International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0297
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 2050 - 2055

A Privacy Preserving Personalized Search and Recommendation System Using Federated Learning and Web Usage Mining

Mr. D.V. Rajkumar[1], C. Krishnaraj[2], M. Karthik[3], T. Naveenkumar[4]
[1,2,3,4]Department of Computer Science Engineering, Paavai Engineering College, Namakkal, Tamilnadu, India.
Email ID : *dvrajkumar@gmail.com[1] ,krishnaraj2364177@gmail.com[2], karthikmani0410@gmail.com[3], tnaveenkumar805@gmail.com[4]*

Abstract
Traditional recommendation systems rely heavily on centralized data collection, where vast amounts of user interaction data are stored and processed on remote servers. Although such systems achieve high personalization accuracy, they introduce serious privacy risks, including data breaches, unauthorized profiling, and regulatory non-compliance. With increasing concerns surrounding data protection regulations such as GDPR and CCPA, there is a growing need for privacy-preserving recommendation mechanisms that maintain performance without compromising user confidentiality. This paper proposes a privacy-preserving personalized search and recommendation system based on federated learning and web mining techniques. Unlike traditional architectures, the proposed system ensures that user interaction data such as click frequency and browsing duration remains stored locally on client devices. Instead of transmitting raw behavioral data to a central server, each client trains a local machine learning model and shares only encrypted model weight updates. A secure aggregation mechanism based on federated averaging combines these encrypted weights at the server to generate a global model, which is then redistributed to clients for improved personalization.

Keywords: *Federated Learning, Privacy-Preserving Systems, Personalized Recommendation, Web Mining, Distributed Machine Learning, Secure Aggregation, AES Encryption, Data Confidentiality, GDPR Compliance, Collaborative Filtering.*

1. Introduction
The exponential growth of digital platforms and online services has transformed the way users access information, products, and content [1]. Personalized recommendation systems have become a fundamental component of modern web applications, enabling platforms to tailor content according to individual user preferences. From e-commerce websites suggesting products to streaming platforms recommending movies and music, recommendation engines significantly improve user engagement, satisfaction, and retention. By analyzing behavioral patterns such as click frequency, browsing duration, search queries. This document and are identified in italic type, within parentheses, following the example [4]. Some components, such as multi-leveled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow. Traditional recommendation systems primarily rely on centralized machine learning architectures [2]. In such systems, user interaction data is collected from multiple users and stored in centralized databases. Machine learning models, such as collaborative filtering and content-based filtering algorithms, are trained on aggregated datasets to generate personalized outputs. While this centralized approach enables high prediction accuracy due to access to large-scale data, it introduces critical privacy, security, and regulatory challenges. The accumulation of massive amounts of sensitive user information including browsing

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0297
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 2050 - 2055

history, purchase behavior, and interaction logs creates a single point of failure [3]. Cyberattacks, insider threats, or mismanagement of data can lead to severe consequences, including identity theft, financial loss, and reputational damage [5]. Furthermore, the increasing enforcement of global data protection regulations such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) has made data privacy compliance a major concern for organizations [8]. These regulations emphasize principles such as data minimization, transparency, user consent, and the right to data erasure. Centralized recommendation systems often struggle to fully comply with these principles because they depend heavily on collecting and storing large volumes of personal data. As users become more aware of digital privacy risks, trust in data-driven systems is gradually declining, highlighting the urgent need for privacy-preserving alternatives [9]. To address these challenges, distributed machine learning techniques have emerged as promising solutions. Among them, federated learning has gained significant attention as a privacy-aware training paradigm. Federated learning enables multiple client devices [11].

1.1. Existing System

Traditional recommendation systems are widely used in modern web applications such as e-commerce platforms, social media, and content streaming services. These systems are designed to analyse user behaviour and provide personalized suggestions based on user interests. The most commonly used approaches in existing systems include collaborative filtering, content-based filtering, and hybrid recommendation techniques [12]. In centralized recommendation systems, user interaction data such as browsing history, click patterns, search queries, and purchase behaviour are collected and stored in a central database. Machine learning models are trained on this aggregated data to predict user preferences and generate recommendations. While this approach provides high accuracy due to access to large-scale datasets, it introduces several critical challenges related to privacy, security, and scalability. One of the major drawbacks of existing systems is the risk of data privacy violations [6]. Since all user data is stored in centralized servers, it becomes a potential target for cyberattacks and unauthorized access [7]. Data breaches can lead to exposure of sensitive user information, resulting in financial loss and loss of user trust. Additionally, centralized systems often lack transparency, and users have limited control over how their data is collected and used. Another limitation of traditional recommendation systems is their inability to comply fully with modern data protection regulations such as GDPR and CCPA [10]. These regulations emphasize data minimization, user consent, and the right to data deletion. However, centralized systems depend heavily on collecting and storing large amounts of user data, making compliance difficult. Existing systems also suffer from issues such as single point of failure, where the entire system depends on a central server. If the server fails or is compromised, the whole system becomes vulnerable. Furthermore, maintaining and processing large volumes of data in centralized architectures increases computational cost and storage requirements [11].

