الجمهورية الشعبية الديمقراطية الجزائرية
People's Democratic Republic of Algeria
وزارة التعليم العالي و البحث العلمي
Ministry of Higher Education and Scientific Research
المدرسة العليا للإعلام الآلي 8 ماي 1945 - سيدي بلعباس
Higher School of Computer Science
8 Mai 1945 - Sidi Bel Abbes

Graduation Thesis

To obtain the diploma of **Engineering Degree**

Field of Study: **Computer Science**

Specialization: **artificial intelligence and data science**

Theme

Privacy-Preserving Federated Random Forest for Healthcare Applications

---

Presented by
Mohamed El Amine Serradj

Defended on: **09, 2025**
In front of the jury composed of

Mr.BENDELLA Mohammed Salih	President of the Jury
Mr. KHALDI Miloud	Thesis Supervisor
Mr. ELHANNANI Souad	Examiner

Academic Year: 2024/2025

Abstract

In the age of data-driven decision making, privacy concerns remain a central obstacle to collaborative machine learning across organizations. This thesis addresses this challenge by presenting the design, implementation, and evaluation of a fully decentralized, privacy-preserving Random Forest model integrated into the FLUTE (Federated Learning Utilities and Tools for Europe) platform. Our approach enables multiple data owners to collaboratively train decision forests without revealing their sensitive data and without relying on a central aggregator.

The proposed system adopts a totally randomized tree construction strategy to maximize privacy budget allocation at the leaf nodes—where prediction occurs—thus improving model utility under differential privacy constraints. Both Laplace and Exponential mechanisms are implemented to support trade-offs between accuracy and interpretability. Secure aggregation and pruning are achieved through multi-party computation using Shamir Secret Sharing, ensuring that class distributions and pruning decisions remain private.

The entire workflow, from training to prediction, is executed in a fully decentralized manner using FLUTE's AIAlgorithm orchestration layer, with each node independently participating in secure aggregation and local inference. We evaluate the system on four benchmark datasets—Adult, Breast Cancer, Diabetes, and Wine—under varying privacy budgets and ensemble sizes. Results show that our decentralized approach achieves competitive accuracy while satisfying strong privacy guarantees.

This thesis demonstrates the feasibility of privacy-preserving federated tree-based learning in realistic settings and provides a modular framework for deploying such models in privacy-critical domains such as healthcare, finance, and smart cities. Future work will focus on adversarial robustness, scalability, and integration with real-time decision pipelines.

Keywords— Federated Learning, Privacy-Preserving Machine Learning,Random Forest,Differential Privacy,Secure Multi-Party Computation,Decentralized Systems

Résumé

À l'ère de la prise de décision fondée sur les données, les préoccupations liées à la vie privée représentent un obstacle majeur à l'apprentissage automatique collaboratif entre organisations. Ce mémoire s'attaque à ce défi en proposant la conception, l'implémentation et l'évaluation d'un modèle de forêt aléatoire entièrement décentralisé et préservant la confidentialité, intégré à la plateforme européenne FLUTE (Federated Learning Utilities and Tools for Europe). Notre approche permet à plusieurs détenteurs de données d'entraîner conjointement des forêts de décision sans exposer leurs données sensibles et sans recourir à un agrégateur central.

Le système proposé adopte une stratégie de construction totalement aléatoire des arbres afin de concentrer l'intégralité du budget de confidentialité sur les nœuds feuilles — là où les prédictions sont effectuées — améliorant ainsi les performances sous contraintes de confidentialité différentielle. Les mécanismes de Laplace et exponentiel sont tous deux implémentés, permettant un compromis entre précision et interprétabilité. L'agrégation sécurisée et l'élagage sont réalisés via des techniques de calcul multipartite utilisant le partage de secrets de Shamir, garantissant la confidentialité des distributions de classes et des décisions d'élagage.

L'ensemble du processus, de l'apprentissage à la prédiction, est exécuté de manière totalement décentralisée grâce au système d'orchestration `AIAlgorithm` de FLUTE, chaque nœud participant indépendamment à l'agrégation sécurisée et à l'inférence locale. L'évaluation sur quatre jeux de données (Adult, Breast Cancer, Diabetes, Wine), sous divers niveaux de confidentialité et tailles d'ensembles, démontre la viabilité et la performance du modèle proposé. Ce travail ouvre la voie à des systèmes d'apprentissage fédéré robustes, privés et exploitables dans des secteurs sensibles comme la santé, la finance et les villes intelligentes.

Keywords— Apprentissage Fédéré, Forêt Aléatoire,Confidentialité Différentielle,Calcul Multi-Party Sécurisé,Systèmes Décentralisés