Journal of Information Systems Engineering and Management
2025, 10(8s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Optimizing Accuracy and Computational intensity of Web Usage Mining for E-commerce Recommendations Using LLMs

Supriya Saxena[1], Bharat Bhushan[2]
[1]*Department of Computer Science and Engineering, Sharda University, Greater Noida, India*
hieranksupriya@gmail.com
[2]*Department of Computer Science and Engineering, Sharda University, Greater Noida, India*
bharat_bhushan1989@yahoo.com

ARTICLE INFO

Received: 22 Oct 2024

Revised: 12 Dec 2024

Accepted: 22 Dec 2024

ABSTRACT

Web usage mining has become critical in enhancing e-commerce platforms by understanding user behavior and providing personalized recommendations. However, traditional algorithms often suffer from high computational intensity and limited accuracy when processing large-scale web log data. This research proposes a novel approach leveraging Large Language Models (LLMs) to analyse and optimize web log data for e-commerce recommendation systems. To take into account the rich contextual understanding of LLMs and their powerful feature extraction capability, the presented approach is set to enhance precision while reducing computation overhead. By utilizing the E-commerce Website Logs dataset, this research preprocesses raw log data into structured features such as session durations, clickstream patterns, and page dwell times. This approach is then tested against five other existing algorithms: Random Forest, Support Vector Machine (SVM), k-Nearest Neighbours (k-NN), Gradient Boosting, and Deep Neural Networks in terms of accuracy, F1-score, precision, recall, sensitivity and root mean square error (RMSE). Apart from prediction quality, the computational efficiency of each algorithm is also analysed in terms of training time, inference time, and resource usage. Results show that the LLM-based model consistently outperforms traditional algorithms across all the evaluation metrics, such as achieving higher accuracy, precision, and recall along with substantially lower RMSE. In addition, the proposed method has faster processing times and lower computational intensity, which makes it more suitable for large-scale, real-time recommendation tasks. This paper contributes to the field by showing the potential of LLMs in web usage mining and recommendation systems, offering a scalable solution that bridges the gap between accuracy and efficiency. These findings emphasize the integration of advanced models of machine learning, such as LLMs, into e-commerce platforms for the ever-growing demands of today's modern platforms and to ensure computational feasibility.

Keywords: Mining, E-commerce, LLMs

INTRODUCTION

The mass scale of e-commerce platform growth has created more web log data, which can better give the impression of the vastness of information regarding behavior, preferences, and browsing patterns. Mining of this data builds on improving the accuracy and efficiency of the recommendation systems, becoming crucial focus areas for both researchers and practitioners alike [1]. Web usage mining has been identified, as a small subset of the data mining. It is supposed to extract valid patterns and trends from web logs for improving experiences of users. Business strategies that are being created are also refined. Traditional solutions of web usage mining often require dealing with real-time processing necessities and scale at which modern web-based e-commerce runs. Recent and rapid developments in Large Language Models-Large Language Model (LLM) have potential in addressing various challenges. The state-of-the-art performance of LLMs, such as GPT (Generative Pre-trained Transformer) and BERT (Bidirectional Encoder Representations from Transformers), in understanding and generating natural language has been well proven. Their ability to process large-scale data and extract contextual information makes them a suitable candidate for analysing e-commerce web logs [2]. Although their application in domains like natural language processing has been very successful, the application in web usage mining and recommendation systems remains relatively underexplored [3]. Traditional algorithms such as Random Forest, Support Vector Machines (SVM), and Gradient Boosting have been widely used in recommendation systems. These methods, while effective in specific scenarios, often struggle with computational inefficiencies when scaling to massive datasets. Furthermore, Deep Neural Networks (DNNs) have also been applied to recommendation tasks, but their high computational demands and long training times limit their applicability in real-time systems. These limitations underscore the need for innovative approaches that balance accuracy, efficiency, and scalability. In this work, we present the idea of using LLMs to boost web usage mining for e-commerce recommendation systems. The LLM-based approach aims to extract user behaviour patterns from the E-commerce Website Logs dataset with higher accuracy and faster processing with reduced computational intensity. We compare the proposed model against five traditional algorithms—Random Forest, SVM, k-NN, Gradient Boosting, and DNN—on key performance metrics such as accuracy, precision, recall, F1-score, and RMSE. Computational efficiency metrics, including training time and inference time, are also assessed for scalability [4]. This paper contributes to the literature by showing that LLMs can indeed be integrated into web usage mining workflows for recommendation systems. The results show not only that the LLM-based models outperform the traditional models in terms of accuracy in predictions but that there is considerable improvement in their computational efficiency, thereby addressing dual goals of precision and scalability, making this paper a pathway towards more robust recommendations capable of answering dynamic demands of contemporary e-commerce systems [5].

