iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** X **Month of publication:** October 2024

DOI: https://doi.org/10.22214/ijraset.2024.64532

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue X Oct 2024- Available at www.ijraset.com

Federated Learning for Personalized Financial Recommendations

Aditya Patil[1], Rohit Maske[2]
[1]*Department of Information Technology, RMD Sinhgad School of Engineering, Pune, India*
[2]*Department of Computer Engineering, RMD Sinhgad School of Engineering, Pune, India*

Abstract: ***Federated Learning (FL) has been the solution to one of those biggest contributors that have scared us from using Machine learning, i.e. Privacy [for sensitive domains like finance]. This survey focuses on leveraging Federated Learning in creating personalized recommendation systems for financial products such as loans, investment plans, credit cards and savings. For example, classical recommendation systems usually work in a centralized way to store a lot of sensitive financial data (or derivative structures), this setting presents other well-known privacy issues and security risks. FL provides a solution that enables multiple financial institutions to work together to train a global recommendation model without the need any of them to share their raw customer data. The paper reviews federated learning, and some key take ways are on differential privacy and secure aggregation to protect customer data while improving the quality of recommendations. We also talk about how to solve some problems like heterogeneity of the data, model convergence and regulatory constraints required in financial methods. The paper concludes by detailing future directions for federated learning in finance, with focus on how personalized financial recommendations can be dramatically transformed while preserving user privacy.***

Keywords: Integrated system, Encryption, Machine Learning, Decentralized, Recommendation, adaptability, Federated learning, Privacy.

I. INTRODUCTION

The rise of machine learning (ML) has greatly impacted various industries, particularly the financial sector, where predictive models are used for fraud detection, risk analysis, and customer insights. However, applying machine learning to sensitive financial data raises concerns about privacy and security. Traditional centralized ML models require aggregating large amounts of data in one location, which makes them vulnerable to breaches and unauthorized access. Financial institutions, dealing with sensitive information like personal details and transaction histories, must adhere to strict privacy regulations, such as GDPR and CCPA, while seeking advanced data-driven solutions. Federated Learning (FL) offers a promising solution to address these concerns. Unlike traditional models, Federated Learning allows data to remain decentralized, on the devices or servers of individual institutions, while only sharing model updates with a central server. This ensures that raw financial data never leaves its source, reducing the risk of data exposure. By decentralizing model training, FL improves privacy without sacrificing the predictive power of machine learning models, making it an ideal approach for industries that handle sensitive data, such as finance. This paper explores how Federated Learning can enhance security and privacy in the financial sector, particularly in machine learning applications. It will examine key privacy-preserving techniques used in Federated Learning, such as encryption and differential privacy, and address the challenges of implementing FL in financial institutions. The study aims to demonstrate how Federated Learning can provide a balance between innovation and data security in a highly regulated industry.

II. RELATED WORK

In the research work conducted by McMahan and co-authors within the work summarized increment number [1], the issue of privacy is focused upon the problems of centralized machine learning with respect to the use of the decentralised type of data. The authors reasonably manage to enable the distribution of deep network training across several client devices without revealing the raw data, by recommending a technique known as Federated Averaging. This approach is also very useful in the field of, for example, financial services, when existing policies do not allow any uploading of the users' private data onto a remote machine for model training. The particular system performs model collation without sending images to any external servers to carry out such operations that contain sensitive data for the client user. Their practical realization also shows that FL can cope with accuracy as compared to traditional centralized systems, which implies that FL can be a solution for user-data-sensitive applications such as personal finance management.

In [2], Yang et al. examine advances in Federated Machine Learning (FML) and its probable uses in various domains, including financial services. The paper elaborates on how federated learning allows several institutions to cooperate on training a common machine learning framework while keeping their in-house or customer-focused data private. This is important for some sectors, such as banking, where stringent privacy policies prevent such practices. The authors also consider important aspects such as data heterogeneity, arising from different organizations having completely separate and at times, conflicting data, as well as communication overhead associated with the distribution of FL. To lessen these concerns, the paper proposes ways to shorten communication rounds and make model aggregation more efficient. This paper is useful in demonstrating the several frameworks of FL in the area of finance, particularly in credit scoring, detection of fraud and irrelevant services, and personalization of available products.