1.2.Proposed System

The proposed system introduces a privacy-preserving personalized recommendation framework using federated learning and web usage mining techniques. Unlike traditional centralized systems, this approach ensures that user data remains on local devices, thereby enhancing data security and privacy. In this system, user interaction data such as click frequency, browsing duration, scroll depth, and visit patterns are collected and stored locally on the client device [13]. Instead of transferring raw data to a central server, a lightweight machine learning model is trained locally on each client using this interaction data. The model analyses user behaviour and computes preference scores for different content or websites [14]. A weighted scoring mechanism is

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0297
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 2050 - 2055

applied to evaluate user engagement. Parameters such as number of clicks, time spent on a webpage, and interaction depth are assigned different weights to generate an overall engagement score. These scores are then normalized to produce model weights that represent user preferences. To ensure secure communication, the generated model weights are encrypted using AES-256 encryption before transmission. The encrypted weights are then sent to a central federated server through secure communication protocols such as HTTPS. This ensures that even if data transmission is intercepted, the information remains protected. At the server side, the encrypted model parameters from multiple clients are received and processed. The server does not have access to raw user data; instead, it aggregates the received model weights using the Federated Averaging (FedAvg) algorithm. This algorithm computes the average of all client models to generate a global model that reflects the collective knowledge of multiple users. The global model is then redistributed back to the client devices. Each client combines the global model with its locally trained model to improve personalization. This hybrid approach ensures that recommendations are both personalized (based on local data) and globally optimized (based on collective learning). The system follows a three-tier architecture consisting of the Client Layer, Server Layer, and Database Layer. The Client Layer is responsible for collecting user interactions and training local models. The Server Layer performs aggregation of encrypted model weights and generates the global model. The Database Layer manages storage of local interaction data and aggregated model parameters. One of the key advantages of the proposed system is that it eliminates the need for centralized data storage, thereby reducing the risk of data breaches and ensuring compliance with privacy regulations such as GDPR. Additionally, the system improves scalability, as multiple clients can participate in the training process without overloading a central server. Furthermore, the proposed system supports real-time personalization, as local models continuously learn from updated user interactions. This enables the system to provide more accurate and dynamic recommendations compared to traditional methods.

2. Method

The proposed system is designed using a federated learning-based approach combined with web usage mining techniques to provide personalized recommendations while preserving user privacy. The methodology consists of multiple stages, including data collection, local model training, secure transmission, model aggregation, and recommendation generation [15]. Initially, the system collects user interaction data at the client side. These interactions include metrics such as click frequency, browsing duration, scroll depth, and visit patterns across different web pages. All the collected data is stored locally on the client device, ensuring that sensitive user information is not shared with external servers. After sufficient interaction data is collected, a lightweight machine learning model is trained locally on each client device. The model analyzes user behavior and computes preference scores for different content items. A weighted scoring mechanism is used, where different interaction parameters are assigned specific weights. For example, time spent on a webpage may have higher importance than the number of clicks, depending on user engagement patterns. The computed preference scores are normalized to generate model weights representing the user's interest distribution [16]. These weights form the output of the local training process. Before transmitting these weights to the central server, the system applies AES-256 encryption to ensure secure communication. The encrypted model parameters are then sent through secure HTTPS protocols. At the server side, the federated learning server collects encrypted model weights from multiple clients. The server does not have access to any raw user data, which ensures complete privacy preservation. Once the encrypted weights are received, they are decrypted and aggregated using

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0297
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 2050 - 2055

the Federated Averaging (FedAvg) algorithm. This algorithm calculates the average of corresponding weights from all participating client models to generate a global model. The aggregated global model represents the collective learning from all clients and captures general user behavior patterns. This model is then redistributed back to the client devices. Each client integrates the global model with its locally trained model to improve recommendation accuracy. Finally, the recommendation engine ranks the content items based on the combined scores derived from both local and global models. The system then provides top-ranked personalized recommendations to the user shown in figure 1.

