# Federated Learning for Privacy Preserving Financial Data Sharing

**Ajay Benadict Antony Raju**

ajaybenadict@gmail.com

**Abstract**

**With the growth of digitalization, most of the financial institutions are focused on the best use of information intelligence for better strategic planning and performance improvement. However, due to the confidential content of financal information, or data, privacy is a major issue of concern and hence effective measures to protect the information while trying to use it should be developed. Thus, Federated Learning (FL) as the approach to share information between the institutions without disclosing sensitive data appears to be quite helpful to overcome these challenges.**

**Federated Learning actually helps in group learning but at the same time, no data is shared with each other. FL also differs from the centralised data storage where a giant financial database is created, but FL enables different institutions train a common model locally on their data and only the gradients or updates of the model are shared among the participating institutions. This approach helps manage risks of the unauthorized access on the data as well as assists in the use of multiple sources of financial data for analysis.**

**This paper focuses on Federated Learning as the method of implementing data sharing in the financial industry with a view of examining how this learning process can transform privacy preservation practices. We address the fundamental concepts of FL and its application in financial environments as well as the advantages, such as increased protection and data privacy compliance. Furthermore, we discuss issues including model accuracy, communication complexity, and the requirement for safe three-way aggregation. With FL, it is possible to incorporate it into the frameworks involved in sharing of financial data, institutions can strengthen their analytical processes incrementally from other institutions while observing the best practices in data privacy.**

**Based on our work, Federated Learning appears as a promising improvement in handling data utility while being more sensitive to privacy issues, which will eventually lead to even more efficient and secure methods of sharing financial data.**

**Keywords: Federated Learning, Privacy-preserving, Financial Data, Data Sharing, Machine Learning, Data Security**

## Introduction

In the current world of complex financial environments, the role of data is indispensable as it fuels decision making, enriches customer experiences and optimizes processes. But with the advancements in the information collection and analysis processes, the issue of confidentiality and security of the data have also assumed importance. Every company deals with large volumes of highly valuable and targeted information such as history of transactions, credit records, and identification details. Therefore, there is a gradual shift to methods that maintain the individuals' privacy during analysis while analyzing data has benefits.

Technique that is becoming a revolutionary solution for this challenge is called Federated Learning (FL). Unlike the conventional machine learning where the raw data is collected at a central place for training the

model, Federated Learning is an effective process through which multiple institutions build a common machine learning model without the need of sharing the data. This solves the issue of privacy in a way that of no information leaves the institution, hence minimizing the instances of leaks and unauthorised access.

The concept of Federated Learning emanates to the idea that uncovers learning mechanisms to support data-driven decisions without the invasion of privacy. In a federated network, each participant trains the model on the data available on his end and sends updated model to the central server instead of the data. This method helps to maintain the confidentiality of financial information and at the same time make a valuable input into the creation of an effective and efficient international mod allowed Pell model.

Nonetheless, Federated Learning in the financial sector has its pitfalls even though it holds a lot of potential. Questions like how to make a model accurate for different inputs, how efficient the communication channels are between institutions and how to protect the accrued models from potential attack are issues that needs to be solved. Also, strict adherence to data protection laws brings more challenges on the same.

This introductory section discusses how Federated Learning must be incorporated into finance data-sharing paradigms to address the necessity of using the data for analytical gains while still preserving users' privacy. And as institutions for finance weave through the challenges of data protection, Federated Learning represents a revolutionary development in safe and valuable data sharing.

**Literature Review:**

Fl has received considerable attention especially when applied for financial data sharing because of the promise it offers in achieving both privacy and data usefulness. FL is agnostic and decentralized method of training an ML model that is both fed with data by several local institutions while the data remain local. This method is advantageous where the data is sensitive or has to undergo strict security measures like in the fields of finance. When McMahan et al. first proposed Federated Averaging in 2017, it laid down the FL as a framework that can allow the participants' training of a shared model without compromising the privacy of each participant's training data while at the same time preserving the quality of the resultant trained model [3].

Subsequent developments in FL have responded to diverse difficulties that are peculiar to the financial domain. For example, Bonawitz et al. (2019) looked into how updates and the sharing of data can be conducted securely such that there is no leakage of information This is important in the case of model updates [2]. Likewise, Zhang et al. (2020 also provided methods for dealing with the non-IID data, which are prevalent in financial datasets because of their heterogeneity [3]. These methods improve the stability of FL in real-world applicability.