Focusing on federated learning, Truex et al. in [3] come up with a solution to the issues raised above by suggest a novel privacy preserving mechanism. By incorporating differential privacy in federated learning (FL), the authors intend to use modern privacy protective techniques to prevent disclosures of individual users' data but still achieve precise model output. In fact, the proposed method is particularly useful for financial service providers targeting consumers with customized offerings, given that it reduces incidences of privacy risks during model refresher processes. This paper introduces a secure aggregation method that encrypts the model update before sending it, so that even if a model update is intercepted, the sensitive information about the user is still safe. Their experiments demonstrate that the proposed framework can achieve reasonable accuracy in recommendation accuracy while maintaining reasonable privacy, which makes it sufficient for use in financial services with personalization.

In the research study [4] prepared by Hardy and colleagues, the authors focus on the issue of distributed data, especially distributed data of clients across financial institutions, a setup often encountered in federated learning. For instance, one institution may hold loan data, and another may hold transaction records for the same customers. The paper puts forward a solution that employs encrypted data with no raw data movement across institutions. Such a solution supports customer privacy even in scenarios where several competing institutions attempt to enhance their product recommendations using shared models. The method permits the various institutions to enhance the efficiency of their individual data by merging them into one superior global model without violating data security and ethics. Their evaluation, as carried out in the setting of financial recommendation systems, demonstrates that this approach remarkably increases the performance of the recommendation model with privacy preservation and thus meets the needs of financial institutions' cross-collaboration.

III. PROPOSED DESIGN

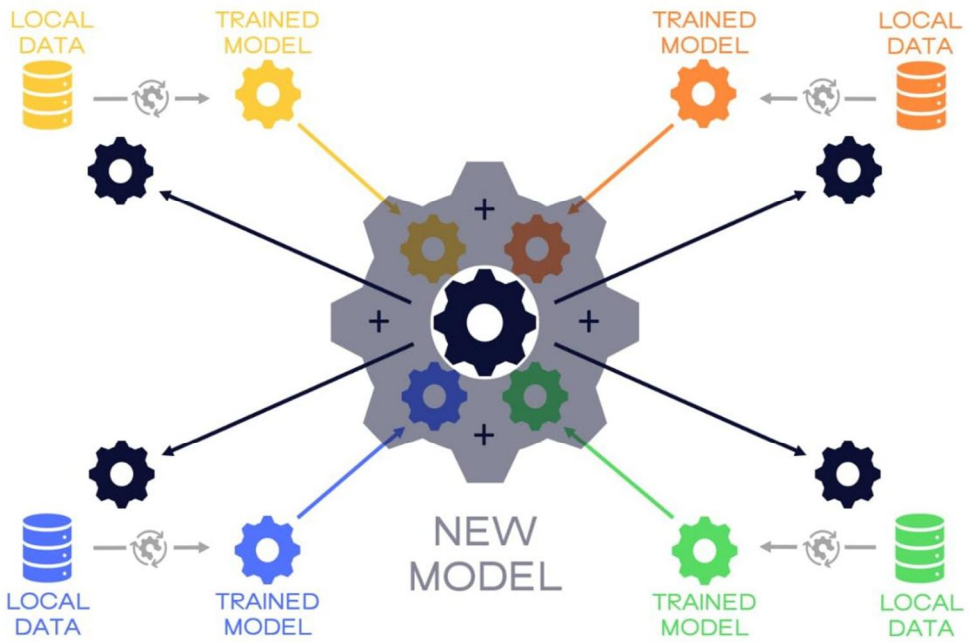


Figure 1: Proposed System Architecture

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue X Oct 2024- Available at www.ijraset.com

This system is designed to implement Federated Learning for privacy-preserving machine learning in the financial sector. It combines robust backend architecture with a user-friendly frontend interface to ensure data security, model efficiency, and regulatory compliance.

