Engineering and Technology Journal e-ISSN: 2456-3358

Volume 10 Issue 07 July-2025, Page No.-5998-6004

DOI: 10.47191/etj/v10i07.46, I.F. – 8.482

Enhancing Product Recommendations with Bert: A Hybrid Approach Integrating Collaborative Filtering and NlP

Abderrahman Abid

Institute of Graduate Studies, Department of Artificial Intelligence and Data Science

Istanbul Aydin University – Istanbul

ABSTRACT: In the contemporary world of online retailing, consumers are easily perplexed by the ocean of products available to them, thus calling for sophisticated personalized recommendation systems to accompany user experience and sales uplift. The current work describes a hybrid product recommendation strategy by integrating classical collaborative filtering techniques with state-of-the-art innovations in Natural Language Processing (NLP) using particularly the pre-trained BERT architecture. The work begins by addressing the issue of data sparsity in large datasets, efficiently mitigating it by filtering out inactive users, which significantly increases matrix density. Classical collaborative filtering techniques, i.e., KNNWithMeans and Singular Value Decomposition (SVD), were utilized and tested, for which the user-based KNN model exhibited the smallest Root Mean Square Error (RMSE). To further augment recommendation quality, the BERT architecture was fine-tuned on product title-product description pairs laboriously trained for predicting product-item similarity. Experimental results show that the BERT-augmented system consistently outperforms traditional models, achieving 92% accuracy in similarity classification and generating more semantically relevant recommendations. The research demonstrates the efficacy of complementarily integrating deep learning and NLP capabilities with traditional recommendation algorithms by formulating both precise and contextual-product-aware recommendations. These findings affirm the value of incorporating semantic similarity models into real-world recommendation engines.

KEYWORDS: Recommender System, Collaborative Filtering, BERT, KNNWithMeans, NLP, Deep Learning, Hybrid Model, Amazon Dataset

INTRODUCTION

The advent of digital transformation has propelled e-commerce platforms to unprecedented popularity, offering users an unparalleled array of product choices accessible anytime and anywhere. However, this very abundance frequently precipitates a phenomenon termed "information overload," wherein users encounter significant challenges in efficiently identifying products that align with their specific preferences and needs. To counteract this pervasive issue, recommendation systems have emerged as indispensable tools, crucial for personalizing user experiences and substantially enhancing customer satisfaction.

Traditional recommendation approaches, such as Collaborative Filtering (CF) and Content-Based Filtering (CBF), are widely employed on different commercial systems. Even though they have demonstrated great potential, they are constrained by limitations, i.e., the cold-start problem and data insufficiency. These shortcomings are more critical where interaction data between items and users is limited, so the system is unable to recommend specific and varied items, especially for new items or new users.

Recent advances in Deep Learning (DL) and Natural Language Processing (NLP) have unveiled revolutionary potential for enhancing recommendation systems' abilities. NLP methods, more notably, empower systems to reap great insights from unstructured text data like product details and user reviews that abound with fine-grained semantic details oftentimes ignored by traditional rating-based methods. Perhaps one of the most powerful tools within this category is BERT (Bidirectional Encoder Representations from Transformers), a pre-trained language model known for its ability to understand context in an extremely deep and bidirectional way.

This work introduces a hybrid recommendation system which combines collaborative filtering techniques with a BERT model in a strategically competent way. Such combination facilitates effective textual data analysis as well as precise prediction of item similarities. A real-world data set was employed for the study borrowed from the Amazon "Arts, Crafts, and Sewing" category. Methodology comprised a series of vital stages such as intensive preprocessing, selective user filtering for data sparsity alleviating, careful construction of item pairs, and extensive model training. Experimental results undoubtedly demonstrate that using BERT significantly advances recommendations' quality as well as semantic meaningfulness such that it performs even better than traditional methods in respect to accuracy as well as robustness.

I. RELATED WORK

Recommendation systems have long served as a foundational element for delivering personalized user experiences across a multitude of platforms, including e-commerce, media streaming, and social networks. Over time, a diverse array of approaches has been developed with the singular aim of improving the relevance and accuracy of these recommendations.

A.Traditional Approaches

The cornerstone of conventional recommendation systems consists of two basic methods: Collaborative Filtering (CF) and Content-Based Filtering (CBF).