However, there are still several issues which are worth discussing. With reference to the FL detriment, Liu et al. (2021) noted obstacles concerning communication overhead and synchronization delays will affect the FL financial systems [3]. Moreover, Wang et al., (2022) also pointed out the lack of efficient schemes in terms of privacy preservation where the authors specified on the differential privacy methods that can enhance the protection strategies consequently for the financial data [5]. Another research direction that emerged in combining Federated Learning with fine technical protocols including cryptographic methods briefly described in article of Yang et al. (2023) is to enhance the security of federated models [6].

These studies together convey the role of FL in transforming financial data sharing and at the same point depict about the research gaps to work upon.

**Problem Statement**

The financial industry requires the use of data to enhance the competitive edge but at the same time there's a need to protect individual privacy. Old-style centralized data sharing pose financial information to a lot of risk such as hacking and unauthorized access [7]. Adding to this difficulty is this growing enforcement of the

General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) to protect private data [8][9].

That's where Federated Learning comes into play as it provides an opportunity for institutions to train models together without sharing the raw data. Nonetheless, there are some difficulties inherent in the use of FL in financial context. Firstly, great efforts have been made to distill knowledge in a decentralized and heterogeneous environment because the existing models' accuracy and effectiveness are critical. It is imperative to note that most often financial data is non-IID or non-stationary which might be instrumental in influencing the results the model as well as its portability [3]. Second, the communication efficiency between the institutions is important; when models are frequently updated and synchronized, high computational and communication costs are inevitable [4]. Last of all, as FL brings enhancements to privacy by design, there is a need to address the weaknesses that come with the model aggregation and to meet new privacy regulations [10][6].

The solution to some these problems is therefore paramount for the proper implementation of Federated Learning in the sharing of financial data. As such, model accuracy has to be improved, communication costs have to be minimized and privacy protection measures have to be strengthened to fully unlock FL's potential for privacy-preserving learning [11][3].

**Solution**

The existing problems of FL in financial data sharing require a multimodal solution that will increase the accuracy of the resulting model, optimize communication, and strengthen the protection of information.

Enhancing the model performance is crucial because of the non-independent and identically distributed (non—IID) characteristics of the data used in the financial applications, which could lead to suboptimal model and reduced generality. Thus, the use of federated transfer learning is possible to prevent this problem. This technique entails the use of models that have to be pre-trained on similar tasks or data before they are federated fine-tuned to enhance performance on various financial data sets [4]. Furthermore, mobbing mechanisms for aggregating the model updated have FedAvg, which can be further enhanced by integrating adaptive sampling, control of local epochs, and extra centralised steps [12]. These approaches assist in attaining a better, and more accurate model even with the complexities that are inherent with uncentralized and varied data.

Improving the efficiency of communication is also an important aspect because whenever there are updates and syncs to the model and/or parameters, it incurs a lot of computation and network load. Some of the measures like the model compression and update aggregation can however help to reduce this burden basically by lowering the amount of information that needs to be exchanged across the network. Techniques such as quantization and pruning reduce the size of the model updates being exchanged between institutions, and hence relieves the communication pressure [30]. In addition, asynchronous federated learning methods enable institutions to update the global model offline, thus eliminating the need for frequent synchronization which lowers system time [14]. Therefore, the following strategies can be employed in order to address the above communication overhead to enhance federated learning: By doing so, the complexity contributed by the communication overhead can be reduced hence improving the performance of federated learning.

Improving the Privacy of Models and Their Combined Outcomes is necessary to eliminate certain risks and follow the requirements of data protection laws. Federated Learning affords inherent privacy gains by centralizing the data; however, more privacy preservation is required. There are technologies like differential privacy that can be incorporated in the same way and prevent individual data contributions from being exposed by adding noise to the model update [10]. There are two additional methods: The first one is secure multi-party computation (MPC), which allows aggregating the model updates without leaking any sensitive information [15]. These combined with the legal regulations concerning data protection like GDPR and CCPA guarantee the protection of the financial data in the course of federated learning [8].

This solution is design to address the problem of Federated Learning in the sharing of financial data by addressing the 3 crucial aspects model accuracy communication efficiency and data privacy. Future work in these domains will be important for refining the impact of and adoption of Federated Learning while preserving customer privacy.

