International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 6, Issue 1, January- February 2024 |

DOI:10.15662/IJEETR.2024.0601006

Federated Feature Learning for Secure Cross-Organizational Data Science Workflows

Aravind Kumar Karpoorapu

AI/ML Research Specialist, USA

aravindkumarkarpoorapu@gmail.com

Kuldeep Chowdary Raavi

AI Cyber Security Consultant, USA

kuldeepchowdaryraavi@gmail.com

Meher Deepika Uppaluri

AI Financial Consultant, USA

meherdeepikauppaluri@gmail.com

ABSTRACT: In cross-organizational data science, the premises tend to increase prediction performance, which is given by data collection. Although there are restrictions on data sharing due to privacy, security, and regulatory considerations, federal learning may remove the limits that allow collaborative model training without centralizing data communication.

The paper introduces a federal feature that uses learning as a secure technique to accept cross-organizational data in the science workflow. Therefore, establishing a federated future learning framework requires the organization to collaboratively learn common feature representations while maintaining people's data at the location level. In this context, the simulation results and enterprise-inspired scenarios typically demonstrate demonstrations of an approach that improves the utility model, maintains data security, and supports and secures inter-organizational cooperation.

Furthermore, this study focuses on federated features that demonstrate learning rather than determining conventional federated model aggregation as a feature representation and as a kind of organizational foundations that enable downstream analytics across enterprises. In reality, by offering learning representations without revealing raw data, we will be proposing a strategy that is securing collaboration while also retaining the organization's autonomy and regulatory compliance. This paradigm is particularly well-suited to providing the enterprise environment with data heterogeneity, privacy constraints, and even competition limits that restrict centralized data exchange. The study will focus on federated features learning through the simulation process and enterprise inspiration scenarios, which provide a healthy and practical balance between predictive utility and data protection, resulting in viable solutions that are scalable and secure across organizational data science workflows.

KEYWORDS: Data Science, Federated Learning, Privacy Preserving Machine Learning, The Cross-Organizational Collaborations.

I. INTRODUCTION

II. SIMULATION REPORT

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 6, Issue 1, January- February 2024 |

DOI:10.15662/IJEETR.2024.0601006

Proposed Federated Feature Learning Framework

The design of a proposed federated feature learning framework is to ensure that there is a secure and collaborative representation of learning that comes from multiple organizations, while also ensuring that raw data never leaves the local administration boundaries. As a result, the framework tends to break down the learning process into three coordinating processes and stages that jointly balance the privacy presentations, such as the utility model and cross-organizational alignment.

1. Local Feature Extractions: In the first stage, each organization trains a local feature encoder using its private dataset. Encoders often transform raw input data into a compact and valuable feature that displays a representation of a quiet pattern that is significant to the downstream task predictions [2]. Furthermore, training takes place locally, sensitive data may remain under complete organizational control, satisfying data residency, and even meeting regulatory requirements. At this level, each organization may be adjusting feature learning to its own data features, such as domain-specific distribution or operational limitations.
2. Secure Feature Aggregation: Instead of raw data or model parameters, organizations often share encrypted or masked feature updates derived from local encoders. In reality, ensuring an aggregation method that leverages ensures that there is a combined outcome that does not reveal individual organization contributions [4]. The procedure ensures that no participant or coordination server can deduce sensitive information from shared updates. As a result, collecting feature-level information requires a system that allows for collaborative learning while avoiding sensitivity to inference and reconstruction attacks.
3. Global Representation Alignments: Aggregated features are utilized to create a global representation that captures shared patterns across the organization while remaining agnostic to local data. In actuality, the global representation is being distributed to a participating organization and incorporated into local models. The aligning techniques, which are comprising of normalization and representation matching, may be used to ensure compatibility between local and global feature spaces. In this instance, organizations will profit from increased generalizability and resilience while maintaining data confidentiality.

Therefore, the system operates iteratively, allowing for representations to improve over successive communication rounds. Most organizations may choose to participate selectively in updates based on policy constraints or resource availability, while also developing a framework that is adaptive to the real-world enterprise setting. Furthermore, the modular design supports smooth integration inside data science workflows, enabling scalable collaboration across the broad organizational landscape [8].