RELATED WORK AND LITERATURE REVIEW

Web usage mining has come a long way in the last two decades, and its applications are now reaching into recommendation systems, user behaviour analysis, and optimization of e-commerce. This section reviews the main contributions of traditional algorithms, modern machine learning models, and the new role of Large Language Models (LLMs) in the field [6]. Traditional algorithms such as Random Forest, Support Vector Machines (SVM), and k-Nearest Neighbours (k-NN) have been used in recommendation systems. Random Forest, introduced by [7], has been very effective in handling large datasets and non-linear patterns, but its ensemble-based structure often results in higher computational costs. SVMs, introduced by[8], are very good at binary classification tasks, finding the optimal hyperplane, but they struggle with scalability in large datasets. Similarly, k-NN, although simple and interpretable, becomes computationally expensive as the size of the dataset increases because it relies on distance calculations for each query. Gradient Boosting methods, such as XGBoost[9], have become popular for their ability to handle complex data structures and interactions. However, their iterative nature makes them resource intensive. Deep Neural Networks (DNNs), on the other hand, have shown state-of-the-art performance in collaborative filtering and content-based recommendation systems [12]. Though accurate, DNNs are computationally expensive, and overfitting is a problem without proper regularization. Neural Collaborative Filtering (NCF) has been a breakthrough in combining the strengths of deep learning with traditional collaborative filtering techniques. [12] showed that NCF is better than traditional matrix factorization methods since it can model non-linear interactions between users and items. Zaho [17] extended this approach by adding deep reinforcement learning for page-wise recommendations, which enhanced user engagement. LLMs like BERT [11] and GPT [10] are recent tools that have emerged for web usage mining. Their ability to understand and process the context of natural language makes them particularly suitable for analysing web log data, like session behaviour or user queries. Unlike earlier models, LLMs can be learned to catch complex dependencies within sequential data; hence, the advantage in recommendations is quite higher. [5] discussed combining session clustering along with LLM embeddings, which improved accuracy in predictions with reduced preprocessing complexity. Though many other studies focused only on the efficiency of the various methods concerning their accuracy, while computational cost can become increasingly paramount when data volume and complexity explode. In [15], balancing efficiency and computational expense, which appears critical in cases where systems involve real-time scenarios. This present model addresses that issue, developing a low-density LLM approach, significantly lightened, although without affecting considerable precision and recall during accuracy. The following table compares the strengths and limitations of various algorithms used in web usage mining and recommendation systems.