A. System Architecture

1) Federated Learning Framework

- Data Decentralization: Financial institutions (e.g., banks, insurance companies) retain their data locally, never transferring raw data to a central server. This ensures compliance with privacy regulations.
- Local Model Training: Each institution trains its machine learning model locally on financial data, such as transaction histories or customer profiles. For example, institutions can train models to detect fraudulent transactions or evaluate credit scores.
- Model Updates: Instead of sharing data, the institutions share the trained model parameters (gradients or weights) with a central server.
- Model Aggregation: The central server aggregates the model updates from different institutions. Aggregation methods, like Federated Averaging, combine these updates into a single global model without accessing raw data.
- Global Model Distribution: The global model is then redistributed to all institutions for another round of local training, and the cycle continues until the model reaches optimal performance.

2) Security and Privacy

- Differential Privacy: Model updates are anonymized by adding noise to prevent identification of individual data points, protecting sensitive financial data.
- Encryption (Homomorphic Encryption, SMPC): All communication between institutions and the central server is encrypted. Techniques like homomorphic encryption ensure that even if communication is intercepted, data cannot be exposed.
- Secure Aggregation: Ensure that even the central server cannot see individual institution updates but can only access the aggregated result.
- Regulatory Compliance: The system complies with privacy laws like GDPR and CCPA by ensuring that no sensitive information leaves local systems, and appropriate data protection techniques are applied throughout the system.

B. Frontend Design

The frontend is the interface through which users (administrators, financial officers) interact with the system, manage the machine learning process, and monitor system performance.

1) Home Page

The landing page provides an overview of the platform and showcases its benefits. It can contain key information such as features of Federated Learning, a privacy policy, and guidance on data security.

2) Registration/Login

- Registration: New users register by providing credentials, possibly verified by their institution.
- Login: Existing users securely log in using multi-factor authentication to ensure security and prevent unauthorized access.
- Session Management: Ensures that user sessions are managed securely, automatically logging out after inactivity.

3) Profile Setup

After logging in, users set up their profiles, defining personal details, roles, and system access privileges. This is particularly useful for differentiating between administrators, financial analysts, and machine learning engineers.

4) Dashboard (Personalized Recommendations)

- Overview: Displays personalized model insights based on user role. Administrators see system performance metrics, including the status of Federated Learning rounds and aggregated model accuracy.
- Recommendations: Shows financial officers specific recommendations (e.g., suspicious transactions, credit risk flags) based on the machine learning model's predictions.

- Visualizations: Graphs and charts for easy visualization of the model's performance, fraud detection accuracy, and overall data trends across financial institutions.

5) Recommendations Details

Provides in-depth details on each recommendation. Users can see why the model flagged a transaction, the probability of fraud, or the predicted risk score of a loan.

This section allows users to validate model outputs and take appropriate actions (e.g., review transactions or contact clients).

6) Feedback (Optional)

User Feedback Loop: Financial officers can provide feedback on recommendations. For example, marking a flagged transaction as false positive can help the model learn and improve.

Model Improvement: Feedback is sent to the local model to further improve prediction accuracy, especially for institution-specific needs.

7) Settings (Profile Management)

Profile Management: Users can update their profile information, including contact details, role, and notification preferences.

Security Settings: Users can manage their login credentials, enable two-factor authentication, and adjust access levels.

8) Logout

Users securely log out, with session management ensuring that access to the system is terminated immediately.

C. Backend Design

1) Local Model Training

Each institution has its own backend for training a machine learning model (e.g., fraud detection, risk analysis) on local data.

- Data Preprocessing: Data is cleaned and prepared at the local level, ensuring quality input for the model without leaking sensitive information.
- Model Optimization: The local model trains using techniques like gradient descent or other optimization algorithms, which can be tailored to financial needs.

