construct a comprehensive recommendation framework. The research explores how multi-source feature fusion can enhance personalization accuracy and robustness. The primary objective of this research is to develop and validate a context-aware recommendation model that effectively combines behavioral, textual, and contextual signals to improve the precision and quality of personalized e-commerce recommendations.

Through data analysis, we identified that (1) integrating BERT-based sentiment features significantly improves the semantic understanding of user preferences, (2) clustering users via K-means enhances the model's ability to capture group-level behavioral patterns, and (3) combining multi-source features through a multi-layer perceptron (MLP) yields the highest performance across all evaluation metrics. These findings suggest that a hybrid approach leveraging textual, behavioral, and contextual data-when fused in a deep learning architecture-can substantially outperform traditional and single-source recommendation methods, highlighting the importance of holistic user representation in personalized recommender systems.

The results of this study have significant implications for the field of personalized recommendation systems. Firstly, the integration of K-means clustering, BERT-based sentiment analysis, and contextual encoding provides a new perspective on multi-source feature fusion, demonstrating improved performance across key metrics. Secondly, the superior results of the proposed model challenge the adequacy of traditional or isolated modeling approaches, suggesting that hybrid models can better capture complex user behaviors. Finally, the demonstrated effectiveness of the K-means+BERT+MLP framework opens new avenues for future research into scalable and interpretable recommendation systems that adapt to evolving user preferences.

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. G. Adomavicius, and A. Tuzhilin, "Context-aware recommender systems," In *Recommender systems handbook*, 2010, pp. 217-253. doi: 10.1007/978-0-387-85820-3_7

2. L. Lü, M. Medo, C. H. Yeung, Y. C. Zhang, Z. K. Zhang, and T. Zhou, "Recommender systems," *Physics reports*, vol. 519, no. 1, pp. 1-49, 2012.
3. J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "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)*, June, 2019, pp. 4171-4186.
4. X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T. S. Chua, "Neural collaborative filtering," In *Proceedings of the 26th international conference on world wide web*, April, 2017, pp. 173-182.
5. M. J. Pazzani, and D. Billsus, "Content-based recommendation systems," In *The adaptive web: methods and strategies of web personalization*, 2007, pp. 325-341. doi: 10.1007/978-3-540-72079-9_10
6. P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "Grouplens: An open architecture for collaborative filtering of netnews," In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, October, 1994, pp. 175-186.
7. J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative filtering recommender systems," In *The adaptive web: methods and strategies of web personalization*, 2007, pp. 291-324. doi: 10.1007/978-3-540-72079-9_9
8. C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to fine-tune bert for text classification?," In *China national conference on Chinese computational linguistics*, October, 2019, pp. 194-206. doi: 10.1007/978-3-030-32381-3_16
9. S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *ACM computing surveys (CSUR)*, vol. 52, no. 1, pp. 1-38, 2019.
10. L. Zheng, V. Noroozi, and P. S. Yu, "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*, February, 2017, pp. 425-434. doi: 10.1145/3018661.3018665