BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Federated learning merupakan pendekatan *machine learning* terdesentralisasi yang memungkinkan pelatihan dilakukan pada perangkat di mana data berada agar privasi data lebih terjaga. Masing-masing perangkat (klien) tidak mengirimkan data lokalnya, melainkan hanya pembaruan model lokal yang dihasilkan dari pelatihan, ke server yang mengagregasikannya menjadi model global [1]. Masalah privasi masih ada karena server, klien, atau penyerang lain dapat melakukan inferensi pada data lokal menggunakan parameter dari model yang dikirim [2].

Homomorphic encryption menjadi solusi yang memungkinkan operasi aritmatika dilakukan pada parameter model yang terenkripsi tanpa perlu dekripsi sehingga server dapat mengagregasi tanpa mengakses data aslinya. Meskipun *homomorphic encryption* memberikan perlindungan privasi yang kuat, penerapannya menghasilkan *overhead* komputasi dan komunikasi yang signifikan [3]. Hu dan Li (2025) mengusulkan MaskCrypt dengan *selective homomorphic encryption* (SHE) yang hanya mengenkripsi sebagian pembaruan model untuk mengurangi *overhead* komunikasi sekaligus memberikan perlindungan privasi yang efektif. Penelitian tersebut mendistribusikan dataset secara acak dan merata kepada seluruh klien, yang disebut sebagai *independent and identically distributed* (IID) [4].

Dalam praktiknya, distribusi data di antara klien dalam *federated learning* dapat sangat beragam/heterogen (non-IID) karena setiap klien secara mandiri mengumpulkan data lokal sesuai dengan preferensi dan ruang pengambilan sampel mereka sendiri [5]. Ketika data tersebar secara non-IID, model lokal memiliki arah yang menyimpang dari model global sehingga akurasi menurun, konvergensi menjadi fluktuatif, dan *loss* tidak menurun secara stabil dibandingkan dengan skenario IID [6, 7]. Li et al. (2018) mengusulkan FedProx yang menambahkan *proximal term* ke fungsi objektif pelatihan lokal klien untuk meningkatkan stabilitas konvergensi dalam kondisi heterogen [8].

Penelitian ini mengeksplorasi integrasi antara mekanisme SHE yaitu MaskCrypt, dengan formulasi yang tahan terhadap heterogenitas data yaitu *proximal term*. Upaya ini dilakukan untuk mencapai keseimbangan optimal antara privasi, efisiensi komunikasi, dan stabilitas performa model dalam skenario *federated learning* yang realistis. Dengan mengombinasikan kedua pendekatan tersebut, diharapkan sistem *federated learning* dapat mengatasi tantangan ganda, yaitu melindungi privasi data sensitif klien melalui enkripsi selektif sambil memastikan konvergensi yang stabil dan akurat meskipun data tersebar secara non-IID di berbagai perangkat.

1.2 Rumusan Masalah

Rumusan masalah yang hendak diselesaikan dalam penelitian ini adalah sebagai berikut.

1. Bagaimana mengintegrasikan formulasi *proximal term* pada MaskCrypt?
2. Bagaimana akurasi dan *loss* MaskCrypt pada distribusi data non-IID setelah diintegrasikan dengan *proximal term*?

1.3 Batasan Permasalahan

Batasan masalah yang mempertajam fokus penelitian ini adalah sebagai berikut.

1. Dataset yang digunakan terbatas pada MNIST dan CIFAR-10 yang tersedia secara publik.
2. Arsitektur model yang digunakan adalah LeNet-5 pada dataset MNIST dan ResNet-18 pada dataset CIFAR-10.
3. Distribusi data pada klien mengikuti kondisi non-IID yang disimulasikan menggunakan distribusi Dirichlet dengan parameter konsentrasi $\alpha = 0,5$.
4. Modifikasi yang dilakukan dalam penelitian ini terbatas pada integrasi *proximal term* ke dalam fungsi objektif pada proses pelatihan lokal klien.
5. Metrik evaluasi yang digunakan adalah akurasi dan *loss* model pada pelatihan.

1.4 Tujuan Penelitian

Tujuan yang hendak dicapai dari penelitian ini adalah sebagai berikut.

1. Mengintegrasikan formulasi *proximal term* pada MaskCrypt.
2. Mengevaluasi akurasi dan *loss* MaskCrypt pada distribusi data non-IID setelah diintegrasikan dengan *proximal term*.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah sebagai berikut.

1. Memberikan kontribusi akademis dalam pengembangan *federated learning* yang menggabungkan SHE (MaskCrypt) dan *proximal term* untuk menciptakan sistem yang aman terhadap ancaman privasi sambil tetap efisien dalam komunikasi pada skenario data non-IID.
2. Menjadi rujukan praktis bagi penelitian penerapan *homomorphic encryption* dalam *federated learning* dengan heterogenitas data, khususnya dalam aplikasi yang melibatkan data sensitif yang memerlukan perlindungan privasi kuat.

1.6 Sistematika Penulisan

Sistematika dari penulisan laporan ini adalah sebagai berikut.

1. Bab 1 PENDAHULUAN
   Bab ini berisi latar belakang masalah, rumusan masalah, batasan permasalahan, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.
2. Bab 2 LANDASAN TEORI
   Bab ini membahas teori yang mendasari penelitian, meliputi konsep *federated learning*, *homomorphic encryption*, non-IID, dan *proximal term* sebagai dasar konseptual dalam perancangan dan implementasi penelitian.
3. Bab 3 METODOLOGI PENELITIAN
   Bab ini menjabarkan metodologi penelitian yang hendak dilakukan dalam menyusun dan mengerjakan penelitian ini, mencakup studi literatur, perancangan, implementasi, dan evaluasi.

4. Bab 4 HASIL DAN DISKUSI

   Bab ini menyajikan hasil implementasi dan evaluasi metode yang diusulkan, yang dianalisis berdasarkan performa model yang meliputi akurasi dan *loss*, serta pengaruh *proximal term* terhadap proses pelatihan.

5. Bab 5 KESIMPULAN DAN SARAN

   Bab ini berisi simpulan dari hasil penelitian yang diperoleh dan saran untuk pengembangan penelitian selanjutnya.