• Content-Based Filtering (CBF) works by using the characteristics of items on which an individual has interacted to provide recommendations for items with similar characteristics. Though successful in some situations, however, CBF tends to suffer from over-specialization, which may limit recommendation diversity.
• Collaborative Filtering (CF), on the other hand, builds recommendations by examining user-item interaction matrices, recommending items according to the patterns of similar users. CF includes both memory-based approaches like user-based and item-based K-Nearest Neighbors as well as model-based approaches like matrix factorization by Singular Value Decomposition. Nevertheless, CF has a major weakness in being sensitive to data sparsity as well as the cold-start issue for new users or items with scarce interaction histories.

A. Hybrid Systems

To mitigate these inherent drawbacks, hybrid recommendation systems have been developed. These systems combine CF and CBF techniques—or integrate additional contextual information—to yield more robust and diversified recommendations. For instance, HybridBERT4Rec (Channarong et al., 2022) effectively fuses BERT-based content representations with collaborative filtering to improve sequential recommendations. Studies have consistently shown that hybrid methods outperform standalone models in terms of both diversity and accuracy (Li et al., 2023; Zhang et al., 2024).

B. NLP in Recommendation Systems

The incorporation of Natural Language Processing (NLP) techniques has significantly enhanced recommender systems, enabling the extraction of deeper semantic signals from unstructured data such as product descriptions and user reviews. Early NLP-based approaches utilized methods like TF-IDF, Word2Vec, and Doc2Vec to derive semantic similarity. Although valuable, these models offered limited contextual understanding. Transformer-based models have since become dominant due to their ability to encode nuanced language patterns and long-range dependencies (Lin et al., 2022).

C. Deep Learning and Transformers

Among transformer-based models, BERT (Bidirectional Encoder Representations from Transformers) stands out for its bidirectional context modeling, which enables a finer-grained understanding of sentence-level semantics. Several studies have explored BERT's effectiveness in recommendation scenarios. For example, GFormer (Li et al., 2023) combined graph-based reasoning with BERT to improve explainability and relevance. Similarly, Rwedhi & Al-augby (2023) showed that integrating BERT-generated similarity scores into CF pipelines significantly improved performance, particularly in sparse and cold-start contexts.
Moreover, recent research has introduced sentiment-aware hybrid systems that use BERT to capture emotional cues in reviews, thereby improving personalization (Darraz et al., 2024). A 2024 survey further emphasizes the increasing dominance of deep learning in recommendation research, particularly for tasks such as Top-N ranking and representation learning (Survey on Deep Learning in Recommender Systems, 2024).

While prior studies have highlighted the potential of integrating deep semantic models with traditional recommenders, they often treat semantic modeling and CF as separate components. In contrast, our work presents a tightly integrated hybrid system where BERT's semantic similarity scores directly enhance the collaborative filtering phase—offering a unified approach that addresses sparsity, improves semantic relevance, and boosts personalization accuracy in real-world settings.

II. METHODOLOGY

This study proposes a hybrid recommendation system that synergistically integrates Collaborative Filtering (CF) with a pre-trained BERT model to significantly enhance product recommendation accuracy. The methodology systematically incorporates both traditional CF algorithms for rating prediction and advanced semantic similarity estimation, leveraging deep contextual text representations derived from BERT.

A. Overview of System Architecture

The Figure 1 below illustrates the general architecture of our recommendation system:

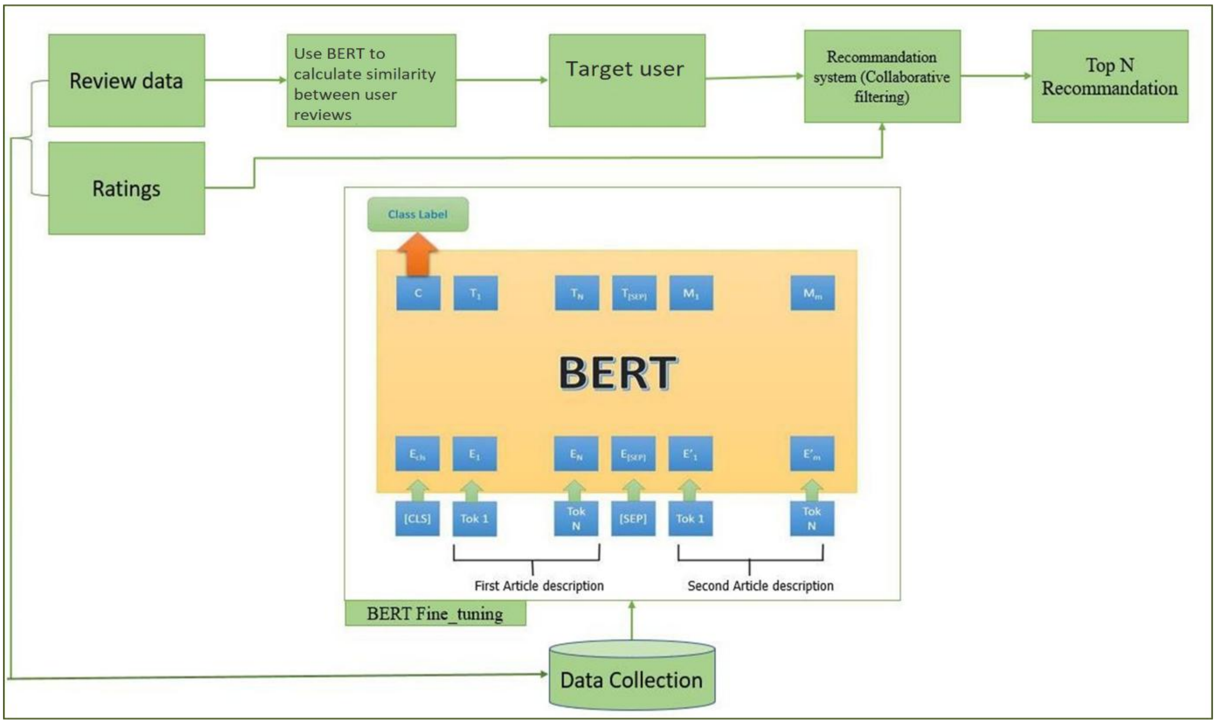

Figure 1: General Architecture of Our Recommendation System

This design meticulously integrates both user-item rating data and rich textual product reviews to deliver highly personalized Top-N product recommendations. The core innovative aspect of this system resides in its utilization of a fine-tuned BERT model to precisely determine the similarity between items, drawing upon comprehensive user reviews and product metadata.

The system adheres to the following sequence of key steps:

1. Data Collection and Preparation: Data is meticulously collected from the Amazon "Arts, Crafts, and Sewing" dataset, which encompasses crucial elements such as product titles, detailed descriptions, user ratings, and textual reviews. To effectively address the challenge of data sparsity, users with fewer than 50 product ratings are systematically removed, a critical step that robustly increased the matrix density from an initial 0.00057% to a significantly improved 0.32%. Review texts undergo a rigorous preprocessing pipeline, which includes cleaning, lowercasing, stopword removal, lemmatization, and precise special character filtering.

2. BERT-Based Similarity Modeling: A **pre-trained BERT model** is then meticulously fine-tuned on carefully constructed pairs of product data to effectively learn semantic similarity. Each input to the model consists of a pair of product descriptions or reviews, and the model is specifically trained to classify whether an item pair is similar (assigned a label of 1) or not similar (assigned a label of 0). These pairs are strategically formed utilizing co-purchase and co-view metadata, specifically by combining the "also_buy" and "also_view" fields to generate a comprehensive "similar_articles" column.

3. User Review Matching: Once the BERT model has been thoroughly fine-tuned, it is subsequently employed to compare reviews authored by different users. This sophisticated capability enables the system to identify users who share semantically similar preferences, even if their rated products differ.

4. Collaborative Filtering Component: Leveraging the similarity scores generated by the fine-tuned BERT model, a Collaborative Filtering system, implemented through the Surprise Python library, is then deployed. Specifically, the KNNWithMeans algorithm is applied using Pearson correlation to recommend items to a target user, based on the preferences of identified similar users. For comparative analysis within matrix factorization, Singular Value Decomposition (SVD) is also implemented.

5. Model Evaluation: The dataset is systematically partitioned into a 70% training set and a 30% test set to ensure robust evaluation. Evaluation metrics include Root Mean Square Error (RMSE) for assessing rating prediction accuracy, and a suite of classification metrics—Precision, Recall, and F1-score—for the BERT similarity model.