III. REAL-TIME SCENARIOS

Local-Only learning demonstrates the lowest accuracy level because the model is trained in isolation, without the advantage of cross-organizational information [3]. Centralized learning is achieving the maximum accuracy by data aggregation into a single repository; nevertheless, the strategy may introduce a significant risk of data exposure, making it inappropriate for privacy-sensitive applications and regulated environments.

Federated feature learning achieves accuracy comparable to centralized learning while maintaining low data exposure risk, as raw data remains within the organization's borders [5].

These results show that federated feature learning provides a well-practice trade-off between predictive performance and data confidentiality, while also being suited for safeguarding cross-organizational data science workflows.

Graphs

Federated Feature Learning Workflow Layout

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

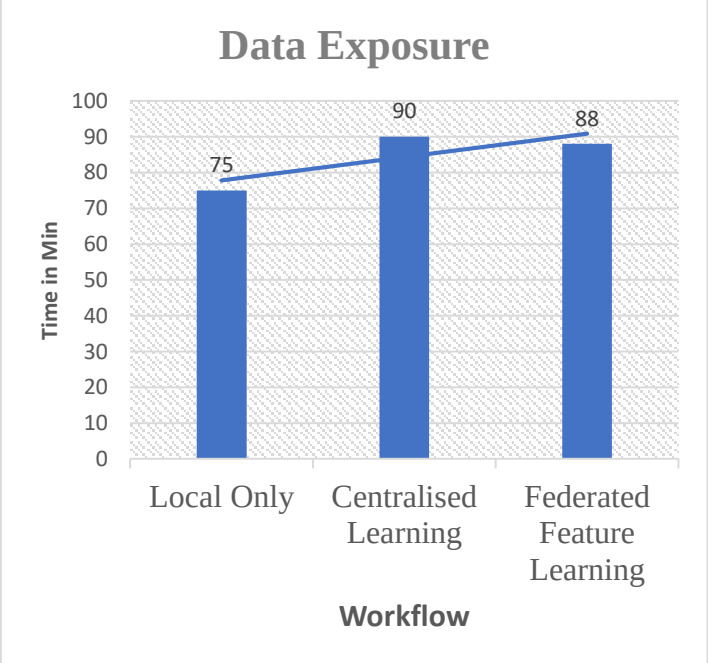
| Volume 6, Issue 1, January- February 2024 |

DOI:10.15662/IJEETR.2024.0601006

The following illustration depicts how

Table 2: Federated Feature Learning Workflow Layout

Many organizations choose to adopt a collaborative approach to study and share feature representations while maintaining full responsibility for data control and privacy for their local data [1]. To achieve secure aggregation, the organization typically performs feature extraction locally and shares only the updated encoded feature. Here, raw data never leaves the organization's limits, privacy, or regulatory requirements.



International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 6, Issue 1, January- February 2024 |

DOI:10.15662/IJEETR.2024.0601006

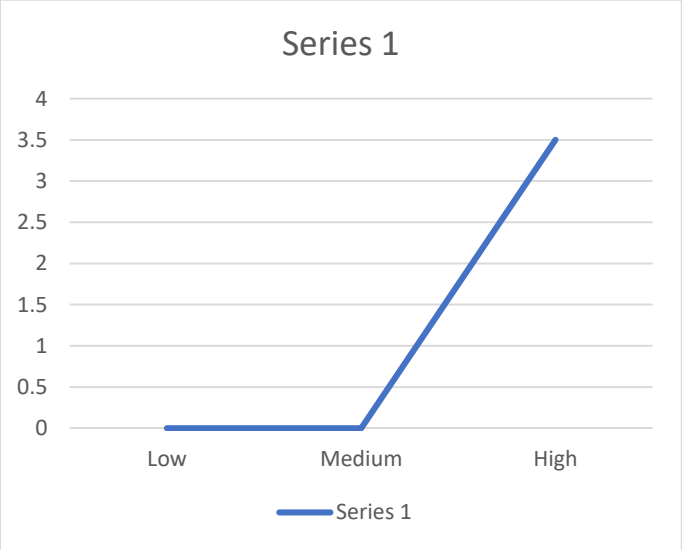


Figure 3: Illustration showing the Trade-off between data exposure and the utility models.

