INTISARI

Sebagian besar data transaksi perbankan bersifat *imbalanced*, dimana hal ini menjadikan model klasifikasi tradisional rentan menghasilkan *false positive* yang tinggi, mengakibatkan biaya investigasi tambahan serta potensi ketidaknyamanan bagi nasabah yang sah. Berbagai pendekatan, seperti *data balancing* melalui teknik *oversampling* ataupun *undersampling*, maupun *data augmentation*, strategi-strategi ini sering kali memperkenalkan *synthetic data*, meningkatkan risiko *overfitting* pada model klasifikasi tersebut. *Federated Learning* (FL)—paradigma yang melatih model kecerdasan buatan secara terdistribusi tanpa memindahkan data mentah—hadir sebagai alternatif dibanding praktik konvensional untuk melakukan *training data* dengan data yang terbatas untuk tiap institusi. Walau demikian, gradien yang dikirim dalam FL masih dapat mengungkap informasi sensitif. Di sinilah *Differential Privacy* (DP) berfungsi: menambahkan *noise* Gaussian terkontrol sebesar σ dan membatasi kebocoran lewat anggaran privasi ε (privacy budget), yaitu batas numerik seberapa banyak informasi tentang individu dapat tersirat dari keluaran algoritma.

Riset ini merancang dan mengevaluasi **Differentially-Private Federated Learning** pada *European Credit Card Fraud Dataset* memakai arsitektur *Convolutional Neural Network*. Sensitivitas gradien dibatasi melalui *clipping*, yakni memotong norma gradien ke ambang tertentu agar setiap pembaruan tidak menghasilkan kebocoran data. Empat skema clipping diuji—**fixed clipping** dan **adaptive clipping** yang ditempatkan di ***client*** ataupun **server**. Masing-masing dikombinasikan dengan empat tingkat *noise* ($\sigma = 0, 0{,}01, 0{,}03,$ dan $0{,}10$) dan disimulasikan pada dua klien selama 50 *communication rounds*. Kinerja diukur menggunakan (ε), *clipping norm*, *accuracy*, dan *Area Under the ROC Curve* (AUC)—metrik yang tangguh untuk data penipuan yang tidak seimbang.

Eksperimen menegaskan keberadaan *privacy–utility trade-off*. Pada tingkat *noise* rendah, seluruh varian model mempertahankan akurasi dan AUC tinggi dengan *privacy budget* yang masih ketat; sebaliknya, penambahan *noise* secara bertahap menaikkan ε sekaligus menurunkan utilitas model. Di antara keempat skema, konfigurasi **adaptive-side adaptive clipping** paling stabil sehingga memberikan kompromi terbaik antara perlindungan privasi formal dan efektivitas deteksi penipuan yang dapat diterapkan di dunia nyata.

Kata kunci: *Federated Learning*, *Differential Privacy*, *Clipping*, *Convolutional Neural Network*, *Credit Card Fraud*.

ABSTRACT

Most banking-transaction datasets are highly *imbalanced*, where such imbalance makes conventional classifiers prone to large numbers of *false positives*, driving up investigation costs and inconveniencing legitimate customers. Typical counter-measures—oversampling, undersampling, or other forms of *data augmentation*—inject synthetic examples, raising the risk of *overfitting* and, critically, leaving the problem of data fragmentation across institutions unsolved.

Federated Learning (FL) offers a broader remedy: each bank trains locally and transmits privacy-preserving model updates—rather than raw records—to a central server, collectively enriching the minority class while keeping customer data on-premise. Yet the exchanged gradients can still leak sensitive information. *Differential Privacy* (DP) mitigates this risk by clipping each gradient to a norm bound C_t and adding calibrated Gaussian *noise* with scale σ; the resulting privacy budget ε upper-bounds what an adversary can infer about any individual.

This study designs and evaluates a **differentially private FL** scheme on the *European Credit-Card Fraud 2013* dataset using a lightweight *Convolutional Neural Network*. Four clipping variants are investigated—**fixed** and **adaptive** clipping placed either on the **client** or on the **server**. Each variant is combined with four *noise* levels ($\sigma = 0, 0.01, 0.03,$ and 0.10) and simulated on two clients over 50 communication rounds. Performance is assessed by the privacy budget (ε), the clipping norm, *Accuracy*, and the *Area Under the ROC Curve* (AUC)—a metric well-suited to highly imbalanced fraud data.

The experiments confirm a clear *privacy–utility trade-off*. At low noise levels, all model variants retain high Accuracy and AUC while keeping a tight privacy budget; conversely, increasing σ enlarges ε and progressively degrades utility. Among the four schemes, the **server-side adaptive clipping** configuration proves the most stable, offering the best compromise between formal privacy guarantees and effective fraud-detection performance.

Keywords: *Federated Learning*, *Differential Privacy*, *Clipping*, *Convolutional Neural Network*, *Credit-Card Fraud*.