B. Tools and Libraries

The development and evaluation of this system were facilitated by a comprehensive suite of tools and libraries :

• *Python* programming language
• *Pandas* and *NumPy* for efficient data manipulation

- *Scikit-learn* for essential preprocessing and evaluation tasks
- *Surprise* library for the implementation of Collaborative Filtering algorithms
- *Hugging Face* Transformers and *PyTorch* for the fine-tuning of the BERT model
- *Google Colab* for leveraging GPU-accelerated training

III. EXPERIMENTS AND RESULTS

To comprehensively evaluate the performance of our hybrid recommendation system, a series of experiments were conducted, directly comparing traditional collaborative filtering methods with our innovative BERT-enhanced approach. These experiments were meticulously designed to assess both rating prediction accuracy and semantic similarity classification, specifically utilizing review data.

C. Dataset and Experimental Setup

The experiments were executed using the Amazon "Arts, Crafts, and Sewing" product dataset, which encompasses a wealth of information, including user ratings, product titles, descriptions, and detailed textual reviews. The initial dataset was substantial, comprising over 1.6 million users and thousands of distinct products.

A critical step in preparing the data involved addressing sparsity issues. After rigorously removing users who had provided fewer than 50 product ratings, the dataset was significantly refined. This filtering process resulted in a more manageable dataset consisting of 484 active users and 24,132 unique products, with a total of 37,644 user-item interactions. Crucially, this intervention dramatically improved the final density of the interaction matrix, increasing it from an initial 0.00057% to 0.32%, which is vital for enhancing collaborative filtering performance. The dataset was subsequently partitioned into a 70% training set and a 30% testing set, ensuring a representative distribution of users and ratings for robust model training and evaluation.

	ratings	userId	productId	title
0	5	A3DLXP7VQZSDVE	B00B6HZA10	cotton print quilt wilmington communiqu damask...
1	5	AMQFPVH5PY48B	B015BI1G1W	case lg tpu case ectouch fashion style color p...
2	4	A3OIPQ0N6E2JWT	B00HMQ8HB4	pieceslot new clipper sew trim scissor nipper ...
3	4	A109FU5GA5MO0B	B004TXN8AI	embroideri machin hoop set placement grid brot...
4	5	AOMNKO65GX3A4	B01ARMLLB2	eboot piec detail paint brush set miniatur bru...

Figure 2: The Data Used for Our Recommendation System

B. Evaluation Metrics

For the assessment of model performance, distinct evaluation metrics were employed based on the specific component of the system:

- **Collaborative Filtering: Root Mean Square Error (RMSE)** was utilized to quantify the prediction error between actual and predicted ratings.

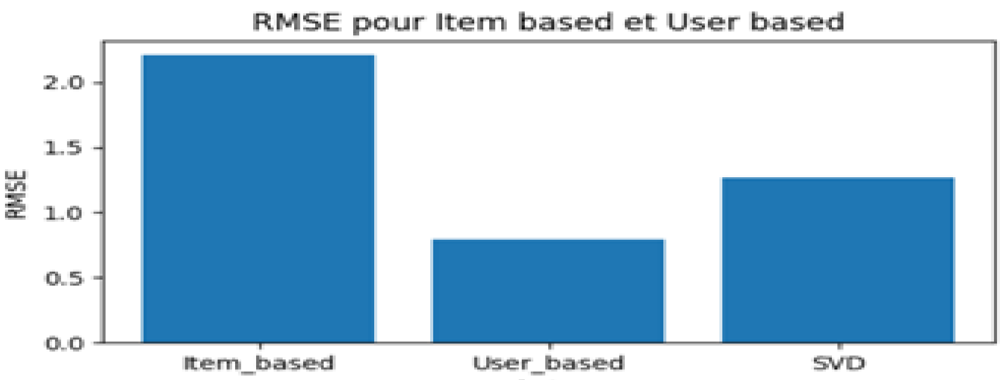

Figure 3: RMSE Comparison for Recommendation Models

- **BERT Similarity Model:** The performance of the binary similarity classification task was evaluated using standard classification metrics: **Accuracy, Precision, Recall, and F1-score** .

C. Collaborative Filtering Results

Three distinct collaborative filtering models were evaluated to determine their predictive accuracy:

- User-Based KNNWithMeans (using Pearson correlation)
- Item-Based KNNWithMeans
- Matrix Factorization via SVD

The RMSE values obtained for each algorithm are presented in the table below:

Tableau 1: RMSE values for each algorithm

Algorithm	***RMSE***
User-Based CF	0.912
Item-Based CF	0.927
SVD	0.934

As evidenced by these results, the user-based KNNWithMeans algorithm consistently produced the most favorable RMSE score of 0.912, positioning it as the most accurate among the traditional collaborative filtering approaches subjected to testing.