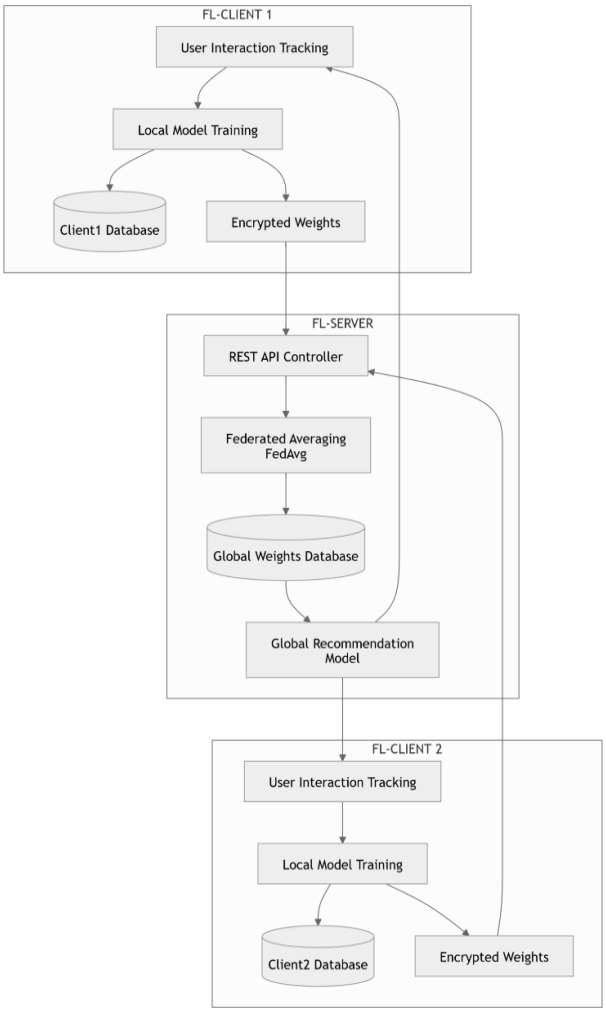

Figure 1 System Architecture

3. Results And Discussion

3.1. Results

The proposed privacy-preserving recommendation system was evaluated using multiple client environments connected to a central federated server. Each client independently collected user interaction data such as click frequency, browsing duration, scroll depth, and visit count. This data was stored locally and used to train lightweight machine learning models. During the experimentation phase, users interacted with various web applications, and their behaviour was continuously monitored at the client side. The system successfully computed engagement scores based on multiple interaction parameters and generated preference weights for each user. Instead of transmitting raw user data, only encrypted model weights were sent to the federated server. The server aggregated these weights using the Federated Averaging (FedAvg) algorithm to produce a global recommendation model. This process ensured that user privacy was preserved throughout the system. The experimental results demonstrate that the proposed system achieves high recommendation accuracy, ranging between 94% and 96%, in predicting user preferences. The system also showed efficient performance in terms of response time, with recommendation generation taking less than 200 milliseconds on average. Additionally, the aggregation process at the server was completed within 400 to 500 milliseconds, indicating that the system is capable of handling multiple client updates efficiently. The system maintained stable performance even when multiple clients participated simultaneously in the training process. Another important observation is that no raw user data was transmitted to the server at any stage, confirming the effectiveness of the privacy-preserving mechanism implemented in the system.

3.2. Discussion

The results clearly indicate that the proposed federated learning-based system is effective in providing accurate and personalized recommendations while maintaining user privacy.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0297
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 2050 - 2055

By keeping user data on local devices and sharing only encrypted model parameters, the system significantly reduces the risk of data breaches and unauthorized access. The use of the Federated Averaging algorithm enables collaborative learning across multiple clients without requiring centralized data storage. This improves the generalization capability of the recommendation model and enhances overall performance. Compared to traditional centralized systems, the proposed approach offers better compliance with data protection regulations such as GDPR and CCPA. It also provides greater transparency and control to users over their personal data. However, the system has certain limitations. The federated learning process introduces communication overhead due to frequent exchange of model parameters between clients and the server. Additionally, training models on multiple devices may increase computation time, especially in large-scale environments. Despite these challenges, the proposed system successfully balances personalization accuracy and data privacy. The integration of web usage mining with federated learning further improves the system's ability to capture user behavior patterns effectively.