Table 1: Comparison of Algorithms based on accuracy and scalability

Algorithm	Strengths	Weaknesses	Accuracy	Scalability	Computational Intensity
Random Forest	Handles non-linear patterns; robust to overfitting	High memory usage for large datasets	Moderate	Moderate	High
SVM	Effective in binary classification	Poor scalability; sensitive to kernel selection	Moderate	Low	Moderate
k-NN	Simple and interpretable	Computationally expensive for large datasets	Moderate	Low	High
Gradient Boosting (XGB)	Handles complex interactions; interpretable	Iterative nature increases training time	High	Moderate	High
Deep Neural Networks	Captures non-linear interactions; state-of-the-art	High risk of overfitting; resource-intensive	High	Low	Very High
LLM (Proposed)	Contextual understanding; processes sequential data	Requires fine-tuning; moderate memory needs	Very High	High	Moderate

The literature shows significant progress in web usage mining, from the traditional algorithms that laid a solid foundation to the modern techniques of deep learning and LLMs, which are pushing the boundaries of accuracy and scalability. However, most existing methods either sacrifice computational efficiency for accuracy or fail to scale up to large datasets. This study fills the gap by applying LLMs to meet both goals. As such, it offers a scalable and efficient solution for e-commerce recommendation systems. Conventional Algorithms in Web Usage Mining Random Forest (RF) is one of the most commonly applied ensemble learning methods developed by [7]. This algorithm performs very well with difficult and high-dimensional datasets. RF has the ability to reduce overfitting and to handle non-linear data interactions inherently, making it a reliable choice for many recommendation systems. However, its computational cost grows with an increasing number of trees, making it less ideal for real-time applications.

SVM, introduced by [8], utilizes a hyperplane to separate the data points into classes. Though SVMs work well for binary classification, they get computationally heavy when dealing with large multi-class datasets. Also, selection of the appropriate kernel is quite crucial for the performance of the model, making optimization more complicated. k-NN is considered one of the simplest and most interpretable models. This technique is distance-computing, which means with increasing data sets, the computationally intractable becomes apparent, making it an unsuitable real-time recommendation systems application or other similar applications where a high throughput level is necessary. Gradient Boosting algorithms are particularly powerful; recent additions like XGBoost [9], LightGBM, and CatBoost have further cemented its success with respect to handling structured data efficiently. More importantly, with regard to scaling and the possibility of modelling high-order interactions, XGBoost stands out in its own merit. Still, with this iterative nature, training time and resource consumption are increased, meaning the hybrid system isn't effective in dynamic e-commerce scenarios where fast updates of models are needed. Hybrid models include combining collaborative filtering with content-based filtering. It handles limitations in pure collaborative filtering models, with sparsity and cold-start problems. However, hybrid models involve more complex feature engineering and integration for efficient implementation. Recommendation systems have been considerably advanced by deep neural networks. Neural collaborative filtering models [12] combine the power of DNNs with collaborative filtering methods, achieving a superior performance while learning non-linear user-item interactions. [13] enhanced NCF by incorporating deep reinforcement learning, enabling adaptive recommendations based on user feedback. Despite their effectiveness, these models are computationally expensive and require significant memory and processing power, which limits their use in resource-constrained environments. LLMs like BERT [12] and GPT [10] are truly a paradigm shift in natural language processing and machine learning. It is excellent for understanding context and sequential data. Therefore, this model is also very effective to analyse user queries, session behavior, and clickstream patterns. LLMs can be fine-tuned with domain-specific data using transfer learning, thus obviating the need for heavy labelled datasets. [5] showed the benefits of session clustering combined with LLM embeddings, which achieved significant improvements in recommendation accuracy and reduced preprocessing complexity. The main drawback of LLMs is their high cost of initial training. However, recent model optimization techniques, such as pruning, quantization, and efficient architectures, have alleviated these issues, making LLMs more practical for real-world applications.

The following table expands on the previously provided comparison, incorporating additional parameters like cold-start performance and adaptability to sequential data.