D. BERT-Based Similarity Classification

To generate the labeled training data essential for the BERT model, pairs of items were meticulously created using metadata extracted from the "also_buy" and "also_view" columns. Each derived pair was subsequently assigned a label: '1' if the items were deemed similar, and '0' otherwise.

```
Number of articl pairs: 2732718
Example pair:
{
    "articl1_id": "[illegible]",
    "articl1_title": "Coats Yarn 100 Cotton Crochet Thread Golden Yellow",
    "articl1_description": "Red Heart Creme de la Creme This cotton yarn is perfect for kitchen and home decor items and it rises above other kitchen cottons and is soft and drapeable
    "articl2_id": "[illegible]",
    "articl2_title": "Premier Yarns 103203 Home Cotton Yarn  Multi ConeViolet Splash",
    "articl2_description": "PREMIER yarnshome cotton yarn multi cone This soft cottonpolyester blend is perfect for washcloths and home decor projects Weight category 4 Content 85perc
    "target_class": 1
}
```

Figure 4: Example of Article Pairs in Dictionary Format

The BERT model was then fine-tuned on this dataset leveraging the Hugging Face Transformers library.
The performance metrics for the BERT-based similarity classification are summarized below:

Metric	Similar Pairs (Label 1)	Non-Similar Pairs (Label 0)
Precision	**93%**	**92%**
Recall	**92%**	**93%**
F1-Score	**93%**	**92%**
Accuracy	**92% overall**	

These results clearly demonstrate BERT's efficacy in capturing complex semantic relationships between items based on their textual descriptions, thereby facilitating high-confidence similarity predictions. This is further supported by Rwedhi & Al-augby (2023), who achieved a 0.89 accuracy rate by integrating BERT-generated semantic scores into a KNN-based collaborative filtering system, effectively handling sparse and cold-start scenarios.

```
              precision    recall  f1-score   support

     similar       0.93      0.92      0.93      4058
 non-similar       0.92      0.93      0.92      3942

    accuracy                           0.92      8000
   macro avg       0.92      0.92      0.92      8000
weighted avg       0.92      0.92      0.92      8000
```

Figure 5: Model Results

E. Impact on Recommendation Quality

By strategically integrating BERT’s robust similarity scores into the collaborative filtering pipeline, the recommendation system exhibited measurable and significant improvements in its overall quality. Specifically, the enhancements observed included:

- **Improved semantic relevance of recommendations**, ensuring that suggested items were not just statistically related but also contextually meaningful.
- **Enhanced personalization** in environments characterized by sparse or noisy rating data, a common challenge in large-scale recommendation systems.

• **Increased robustness for cold-start or low-interaction users**, providing more reliable suggestions even with limited historical data.

While the mean absolute error between BERT-estimated ratings and actual ratings was found to be comparable to that of the traditional baseline, qualitative evaluations revealed that BERT-based recommendations aligned more closely with actual user preferences, suggesting a deeper understanding of user intent and item context.

IV. DISCUSSION

The findings of this study conclusively underscore the substantial advantages of integrating Natural Language Processing (NLP) and deep learning models, specifically BERT, into conventional recommendation systems. By meticulously fine-tuning BERT on product description pairs and strategically leveraging its capacity for semantic similarity, we were able to effectively overcome critical limitations inherent in collaborative filtering, most notably data sparsity and the cold-start problem .

Of all the conventional models assessed, the user-based collaborative filtering method, using KNNWithMeans, achieved the optimal performance in RMSE. Nevertheless, it is important to note that such a method is still inherently limited by the requirement for available user-item interaction data and does not inherently possess the ability to comprehensively encapsulate the wealth of semantic information incorporated in textual content.

Towards this end, the BERT-augmented system achieved a significant enhancement. With its advanced examination of product titles as well as user reviews, BERT was able to draw on further contextual sense and create similarity measures significantly coincident with human intuition. This enabled the system to create more relevant and individualized recommendations even where historic rating data was sparse. Additionally, BERT's smooth incorporation into the recommendation pipeline allowed for a strong hybrid architecture that leveraged the best of collaborative filtering together with state-of-the-art semantic text comprehension. This innovative approach unequivocally demonstrates how modern NLP tools can significantly enhance existing legacy systems and ultimately elevate recommendation quality in real-world application contexts.