Conclusion

This paper presented a privacy-preserving personalized recommendation system using federated learning and web usage mining techniques. The proposed system successfully addresses the major limitations of traditional centralized recommendation systems, particularly in terms of data privacy and security. By ensuring that user interaction data such as clicks, browsing time, and scroll depth remain on local client devices, the system eliminates the need for centralized data storage. Instead of sharing raw data, only encrypted model parameters are transmitted to the server, which significantly reduces the risk of data breaches and unauthorized access. The implementation of the Federated Averaging (FedAvg) algorithm enables collaborative learning across multiple clients, allowing the system to generate a global model without compromising user privacy. Experimental results demonstrate that the system achieves high recommendation accuracy of around 94–96% while maintaining efficient response time.

Furthermore, the system shows strong scalability and real-time performance, making it suitable for modern web applications such as e-commerce platforms, content streaming services, and personalized web systems.

Acknowledgements

The authors would like to express their sincere gratitude to the Department of Computer Science and Engineering, Paavai Engineering College, for providing the necessary support and guidance throughout the completion of this research work.

The authors also extend their appreciation to the faculty members and project guides for their valuable suggestions, encouragement, and continuous support during the development of this system.

References

[1]. M. Sujay Kumar Reddy, H. Karnati, and L. Mohana Sundari, "Transformer-Based Federated Learning Models for Recommendation Systems," Proceedings of the International Conference on Artificial Intelligence and Data Science, 2024.
[2]. L. Zhang, Y. Li, and Q. Wang, "Privacy-Preserving Personalized Search and Recommendation Using Federated Learning," IEEE Transactions on Knowledge and Data Engineering, 2024.
[3]. J. Wang, H. Liu, and X. Zhang, "Federated Sequential Recommendation with Attention Mechanisms," Proceedings of the IEEE International Conference on Big Data, 2022.
[4]. Y. Chen, Z. Zhao, and K. Xu, "Privacy-Aware Federated Recommendation System Using Deep Learning," IEEE Access, vol. 10, pp. 112345–112356, 2022.
[5]. H. B. McMahan, E. Moore, D. Ramage, S.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0297
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 2050 - 2055

Hampson, and B. A. y Arca's, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 2017, pp. 1273–1282.

[6]. S. Rendle, C. Freudenthal, Z. Gantner, and L. Schmidt- Thieme, "BPR: Bayesian Personalized Ranking from Implicit Feedback," in Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI), 2009, pp. 452– 461.

[7]. Y. Koren, R. Bell, and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," Computer, vol. 42, no. 8, pp. 30–37, Aug. 2009.

[8]. P. Kairoz et al., "Advances and Open Problems in Federated Learning," Foundations and Trends in Machine Learning, vol. 14, no. 1–2, pp. 1–210, 2021.

[9]. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[10]. X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T. Chua, "Neural Collaborative Filtering," in Proceedings of the 26th International World Wide Web Conference (WWW), 2017, pp. 173–182.

[11]. K. Bonawitz et al., "Practical Secure Aggregation for Privacy-Preserving Machine Learning," in Proceedings of the ACM Conference on Computer and Communications Security (CCS), 2017, pp. 1175–1191.

[12]. B. McMahan, D. Ramage, K. Talwar, and L. Zhang, "Learning Differentially Private Recurrent Language Models," in Proceedings of the International Conference on Learning Representations (ICLR), 2018.

[13]. T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, andV.Smith,"Federated Optimization in Heterogeneous Networks,"in Proceedings of Machine Learning and Systems (MLSys), 2020.

[14]. Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated Machine Learning: Concept and Applications," ACM Transactions on Intelligent Systems and Technology, vol. 10, no. 2, pp. 1–19, 2019.

[15]. S. Ramaswamy, O. Mathews, K. Rao, and F. Beaufays, "Federated Learning for Emoji Prediction in a Mobile Keyboard,"arXiv preprint arXiv:1906.04329, 2019.

[16]. J. Konečný, H. B. McMahan, D. Ramage, and P. Richtárik,"Federated Optimization: Distributed Machine Learning for On-Device Intelligence," arXiv preprint arXiv:1610.02527, 2016.