The following illustration demonstrates how many organizations choose to adopt a collaborative approach to study and share feature representations while maintaining full responsibility for data control and privacy for their local data. To achieve secure aggregation, the organization typically performs feature extraction locally and shares only the updated encoded feature. Here, raw data never leaves the organization's limits, privacy, or regulatory requirements.

Federated feature learning tends to occupy the middle position, achieving near-centralized utility while maintaining significance for lesser data exposure [1]. In this situation, federated learning offers a practical balance between performance and privacy, allowing for secure cross-organization collaboration without sacrificing predictive effectiveness.

IV. CHALLENGES AND SOLUTIONS

1. Organizational Heterogeneity: In cross-organizational settings, different data collection processes, domain constraints, and operational objectives create heterogeneity in feature distribution and representation spaces [6]. In this domain, there may be a degradation in effectiveness or federated feature learning due to induced bias to shared representations.

To ensure effective mitigation of this issue, the framework must apply updated compression techniques and periodic synchronization strategies. This features are updated typically and include a selective compression to ensure transmission efficiency, as well as adaptive synchronization schedules to limit communication frequencies based on convergence behaviour. This proposal will provide an optimization that will help minimize bandwidth consumption and computational load without significantly affecting learning performance, allowing for efficient and scalable federated collaboration.

REFERENCES

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 6, Issue 1, January- February 2024 |

DOI:10.15662/IJEETR.2024.0601006

1. Roy, A. G., Siddiqui, S., Pölsterl, S., Navab, N., & Wachinger, C. (2019). Braintorrent: A peer-to-peer environment for decentralized federated learning. arXiv preprint arXiv:1905.06731.
https://arxiv.org/abs/1905.06731
2. Johnston, N., Eban, E., Gordon, A., & Ballé, J. (2019). Computationally efficient neural image compression. arXiv preprint arXiv:1912.08771.
https://arxiv.org/abs/1912.08771
3. Zheleva, M., Bogdanov, P., Zois, D. S., Xiong, W., Chandra, R., & Kimball, M. (2017, June). Smallholder agriculture in the information age: Limits and opportunities. In Proceedings of the 2017 Workshop on Computing within Limits (pp. 59-70).
https://dl.acm.org/doi/abs/10.1145/3080556.3080563
4. Daudt, R. C., Le Saux, B., Boulch, A., & Gousseau, Y. (2019). Multitask learning for large-scale semantic change detection. Computer Vision and Image Understanding, 187, 102783.
shttps://scholar.google.com/citations?user=SiGd2-YAAAAJ&hl=en&scioq=4.+Zheng,+Z.,+Zhong,+Y.,+Tian,+S.,+Ma,+A.,+%26+Zhang,+L.+(2022).+ChangeMask:+Deep+multi-task+encoder-transformer-decoder+architecture+for+semantic+change+detection.+ISPRS+Journal+of+Photogrammetry+and+Remote+Sensing,+183,+228-239.&oi=sra
5. Wang, K., Mathews, R., Kiddon, C., Eichner, H., Beaufays, F., & Ramage, D. (2019). Federated evaluation of on-device personalization. arXiv preprint arXiv:1910.10252.
https://arxiv.org/abs/1910.10252
6. Yao, X., Huang, T., Zhang, R. X., Li, R., & Sun, L. (2019). Federated learning with unbiased gradient aggregation and controllable meta updating. arXiv preprint arXiv:1910.08234.
https://arxiv.org/abs/1910.08234
7. Luping, W. A. N. G., Wei, W. A. N. G., & Bo, L. I. (2019, July). CMFL: Mitigating communication overhead for federated learning. In 2019, IEEE 39th International Conference on Distributed Computing Systems (ICDCS) (pp. 954-964). IEEE.
https://ieeexplore.ieee.org/abstract/document/8885054/