**Conclusion**

FL is certainly one of the most revolutionary novelties in the area of privacy-preserving FinTech information exchange. By enabling institutions to collaboratively train machine learning models without centralizing sensitive data, FL addresses a critical challenge in the financial sector: This is due to the conflicting demands being made of information gain on the one hand, and strict privacy limitations on the other. It reduces the vulnerabilities inherent with improper data sharing and protects individuals' rights, making it a safer and more efficient way to approach data-sharing than place it in a centralized hub.

As one of the major advantages of Federated Learning, it is possible to state the work with non-homogeneous and spread-out datasets and the required protection of data sources. This is especially important in the financial industry since data is non- IID and is sensitive by nature. Several approaches regarding FL have shown that it can be capable of addressing the challenges presented by heterogeneous financial data, including federated transfer learning and other forms of aggregation learning on the device. These advancements help make the models federated more relevant and easily applicable across institutions to provide a reliable fedarated base to improve decision making and other operations.

Although the concept of FL is well demonstrated in financial environment, the achievement of its implementation is not without some certain forms of challenges. However, communication overhead is still a big issue because of the huge load of data that is required to be transferred for model update and synchronization. Solving this problem through model compression and asynchronous learning methodology has a tendency of making the federated systems more feasible and sustainable due to the reduction of computational and network loads. However, there is the need to improve the privacy protection through the use of categories such as differential privacy and secure multi-party computation to protect sensitive data during the model aggregation processes and to meet new data protection laws.

The adoption of Federated Learning in the financial data-sharing frameworks will have a transformative impact on institutions' cooperation and data leverage. It provides the course to follow which complies with strict privacy regulations and, at the same time, opens more horizons for progress and optimization. With a growing awareness among financial institutions on the importance of employing data as a source of competitive advantage, use of FL can play a crucial role in offering such an advantage particularly in providing a secure means of accomplishing collaborative analytics.

More efforts in research and development of Federated Learning methodologies will be required to overcome these current issues and extend the spectrum of FL's use. Flawless advancements in model precision, operational communication, as well as development in preservation of individual privacy will shape the next future progression of FL to be a stronger, more efficient solution for decentralized sharing of data. It is clear that if these implementations are adopted the financial institutions will be able to incorporate beneficial data metrics without compromising the confidentiality of the data.

In conclusion, it is possible to state that Federated Learning is a breakthrough as for sharing financial data. Due to its capability to safely and securely manage important data and allow model training with sensitive information shared among the collaborative participants, it marks an improvement to some of the grievances that affect privacy of data and data-based decisions. The announced technology improvements should make the financial sector's approach even more profound, creating unprecedented opportunities for effective and safe use of data.

**References**


1. McMahan, B. et al. (2017). "Communication-Efficient Learning of Deep Networks from Decentralized Data."
2. Bonawitz, K. et al. (2019). "Towards Federated Learning at Scale: System Design."
3. Zhang, Y. et al. (2020). "Federated Learning with Non-IID Data."
4. Liu, Y. et al. (2021). "Challenges and Solutions in Federated Learning Communication Efficiency."
5. Wang, J. et al. (2022). "Enhancing Privacy in Federated Learning with Differential Privacy."
6. Yang, L. et al. (2023). "Cryptographic Enhancements in Federated Learning."
7. Alavi, M., & Leidner, D. E. (2001). "Reviewing the Past, Present, and Future of Information Systems Research." *MIS Quarterly*, 25(3), 431-459.
8. Voigt, P., & Von dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR)*. Springer.
9. California Consumer Privacy Act of 2018. *Cal. Civ. Code 1798.100 et seq.*
10. Dwork, C. (2008). "Differential Privacy: A Survey of Results."
11. Kairouz, P. et al. (2019). "Advances and Open Problems in Federated Learning."
12. Smith, V. et al. (2018). "Robust Federated Averaging Algorithms."
13. Lin, J. et al. (2018). "Model Compression Techniques for Federated Learning."
14. Acar, A. et al. (2021). "Asynchronous Federated Learning Approaches."
15. Lindell, Y., & Pinkas, B. (2009). "Secure Multiparty Computation for Privacy-Preserving Data Analysis."