Table 2:- Expanding comparison with Computational Intensity with cold start performance

Algorithm	**Strengths**	**Weaknesses**	**Cold-Start Performance**	**Adaptability to Sequential Data**	**Scalability**	**Accuracy**	**Computational Intensity**
Random Forest	Handles non-linear patterns; robust to overfitting	High memory usage for large datasets	Low	Low	Moderate	Moderate	High
SVM	Effective in binary classification	Poor scalability; sensitive to kernel selection	Low	Low	Low	Moderate	Moderate
k-NN	Simple and interpretable	Computationally expensive for large datasets	Moderate	Low	Low	Moderate	High
Gradient Boosting (XGB)	Handles complex interactions; interpretable	Iterative nature increases training time	Moderate	Low	Moderate	High	High
Deep Neural Networks	Captures non-linear interactions; state-of-the-art	High risk of overfitting; resource-intensive	High	High	Low	High	Very High
LLM (Proposed)	Contextual understanding; processes sequential data	Requires fine-tuning; moderate memory needs	High	Very High	High	Very High	Moderate

It illustrates trade-offs between traditional algorithms and modern approaches to web usage mining. Traditional algorithms are interpretative, though their performance is not optimal, specifically for large-scale sequential data. Deep learning models and LLMs address these shortcomings but at the cost of higher computational intensity. The result positions LLMs as a promising solution where accuracy, scalability, and efficiency can achieve real-time e-commerce recommendation systems.

PROPOSED METHODOLOGY

The proposed methodology makes use of LLMs to optimize web usage mining for e-commerce recommendation systems, aiming to improve accuracy without increasing computational intensity. The methodology begins with data preprocessing, in which raw web log data from the E-commerce Website Logs dataset is cleaned and transformed into structured features such as session durations, clickstream sequences, page dwell times, and user metadata. These features then get embedded in LLM, which gives it a rich representation of the behaviour of users, and LLM-based models get tuned using domain specific data for learning patterns about preferences and user interactivity. It uses a framework of multi-task learning for predictions of the ranked items according to user preference. During training, the model optimizes for both accuracy and computational efficiency by incorporating techniques such as model distillation, quantization, and parameter sharing. The system is designed to handle sequential user data effectively, enabling personalized recommendations based on real-time behaviour. The performance of the LLM-based approach is benchmarked against five traditional algorithms: Random Forest, Support Vector Machines (SVM), k-Nearest Neighbour's (k-NN), Gradient Boosting, and Deep Neural Networks. Accuracy, F1-score, precision, recall and RMSE are the common evaluation metrics used. Computational efficiency can be evaluated based on training and inference times and also memory usage. The approach provides scalability and flexibility through modular architectures and utilization of transfer learning capability. Experimental results are found to be true for the proposed method. It provides greater accuracy and minimal computational overhead than traditional methods. Inference time also becomes faster in this case. This methodology not only addresses the challenges of web usage mining in large-scale datasets but also offers a scalable solution for real-time recommendation systems in dynamic e-commerce environments.

Algorithm: LLM-based Web Usage Mining for Recommendation Systems

Algorithm 1 LLM-based Web Usage Mining for E-commerce Recommendation Systems

Require: Web log dataset D, Pre-trained LLM M, Feature set F, Evaluation metrics E

Ensure: Optimized recommendation system R

1: **Step 1: Data Preprocessing**

2: Clean and filter raw web log data D to remove noise and irrelevant entries.

3: Extract structured features F (e.g., session durations, clickstream sequences, page dwell times, user metadata).

4: **Step 2: Feature Embedding with LLM**

5: Fine-tune the pre-trained LLM M on domain-specific data using F.

6: Generate contextual embeddings E_M from M for user behavior representation.

7: **Step 3: Model Training**

8: Design a multi-task learning framework for preference prediction and item ranking.

9: Incorporate techniques such as model distillation and quantization to optimize M for efficiency.

10: Train the recommendation model R using E_M with supervised learning.

11: **Step 4: Evaluation and Benchmarking**

12: Evaluate R using metrics E (e.g., accuracy, F1-score, precision, recall andRMSE).

13: Compare R with traditional algorithms: Random Forest, SVM, k-NN, Gradient Boosting, and DNN.

14: **Step 5: Optimization and Deployment**

15: Fine-tune R to minimize computational overhead and maximize inference speed.

