Federated Learning Framework for Privacy-Preserving Credit Default Prediction under Dynamic Economic Regimes

Anil Natarajan
Department of Computer Science, Colorado State University, Fort Collins, CO, USA.
anilnatarajan174@colostate.edu

Jerome Marsh
Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA.
jerome.work@ku.edu

Abstract

This paper presents a comprehensive systems-level analysis of a federated learning framework designed for privacy-preserving credit default prediction under dynamic economic regimes. The proliferation of machine learning in financial risk assessment has raised significant concerns regarding data privacy, regulatory compliance, and model robustness, particularly when economic conditions shift unpredictably. We propose an architectural paradigm that integrates horizontal federated learning with differential privacy and secure aggregation to enable collaborative model training across financial institutions without exposing sensitive client data. The framework explicitly addresses the challenge of concept drift induced by changing macroeconomic environments by incorporating adaptive model recalibration mechanisms and regime detection modules. We examine structural trade-offs between privacy guarantees, model accuracy, communication efficiency, and computational overhead, drawing on cross-domain comparisons from healthcare and telecommunications. Governance considerations are emphasized, including fairness audits, bias mitigation, and accountability under evolving regulatory frameworks such as the General Data Protection Regulation and the Fair Credit Reporting Act. Infrastructure requirements for deployment at scale are discussed, including edge computing support, asynchronous updates, and energy sustainability. Through analytical discussion and illustrative case scenarios, we demonstrate that a carefully designed federated system can achieve robust, interpretable, and fair credit default predictions while preserving individual privacy. The paper concludes with forward-looking recommendations for policy design, standardization, and interdisciplinary research to operationalize privacy-preserving financial AI in volatile economic climates.

Keywords

federated learning, privacy preservation, credit default prediction, dynamic economic regimes, differential privacy, financial infrastructure, fairness, concept drift.

1. Introduction

Credit default prediction is a cornerstone of modern financial risk management, enabling lenders to assess the probability that a borrower will fail to meet contractual obligations. Traditional approaches rely on centralized collection of vast amounts of sensitive personal financial data, which creates significant privacy vulnerabilities and exposes institutions to

regulatory penalties. The emergence of federated learning as a distributed machine learning paradigm offers a promising alternative that allows multiple financial entities to collaboratively train a predictive model without sharing raw data [1]. However, the financial domain presents unique challenges not fully addressed by generic federated learning frameworks. Economic regimes are inherently dynamic, characterized by periods of expansion, recession, volatility, and structural change, all of which can alter the statistical relationships between borrower attributes and default events. A model trained during a stable period may fail catastrophically when a market downturn occurs, leading to systemic risk amplification. Furthermore, the stringent privacy requirements imposed by regulations such as the General Data Protection Regulation and the California Consumer Privacy Act demand that any distributed learning system incorporate robust privacy-preserving mechanisms, including differential privacy and secure multi-party computation [2,3]. Balancing these competing objectives—privacy, accuracy, fairness, and adaptability—requires a holistic systems design approach that extends beyond algorithmic innovation to encompass governance, infrastructure, and sustainability. This paper proposes and examines a federated learning framework tailored for credit default prediction under dynamic economic regimes, focusing on architectural decisions, trade-offs, deployment constraints, and policy implications. The analysis draws on insights from recent advances in privacy-preserving machine learning [4], a benchmark design study for economic stress early warning [5], and cross-sector experiences from healthcare and telecommunications where federated learning has been deployed in similarly sensitive and dynamic environments [6,7].

2. Background and Related Work

Federated learning was originally introduced as a means to train deep learning models on decentralized data residing on mobile devices, with the server aggregating model updates rather than raw data [1]. Since then, the paradigm has been extended to cross-silo settings where a small number of organizations participate in collaborative training, a configuration that aligns naturally with the structure of the banking industry. In financial risk modeling, several studies have explored the application of federated learning for credit scoring and fraud detection, demonstrating that federated models can achieve accuracy comparable to centralized models while reducing data exposure [8]. However, these works often assume static data distributions and do not account for the volatility of economic regimes. Concept drift is a well-studied phenomenon in online learning, but its integration with federated learning remains an active research area. Some approaches propose periodic retraining using sliding windows or ensemble methods, but these can be costly and may introduce privacy leakage if model updates reveal temporal patterns [9]. Differential privacy has emerged as the de facto standard for providing mathematical guarantees against inference attacks on model updates, with the added challenge of balancing noise injection with model utility [2,10]. Secure aggregation protocols, such as those based on secret sharing, further protect individual updates from being inspected by the server [11]. In the context of credit default prediction, fairness is a critical dimension because biased models can disproportionately deny loans to protected groups, perpetuating socioeconomic inequities. Federated learning introduces additional fairness complexities when participating institutions serve different demographic populations and data distributions are heterogeneous [12]. Recent research has proposed fairness-aware aggregation rules and adversarial debiasing techniques, but these have not been systematically evaluated under shifting economic conditions [13]. The present work builds on these foundations by proposing a framework that explicitly incorporates regime