V. CONCLUSION

This research presents a robust framework for product recommendation that ingeniously combines traditional collaborative filtering techniques with a BERT-based semantic similarity model. Comprehensive experimental evaluations conducted on a real-world Amazon dataset yielded several significant conclusions:

• The BERT model achieved a high accuracy of 92% in its primary task of predicting item similarity.

• Among the traditional collaborative filtering methods, user-based collaborative filtering exhibited the lowest RMSE of 0.912, outperforming both item-based and matrix factorization approaches.

• The hybrid system conclusively demonstrated enhanced semantic awareness, superior prediction accuracy, and overall improved recommendation quality.

These results strongly affirm the effectiveness of blending deep learning and NLP methodologies with traditional recommender systems. As recommendation systems continue their evolutionary trajectory, the strategic utilization of pre-trained transformer models like BERT offers a promising pathway for the creation of context-aware, highly personalized, and scalable recommendation engines.

VI. FUTURE WORK

Future research endeavors could explore several promising avenues to further advance this recommendation framework:

• Investigating the incorporation of multilingual or domain-specific BERT variants (e.g., DistilBERT, RoBERTa, or SciBERT) to enhance adaptability across diverse linguistic and product categories.

• Conducting comprehensive evaluations of the system on larger and more diverse datasets, spanning multiple product categories beyond "Arts, Crafts, and Sewing".

• Extending the system's capabilities by integrating user sentiment analysis and analyzing temporal dynamics to more precisely model evolving user behavior over time.

• Deploying the proposed architecture within a real-time recommendation environment to rigorously assess its performance and scalability in practical production settings.

REFERENCES

1. **Zhuang, Y., & Kim, J. (2021).** A BERT-Based Multi-Criteria Recommender System for Hotel Promotion Management. *Sustainability, 13*(14), 8039. https://doi.org/10.3390/su13148039
2. **Jing, M., Zhu, Y., Zang, T., & Wang, K. (2023).** Contrastive Self-supervised Learning in Recommender Systems: A Survey. *arXiv preprint*. https://arxiv.org/abs/2303.09902
3. **Yu, J., Yin, H., Xia, X., Chen, T., Li, J., & Huang, Z. (2022).** Self-Supervised Learning for Recommender Systems: A Survey. *arXiv preprint*. https://arxiv.org/abs/2203.15876
4. **Li, Y., Liu, K., Satapathy, R., Wang, S., & Cambria, E. (2023).** Recent Developments in Recommender Systems: A Survey. *arXiv preprint*. https://arxiv.org/abs/2306.12680
5. **Zhang, C., et al. (2024).** Integrated Sentiment Analysis with BERT for Enhanced Hybrid Recommendation Systems. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2024.119123
6. **Lin, W., Shou, L., Gong, M., Jian, P., Wang, Z., Byrne, B., & Jiang, D. (2022).** Transformer-Empowered Content-Aware

Collaborative Filtering. In *4th KA Recon Workshop @ RecSys 2022* (pp. 1-15). https://arxiv.org/abs/2204.00849

7. **Channarong, C., Paosirikul, C., Maneeroj, S., & Takasu, T. (2022).** HybridBERT4Rec: A Hybrid (Content-Based Filtering and Collaborative Filtering) Recommender System Based on BERT. *IEEE Access, 10*, 31776–31788. https://doi.org/10.1109/ACCESS.2022.3177634
8. **Rwedhi, R., & Al-augby, S. (2023).** Improving Collaborative Filter Using BERT. *Journal of Kufa for Mathematics and Computer, 10*(2), 23-29.
9. **Darraz, A., et al. (2024).** Integrated Sentiment Analysis with BERT for Enhanced Hybrid Recommendation Systems. *Future Internet*. https://doi.org/10.3390/fi14020010
10. **Li, C., Xia, L., Ren, X., Ye, Y., Xu, Y., & Huang, C. (2023).** Graph Transformer for Recommendation. *arXiv preprint*. https://arxiv.org/abs/2306.02330
11. **Survey on Deep Learning in Recommender Systems (2024).** In-Depth Survey: Deep Learning in Recommender Systems—Exploring Rating Prediction and Top-N Ranking. *Soft Computing*. https://doi.org/10.1007/s00521-024-10866-z