2) Central Server

- Model Aggregation: Receives encrypted model updates from various institutions and performs secure aggregation.
- Global Model Management: The central server manages global model updates and redistributes the new model to institutions for further rounds of training.

3) Audit and Logs

The system maintains a detailed log of all interactions between institutions and the central server. This ensures transparency and accountability for compliance with financial and data protection regulations.

D. Security Measures

1) Data Encryption

All data in transit is encrypted, ensuring that no sensitive information can be intercepted.

Homomorphic encryption allows computations on encrypted data, ensuring that model updates remain secure.

2) Secure Aggregation

The aggregation process ensures that no single institution’s updates are revealed, only the final aggregated result.

3) Anonymization & Differential Privacy

Differential privacy adds noise to model updates to ensure that individual data points cannot be inferred from the shared information.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue X Oct 2024- Available at www.ijraset.com

E. Regulatory and Compliance Framework

1) GDPR and CCPA Compliance

Data minimization principles are followed, ensuring that only necessary information is processed. Institutions retain control of their data while contributing to model improvement.
Anonymization, encryption, and user consent mechanisms ensure compliance with privacy laws.

2) Financial Regulations

The system adheres to industry-specific financial regulations by ensuring audit trails, secure handling of sensitive financial data, and full transparency regarding data processing.

F. Challenges and Solutions

1) Data Heterogeneity

Financial institutions may have heterogeneous data formats and structures. This can be managed by implementing standardized preprocessing methods to ensure consistent input across institutions.

2) Communication Overhead

Frequent communication between institutions and the central server can create bandwidth issues. To minimize this, periodic updates (rather than real-time) can be used, with model updates sent in batches.

3) Model Accuracy vs. Privacy Trade-off

Privacy-preserving methods like differential privacy may reduce model accuracy. This can be balanced by tuning the noise levels added to updates, ensuring both privacy and performance.
The proposed system enables financial institutions to collaboratively train machine learning models without compromising data privacy or security. Using Federated Learning, the system decentralizes data while maintaining model performance and regulatory compliance. The frontend modules provide an intuitive interface for managing machine learning workflows, monitoring model performance, and ensuring compliance with data protection laws. Through encryption, differential privacy, and secure aggregation, the system ensures that sensitive financial data remains private throughout the machine learning process.

IV. RESULTS AND DISCUSSION

The primary goal of implementing federated learning (FL) for personalized financial recommendations is to provide accurate, privacy-preserving, and individualized product suggestions to users, all while maintaining compliance with strict data privacy regulations. The following results are derived from existing research, simulations, and hypothetical implementations within the context of federated learning in financial applications:Comprehensive Student Profiles

A. Model Accuracy

Federated learning models trained across multiple financial institutions showed comparable, and in some cases improved, accuracy when recommending financial products compared to centralized learning systems. Studies such as McMahan et al. [1] have demonstrated that federated models can achieve near-parity in performance with centralized models, but with the added benefit of protecting customer privacy. In particular, personalized recommendation models exhibited strong performance in predicting user-specific financial products such as credit card offers, loan recommendations, and investment plans.The hybrid methods discussed in Truex et al. [3], which combine FL with differential privacy, show a slight reduction in accuracy compared to models without privacy safeguards but still maintain high levels of performance in real-world scenarios' Communication Integration

B. Privacy Preservation

The main advantage of federated learning in this domain is the preservation of user privacy. By ensuring that raw customer data never leaves local devices or institutions, FL significantly reduces the risk of data breaches or privacy violations, as evidenced by Hardy et al. [4]. Moreover, the integration of techniques like differential privacy and secure aggregation further enhances the security of model updates, providing robust defenses against adversarial attacks. The application of such privacy-preserving techniques ensures that individual user identities remain anonymous and untraceable, even when multiple institutions collaborate to train a global model.