detection, adaptive recalibration, and fairness constraints, while maintaining rigorous privacy guarantees.

3. System Architecture and Design Considerations

The proposed framework adopts a hierarchical federated architecture in which participating financial institutions each host a local data store and a local model replica, while a central coordinator orchestrates the training process. Instead of transmitting raw client records, each institution computes model updates such as gradient vectors or weight deltas, applies local differential privacy noise, and submits the perturbed updates to the coordinator. The coordinator then performs secure aggregation using a threshold-based protocol to combine updates from a quorum of institutions before generating a global model [11]. This architecture is designed to be resilient to client dropouts, which are common in real-world deployments due to network interruptions or maintenance windows. A key design choice involves the trade-off between communication efficiency and privacy loss. Reducing the frequency of communication rounds decreases bandwidth consumption but may prolong convergence and increase the total privacy budget consumed under a fixed epsilon-differential privacy framework. Adaptive communication scheduling based on model convergence metrics and economic regime indicators can mitigate this tension. For example, during periods of high volatility, more frequent updates might be necessary to capture rapidly shifting relationships, whereas stable regimes allow for sparser communication. The framework includes a dedicated regime detection module that monitors macroeconomic indicators such as interest rate changes, credit spread dynamics, and unemployment trends, and triggers recalibration events when a regime shift is detected [14]. This module operates on anonymized aggregate statistics from participating institutions, preserving privacy while enabling timely adaptation. The coordinator is also responsible for maintaining a historical repository of global model snapshots, which can be used to evaluate backtesting performance and to roll back to a stable state if a new model exhibits degradation.

4. Privacy-Preserving Mechanisms and Trade-offs

Two primary privacy preservation mechanisms are incorporated into the framework: local differential privacy and secure aggregation. Under local differential privacy, each institution adds calibrated noise to its model update before transmission, ensuring that the update does not reveal information about any individual client's data beyond a predefined privacy budget epsilon [2,10]. The choice of epsilon involves a fundamental trade-off: a lower epsilon provides stronger privacy protection but introduces greater noise, which can reduce model accuracy and slow convergence. In financial applications, the stakes are high because even modest accuracy losses can translate into significant misclassification costs. Moreover, the privacy budget must be carefully allocated across multiple communication rounds, as cumulative privacy loss under composition theorems can erode guarantees [15]. The framework employs a privacy accountant to track the total epsilon spent and to halt training when the budget is exhausted, forcing a reset of the global model or a renegotiation of privacy parameters among participants. Secure aggregation, implemented via Shamir's secret sharing or more efficient additive secret sharing, prevents the coordinator from observing any individual institution's update, even in perturbed form [11]. This adds a layer of protection against a malicious or compromised coordinator, but at the cost of increased computational overhead and communication latency. The combination of both mechanisms provides defense in depth, although it also amplifies the noise due to the interaction between differential privacy and secure aggregation's rounding errors [16]. An important governance implication

is that participating institutions may have heterogeneous privacy requirements based on local regulations or internal policies. The framework supports configurable privacy budgets per institution, allowing each entity to set its own epsilon threshold. However, heterogeneity can lead to unequal contributions to the global model, potentially biasing the model toward institutions with looser privacy constraints. Addressing this imbalance requires careful weighting of updates during aggregation, a topic that intersects with fairness.