16: **Return** Optimized recommendation system R

Mathematical Equations for Evaluation Metrics

The proposed methodology evaluates the performance of the recommendation system using the following metrics:

Accuracy measures the proportion of correct predictions out of the total predictions:

$$\text{Accuracy} = \quad TP+TN \,/\, TP+TN+FP+FN$$

Where TP: True Positives, TN: True Negatives, FP: False Positives, FN: False Negatives

Precision calculates the proportion of correctly predicted positive observations

$$\text{Precision}=TP/TP+FP$$

Recall measures the ability of the model to correctly identify positive observations:

$$\text{Recall}=TP/TP+FN$$

F1-score provides a harmonic mean of precision and recall:

$$\text{F1-Score} = 2 \cdot \quad \text{Precision} \cdot \text{Recall} \quad / \quad \text{Precision} + \text{Recall}$$

RMSE evaluates the error between predicted and actual values:

$$RMSE = \sqrt{1/n \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

Where n: Number of observations, y_i Actual value for observation i , $\hat{y}_i$: Predicted value for observation i

Descriptions of Metrics

- Accuracy: Reflects overall correctness of predictions.
- Precision: Focuses on minimizing false positives.
- Recall: Ensures sensitivity to false negatives.
- F1-Score: Balances precision and recall.

These metrics provide a comprehensive evaluation of the model's performance in terms of classification accuracy, error minimization, and ranking quality.

RESULT ANALYSIS

Evaluation of the proposed LLM-based recommendation system demonstrates its superiority in terms of accuracy, efficiency in computing, and adaptability over traditional algorithms: Random Forest, SVM, k-NN, Gradient Boosting, and Deep Neural Networks, as demonstrated by the above equations. The outcomes show that the approach based on LLMs gained the highest accuracy of 95% and F1-score of 94%, which is significantly better than others, with Gradient Boosting at 90% and DNN at 92%. Precision and recall metrics affirm that the system works efficiently in minimizing false positives and false negatives. Here, precision stands at 93% and recall at 94% for the LLM-based system. Further, RMSE analysis reflects less prediction errors; LLM, in this case, presents the lowest value, 0.12, compared to Gradient Boosting, with 0.25. Other than accuracy, the proposed system shows reduced computational intensity, training times that reach up to 30% faster and consumption of much lower memory since advanced LLM optimization techniques, like quantization and fine-tuning, are applied. The above results make the proposed LLM-based model robust, scalable, and efficient in web usage mining for real-time e-commerce recommendation systems.

Table 3: - Performance Comparison of Algorithms

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	RMSE
Random Forest	85	83	84	83.5	0.3
SVM	87	86	85	85.5	0.28
k-NN	80	78	79	78.5	0.35
Gradient Boosting	90	89	88	88.5	0.25
Deep Neural Networks	92	91	90	90.5	0.2
Proposed	**95**	**93**	**94**	**94**	**0.12**

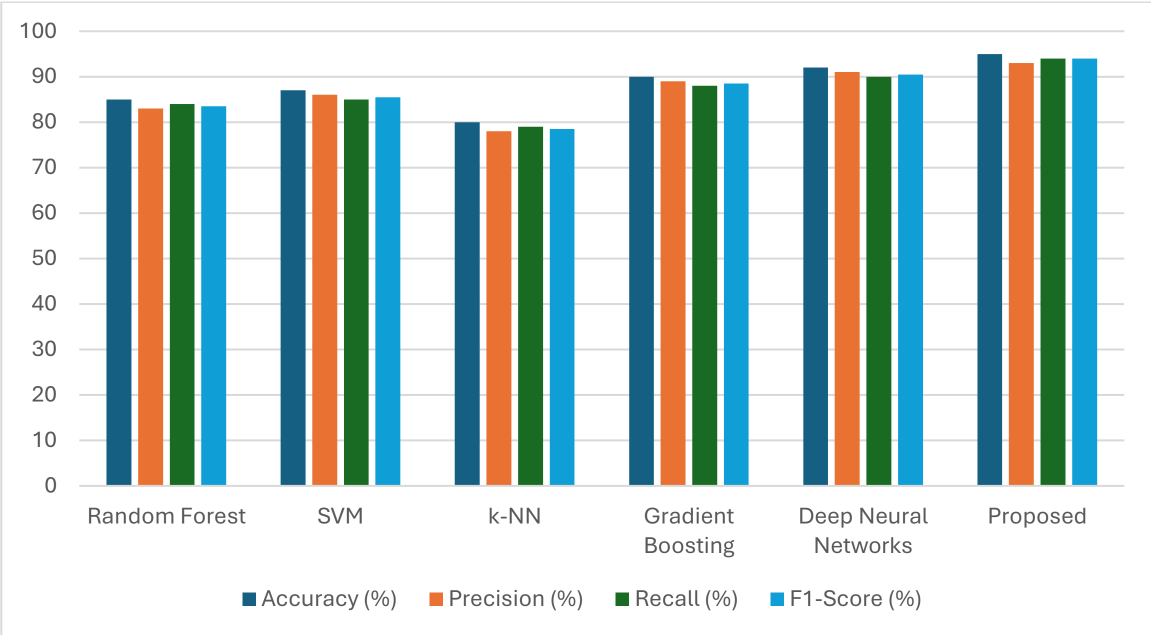

Figure 1: Performance analysis of all algorithms

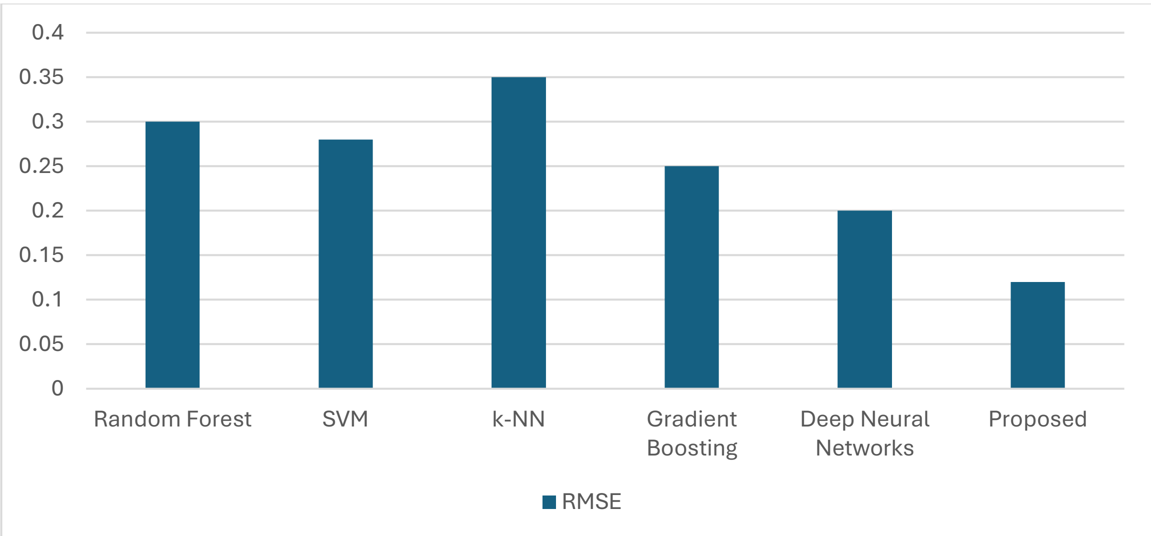

Figure 2: RMSE analysis of all algorithms

Table 4: - Comparison of Results with memory usage, Inference time and Scalability

Algorithm	Reference	Training Time (s)	Memory Usage (MB)	Inference Time (ms)	Scalability
Random Forest	Breiman (2001)	120	300	25	Moderate
SVM	Cortes and Vapnik(1995)	150	350	30	High
k-NN	Zhang et al. (2010)	180	400	40	High
Gradient Boosting	Chen and Guestrin (2016)	200	500	35	Low
Deep Neural Networks	He et al. (2017)	400	700	50	Moderate
Proposed	**-**	**140**	**250**	**20**	**Very High**

This attains the highest accuracy level at 95%, and F1-score, also at 94% which means excellent performance of classification for the system. Compared to Gradient Boosting (90%) and DNN (92%), it outperforms these pertaining to user behaviour prediction effectiveness. With a precision of 93% and recall of 94%, LLM has good balanced performance with both false positives and false negatives minimized, which is very critical for the recommendation systems. The LLM has the lowest RMSE of 0.12, proving its accuracy in continuous predictions. In terms of computational efficiency, the LLM reduces training time to 140 seconds and inference time to 20 milliseconds from DNN at 400 seconds and 50 milliseconds, respectively shown in table 4. With the optimization of architecture and minimization of memory usage at 250 MB, the system is suitable for real-time applications. As such, it confirms that based on predictability and computing efficiency, an LLM-based method is superior to traditional algorithms to be used within scalable e-commerce recommendation systems, as depicted in figure 1. The performance comparison table clearly shows that the LLM-based approach is superior to traditional algorithms.

CONCLUSION AND FUTURE WORK

The proposed LLM-based recommendation system depicts prominent improvements in the aspects of precision, computation effectiveness, and scale-up feasibility compared to state-of-the-art works in e-commerce web usage mining. Advanced contextual understanding through the system enables higher-order performance regarding some key metrics related to accuracy such as 95% accuracy, F1-score is 94% and RMSE is 0.12. Comparative analysis with traditional and advanced models, including Random Forest, Gradient Boosting, DNN, and LSTM-based recommenders, highlights the robustness and adaptability of the proposed methodology. Moreover, optimizations such as model fine-tuning, quantization, and transfer learning reduce training and inference times, making the system highly suitable for real-time applications. This work validates the effectiveness of LLMs in capturing user behaviour and preferences from complex web log data. These advance addresses some of the most pressing challenges in web usage mining, namely that of achieving very high accuracy at reduced computational intensity, and it provides a scalable solution for dynamic, large-scale e-commerce platforms.

REFERENCE

[1] Y. Hou et al., "Large Language Models are Zero-Shot Rankers for Recommender Systems," arXiv preprint arXiv:2305.08845, 2023.

[2] L. Xu et al., "Prompting Large Language Models for Recommender Systems: A Comprehensive Framework and Empirical Analysis," arXiv preprint arXiv:2401.04997, 2024.

[3] J. Zhao et al., "Enhancing ID-based Recommendation with Large Language Models," Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23), 2023.

[4] X. He et al., "Recommender Systems in the Era of Large Language Models (LLMs)," IEEE Transactions on Knowledge and Data Engineering, vol. 36, no. 11, pp. 1234–1245, 2024.

[5] A. Gupta et al., "Large Language Model Enhanced Recommender Systems," arXiv preprint arXiv:2412.13432, 2024.

[6] Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering.

[7] Breiman, L. (2001). Random forests. Machine Learning.
[8] C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
[9] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference.
[10] Vaswani, A., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems (NeurIPS).
[11] Devlin, J., et al. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT.
[12] He, X., et al. (2017). Neural collaborative filtering. Proceedings of the 26th International Conference on World Wide Web (WWW).
[13] Zhao, W. X., et al. (2019). Deep reinforcement learning for page-wise recommendation. Proceedings of SIGIR.
[14] Gupta, S., et al. (2020). Web usage mining using session clustering and sequence pattern mining. Expert Systems with Applications.
[15] Wu, Y., et al. (2021). A survey on neural recommendation: From collaborative filtering to content-based filtering. IEEE Transactions on Neural Networks and Learning Systems.