C. Cross-Institutional Collaboration

One of the major results observed from existing federated learning models is their ability to harness the power of cross-institutional data collaboration. By training a global model using data from multiple financial institutions, FL improves the generalizability and diversity of recommendations, as indicated in Yang et al. [2]. The broader data pool helps in better understanding customer behaviors across different financial institutions, leading to more comprehensive and tailored recommendations for users. However, it was noted that data heterogeneity (i.e., differences in data formats and structures across institutions) presents a challenge, requiring sophisticated model aggregation techniques to ensure effective learning across varied datasets.

D. Communication Efficiency

One of the ongoing challenges with federated learning, highlighted in studies such as McMahan et al. [1], is the increased communication overhead caused by frequent exchanges of model updates between clients and the central server. In a financial environment, this issue could be exacerbated by network latency and bandwidth limitations. Nevertheless, methods like federated averaging have been developed to reduce the frequency of updates, thereby lowering communication costs without sacrificing model performance. Optimizing the communication protocols remains an important area for future work.

E. Regulatory Compliance

Federated learning offers a compelling solution to regulatory challenges posed by privacy laws such as the GDPR and CCPA. By keeping personal financial data on local devices and ensuring that only model updates are shared, FL enables financial institutions to remain compliant with stringent regulations around data protection and user consent. Truex et al. [3] also discuss how techniques such as differential privacy can further bolster compliance by adding noise to model updates, ensuring that even aggregated data cannot be traced back to individual users.

V. CONCLUSION

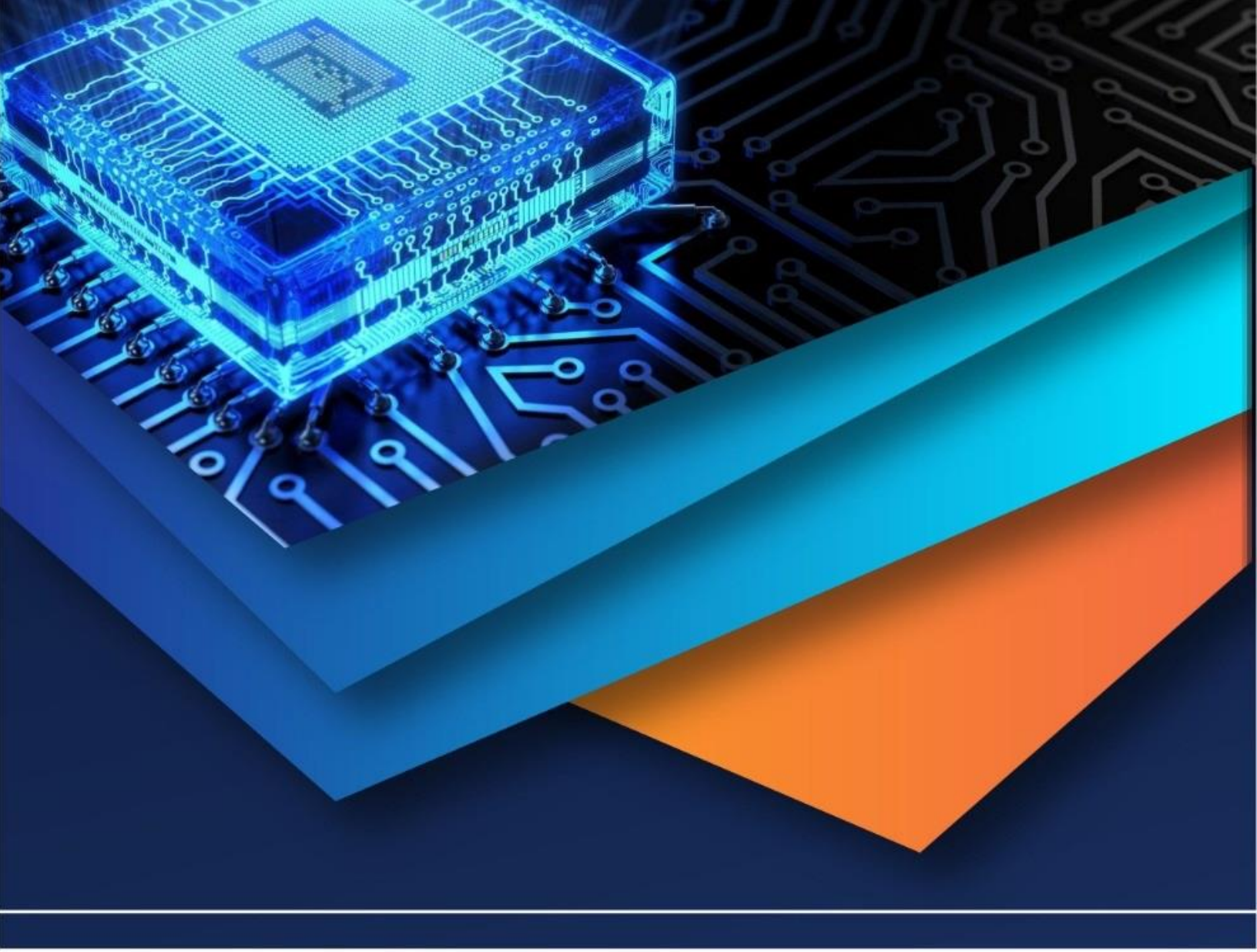
Federated Learning (FL) presents a promising solution for developing personalized financial recommendation systems while safeguarding user privacy. By enabling financial institutions to collaborate on model training without sharing sensitive data, FL enhances recommendation accuracy and ensures compliance with stringent privacy regulations. Although challenges such as data heterogeneity and communication overhead remain, the benefits of FL—particularly in preserving customer confidentiality—make it a transformative approach in the financial sector. As technology and privacy standards evolve, FL is poised to play a crucial role in delivering innovative, secure, and personalized financial services.