5. Dynamic Regime Adaptation and Robustness

Economic regimes are not stationary; they evolve in response to policy changes, market shocks, and structural transformations. A credit default model trained exclusively on historical data from a growth period will likely fail to generalize during a recession because the underlying distribution of borrower risk factors shifts. The framework incorporates a two-tier adaptation mechanism. At the global level, the regime detection module continuously evaluates a set of macroeconomic time series and flags regime changes when statistically significant alterations in the joint distribution of these signals are observed [14]. Upon detection, the coordinator initiates an adaptive recalibration phase in which participating institutions perform additional local training rounds using the most recent data, while the server adjusts the aggregation weights to emphasize institutions that are experiencing the regime change most acutely. At the local level, each institution can implement an internal drift detection algorithm, such as the Page-Hinkley test or the Kolmogorov-Smirnov test, applied to the distribution of model outputs or feature residuals. If local drift is detected, the institution can request a personalized model update from the global server, effectively blending global and local knowledge [9]. This hierarchical approach balances the need for rapid adaptation with the computational and communication costs of frequent global updates. Robustness is further enhanced by employing ensemble methods where multiple global model snapshots from different regimes are combined using a dynamic weighting scheme that reflects current economic conditions. A recent benchmark design study [5] has demonstrated the importance of evaluating early warning systems under economically credible stress scenarios, highlighting that models which perform well in backtests may fail in out-of-sample crises. The framework's adaptation mechanisms are validated through continuous online evaluation using holdout data that spans multiple economic cycles, though real-world deployment requires careful calibration to avoid overfitting to transient fluctuations.

6. Governance, Fairness, and Policy Implications

The deployment of a federated credit default prediction system raises profound governance questions regarding accountability, transparency, and fairness. Unlike centralized models where a single entity is responsible for the model's decisions, a federated system distributes responsibility across multiple institutions and a coordinating body. Clear contractual agreements must define the roles, liabilities, and decision rights of each participant. In the event of a model failure that leads to discriminatory lending practices or systemic losses, allocating blame becomes legally complex. The framework should include an audit trail that records the aggregation process and the contributions of each institution, while respecting privacy constraints. Cryptographic techniques such as zero-knowledge proofs can enable verifiability without revealing sensitive data [17]. Fairness in federated credit scoring is particularly challenging because disparities can arise from multiple sources: heterogeneous data distributions, differential privacy noise that affects minority groups disproportionately, and aggregation weights that favor larger institutions. The framework incorporates fairness constraints at the aggregation step by penalizing model updates that increase disparate impact

across protected attributes, such as race, gender, or age [12,13]. However, measuring fairness in a distributed setting requires careful design because the server does not have access to the underlying demographic labels. Instead, each institution can compute local fairness metrics and submit them as part of the update, aggregated in a privacy-preserving manner using secure computation. Policy implications extend to regulatory compliance with laws such as the Equal Credit Opportunity Act, which requires lenders to provide adverse action notices explaining the specific reasons for denial. A federated model that is a black-box ensemble may make it difficult to produce such explanations. Therefore, the framework must support interpretability techniques, such as feature importance scores computed from the global model and distributed to institutions, albeit with potential privacy leakage. Recent policy discussions have called for the establishment of shared infrastructure standards for privacy-preserving financial AI, including certification of aggregation protocols and regular fairness audits by independent third parties [18].

7. Deployment and Infrastructure Sustainability

Moving from theoretical design to real-world deployment demands careful attention to infrastructure, scalability, and sustainability. Financial institutions typically operate legacy IT systems with varying levels of computational capacity. The federated learning framework must accommodate heterogeneity in hardware, network bandwidth, and latency. Asynchronous training protocols allow institutions with slower connections to participate without blocking faster peers, but they introduce staleness issues that can degrade model convergence [19]. The framework employs a staleness-tolerant aggregation algorithm that weights updates based on their recency and the variance of the local data. Communication costs represent a major bottleneck, especially when models are large or when differential privacy noise requires high-precision aggregation. Compression techniques such as gradient quantization and sparse updates can reduce bandwidth usage at the expense of additional noise. Energy consumption is an emerging concern, as the computational overhead of local differential privacy and secure aggregation can be substantial, particularly for institutions processing millions of records. Green computing principles should guide the selection of cryptographic primitives and training schedules, favoring lightweight protocols such as low-depth circuits for secure aggregation [20]. Edge computing can offload some computational burden from central servers by performing partial aggregation at regional nodes, reducing latency and improving fault tolerance. Sustainability also involves model maintenance over time: models must be periodically retrained, privacy budgets reset, and participant credentials updated. The framework includes a lifecycle management module that schedules retraining events, monitors model drift, and handles the orderly addition or removal of institutions. Cross-domain lessons from healthcare, where federated learning has been deployed across hospitals with strict data governance, provide valuable insights into building trust among participants and ensuring data quality [6]. In the financial sector, the high economic stakes necessitate more rigorous stress testing of the infrastructure itself, including simulations of network partitions, adversarial attacks, and regulatory audits.

8. Conclusion

This paper has presented a comprehensive federated learning framework for privacy-preserving credit default prediction that is explicitly designed to operate under dynamic economic regimes. The architecture integrates local differential privacy, secure aggregation, adaptive regime detection, and fairness constraints within a hierarchical system that balances competing objectives. We have examined structural trade-offs between privacy guarantees,

model accuracy, communication efficiency, and computational overhead, emphasizing that no single configuration is optimal across all contexts. Governance and policy considerations are paramount, requiring clear accountability structures, auditable processes, and mechanisms for interpretability and non-discrimination. Deployment challenges related to infrastructure heterogeneity, energy sustainability, and lifecycle management underscore the need for close collaboration between financial institutions, regulators, technology providers, and academic researchers. Future work should focus on empirical validation of the framework using real-world credit data from multiple economic cycles, development of standardized benchmarking protocols that incorporate economic credibility [5], and exploration of decentralized federated topologies that eliminate the central coordinator to further reduce trust assumptions. As financial markets continue to evolve and privacy regulations tighten, the proposed framework offers a principled pathway toward responsible AI in credit risk assessment.

References

1. McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS) (pp. 1273–1282). PMLR.

2. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS) (pp. 308–318). ACM. https://doi.org/10.1145/2976749.2978318

3. Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science, 9(3–4), 211–407. https://doi.org/10.1561/0400000042

4. Bhowmick, A., Duchi, J., Freudiger, J., Kapoor, G., & Rogers, R. (2018). Protection against reconstruction and its applications in private federated learning. arXiv preprint arXiv:1812.00984.

5. Liu, T. (2026). Leakage-safe benchmark design for market-stress early warning: An economically credible evaluation.

6. Sheller, M. J., Edwards, B., Reina, G. A., Martin, J., Pati, S., Kotrotsou, A., ... & Bakas, S. (2020). Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Scientific Reports, 10, 12598. https://doi.org/10.1038/s41598-020-69250-1

7. Niknam, S., Dhillon, H. S., & Reed, J. H. (2020). Federated learning for wireless communications: motivation, opportunities, and challenges. IEEE Communications Magazine, 58(6), 46–51. https://doi.org/10.1109/MCOM.001.1900649

8. Long, G., Tan, Y., Jiang, J., & Zhang, C. (2020). Federated learning for credit scoring: A privacy-preserving approach. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1–8). IEEE. https://doi.org/10.1109/IJCNN48605.2020.9207607

9. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. ACM Computing Surveys, 46(4), 1–37. https://doi.org/10.1145/2523813

10. Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In Proceedings of the 3rd Theory of Cryptography Conference (TCC) (pp. 265–284). Springer. https://doi.org/10.1007/11681878_14

11. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., ... & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS) (pp. 1175–1191). ACM. https://doi.org/10.1145/3133956.3133982

12. Mohri, M., Sivek, G., & Suresh, A. T. (2019). Agnostic federated learning. In Proceedings of the 36th International Conference on Machine Learning (ICML) (pp. 4615–4625). PMLR.

13. Zafar, M. B., Valera, I., Gomez Rodriguez, M., & Gummadi, K. P. (2017). Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In Proceedings of the 26th International Conference on World Wide Web (WWW) (pp. 1171–1180). ACM. https://doi.org/10.1145/3038912.3052660

14. Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica, 57(2), 357–384. https://doi.org/10.2307/1912559

15. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhao, S. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210. https://doi.org/10.1561/2200000083

16. Goryczka, S., & Xiong, L. (2018). A comprehensive comparison of multi-party secure additions with differential privacy. IEEE Transactions on Dependable and Secure Computing, 15(6), 1061–1074. https://doi.org/10.1109/TDSC.2017.2715189

17. Goldwasser, S., Micali, S., & Rackoff, C. (1989). The knowledge complexity of interactive proof systems. SIAM Journal on Computing, 18(1), 186–208. https://doi.org/10.1137/0218012

18. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

19. Xie, C., Koyejo, O., & Gupta, I. (2019). Asynchronous federated optimization. arXiv preprint arXiv:1903.03934.

20. Juuti, M., Kaski, S., & Asokan, N. (2020). Efficient secure aggregation for privacy-preserving federated learning. In Proceedings of the 2020 IEEE European Symposium on Security and Privacy (EuroS&P) (pp. 465–480). IEEE. https://doi.org/10.1109/EuroSP48549.2020.00038