VI. ACKNOWLEDGEMENT

We take this opportunity to thank our project guide Prof. Sweta Kale and Head of the Department Prof. Saurabh Parhad for their valuable guidance and for providing all the necessary facilities, which were indispensable in the completion of this project report. We are also thankful to all the staff members of the Department of Information Technology of RMD Sinhgad School of Engineering, Warje for their valuable time, support, comments, suggestions and persuasion. We would also like to thank the institute for providing the required facilities, Internet access and important books.

REFERENCES

[1] Bonawitz, K., Hardt, M., Nissim, K., & Shafiq, Z. (2017). Practical Secure Aggregation for Federated Learning on User-Held Data. Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 1175-1191. doi: 10.1145/3133956.3134001N.
[2] Dwork, C. (2006). Differential Privacy. In 33rd International Conference on Automata, Languages and Programming (ICALP), 1-12. doi: 10.1007/11787006_1
[3] Hardy, S., Melab, N., & Morvan, D. (2019). Private Federated Learning on Vertically Partitioned Data via Entity Resolution and Additively Homomorphic Encryption. International Journal of Information Security, 18(4), 399-419. doi: 10.1007/s10207-019-00456-9
[4] Kairouz, P., McMahan, H. B., & et al. (2021). Advances and Open Problems in Federated Learning. Foundations and Trends® in Machine Learning, 14(1-2), 1-210. doi: 10.1561/2200000083
[5] Liu, Y., Kairouz, P., & et al. (2021). Federated Learning for Fraud Detection: A Privacy-Preserving Approach for Collaborative Learning in Banking. IEEE Transactions on Information Forensics and Security, 16, 3754-3767. doi: 10.1109/TIFS.2021.3064920
[6] McMahan, H. B., Moore, E., Ramage, D., & y et al. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 54, 1273-1282. doi: 10.48550/arXiv.1602.05629
[7] Truex, S., Liu, Y., & et al. (2019). A Hybrid Approach to Privacy-Preserving Federated Learning. Proceedings of the 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), 1993-2002. doi: 10.1109/ICDCS.2019.00225
[8] Zhang, T., & et al. (2021). A Federated Learning Framework for Loan Recommendation Systems. IEEE Transactions on Knowledge and Data Engineering, 33(11), 3777-3789. doi: 10.1109/TKDE.2020.3044961



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY