BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Federated Learning (FL) menjadi salah satu pendekatan untuk melatih model secara kolaboratif tanpa memindahkan *raw data* dari *client* ke *server* pusat. Pada skema dasar seperti FedAvg, *server* mengirimkan model global kepada *client*, *client* melakukan *training* secara lokal, kemudian *server* mengagregasi pembaruan model dari *client* terpilih [2]. Meskipun demikian, karakteristik data pada *edge device* tetap dapat menurunkan performa FL. Data pada setiap *client* sering kali tidak seimbang, berjumlah sedikit, dan hanya merepresentasikan sebagian kelas dari distribusi global [3, 4, 5].

Data scarcity dan *label skew* merupakan dua kondisi yang banyak dibahas pada FL. *Data scarcity* dapat membuat model lokal terlalu menyesuaikan diri terhadap sampel yang terbatas, sedangkan *label skew* dapat menyebabkan pembaruan model lokal bias terhadap kelas yang tersedia pada *client* tertentu. Kombinasi kedua kondisi tersebut dapat menghasilkan *local model drift* dan menurunkan kualitas model global. FLea dikembangkan untuk mengatasi kondisi tersebut melalui *global feature buffer* yang menyimpan *activation-target pair* dari beberapa *client*. *Activation* tersebut digunakan dalam *feature augmentation* untuk memperkaya *local training*, sedangkan *distilling loss* berbasis KL-divergence dan *decorrelation loss* digunakan untuk mengurangi *local model drift* serta keterkaitan antara *intermediate activation* dan *raw input* [1, 6, 7].

Mekanisme *feature sharing* dapat membantu *client* memperoleh informasi tambahan tanpa membagikan *raw data* secara langsung. Namun, *intermediate activation* tetap merupakan representasi yang dihasilkan dari data lokal, sehingga informasi yang dibagikan masih perlu diperlakukan sebagai objek yang sensitif. Parameter model, gradien, dan *activation* dapat membawa informasi tentang *raw input* dalam kondisi tertentu [8, 9, 10, 11]. Oleh karena itu, *feature sharing* perlu dilihat tidak hanya dari sisi utilitas, tetapi juga dari sisi pengurangan informasi yang dapat dikaitkan kembali dengan *input*. Risiko tersebut dapat dianalisis melalui keterkaitan antara *raw input* dan representasi yang dibagikan. *Distance correlation* digunakan sebagai indikator *information leakage* untuk mengukur keterkaitan pada *intermediate representation* terhadap *raw input* [12, 13].

Differential Privacy (DP) menawarkan kerangka untuk membatasi informasi yang dapat diungkap dari *output* suatu mekanisme melalui penambahan *noise* terkontrol [14, 15]. Pada sisi *client*, pendekatan lokal dapat diterapkan sebelum informasi dikirimkan kepada *server*, sehingga representasi yang diterima *server* bukan lagi nilai asli yang langsung dihasilkan dari data lokal [16, 17]. Pada objek berbentuk *activation vector* numerik, *Gaussian Mechanism* dapat diterapkan setelah sensitivitas vektor dibatasi melalui *L2 clipping* [15, 18].

Berdasarkan permasalahan *privacy* pada *feature buffer*, penelitian ini menerapkan *Gaussian Mechanism* pada *intermediate activation* di sisi *client* sebelum *activation* tersebut dibagikan kepada *server*. Mekanisme ini menggabungkan *L2 clipping* dan penambahan *Gaussian noise* untuk membatasi sensitivitas serta mengurangi keterkaitan antara *input* dan *activation* yang dilepas. Kontribusi penelitian ini adalah merancang dan mengevaluasi penerapan *Gaussian Mechanism* pada *feature buffer* dalam skema *Federated Learning*, serta menganalisis kompromi antara perlindungan informasi dan utilitas model berdasarkan akurasi klasifikasi dan *distance correlation*.

1.2 Rumusan Masalah

1. Bagaimana *Gaussian Mechanism* pada sisi *client* diterapkan untuk mengurangi risiko *privacy* pada *intermediate activation* sebelum *activation* tersebut dibagikan melalui *feature buffer* dalam skema *Federated Learning* untuk klasifikasi aktivitas manusia?

2. Bagaimana pemilihan batas *clipping C* dan nilai δ sebagai batas probabilitas kegagalan tambahan dalam jaminan *privacy*, serta pengaruh nilai *privacy budget* (ε) pada *Gaussian Mechanism*, terhadap akurasi klasifikasi dan *distance correlation* sebagai ukuran keterkaitan antara *input* dan *activation* pada *feature buffer*?

1.3 Batasan Permasalahan

1. *Dataset* yang digunakan terbatas pada UCI-HAR untuk klasifikasi enam aktivitas manusia berbasis sensor *smartphone*.

2. Model klasifikasi yang digunakan terbatas pada HARNet.

3. Skenario *Federated Learning* difokuskan pada *data scarcity* dan *label skew* dengan pembagian data non-IID.
4. Mekanisme *privacy* dibatasi pada penerapan *Gaussian Mechanism* terhadap *intermediate activation* pada sisi *client* sebelum masuk ke *feature buffer*.
5. Eksperimen dilakukan sebagai simulasi *Federated Learning* lokal, bukan sebagai implementasi jaringan terdistribusi nyata.
6. Seluruh konfigurasi eksperimen menggunakan satu *seed*, yaitu *seed* 0, dengan pembagian data, inisialisasi model, dan urutan pemilihan *client* yang sama.
7. Evaluasi utama dibatasi pada akurasi klasifikasi dan *distance correlation*.

1.4 Tujuan Penelitian

1. Mengimplementasikan *Gaussian Mechanism* pada sisi *client* untuk mengurangi risiko *privacy* pada *intermediate activation* sebelum *activation* tersebut dibagikan melalui *feature buffer* dalam skema *Federated Learning* untuk klasifikasi aktivitas manusia.
2. Menganalisis pemilihan batas *clipping* C dan nilai δ sebagai batas probabilitas kegagalan tambahan dalam jaminan *privacy*, serta pengaruh nilai *privacy budget* (ε) pada *Gaussian Mechanism* terhadap akurasi klasifikasi dan *distance correlation* sebagai ukuran keterkaitan antara *input* dan *activation* pada *feature buffer*.

1.5 Manfaat Penelitian

1. Memberikan kontribusi akademis dalam kajian *privacy-utility trade-off* pada mekanisme *feature sharing* dalam *Federated Learning*.
2. Menjadi referensi praktis bagi penerapan *Gaussian Mechanism* pada *feature buffer* untuk klasifikasi aktivitas manusia berbasis data sensor.
3. Menyediakan hasil evaluasi mengenai pemilihan batas *clipping* C dan nilai δ sebagai batas probabilitas kegagalan tambahan dalam jaminan *privacy*, serta pengaruh nilai *privacy budget* (ε) terhadap akurasi klasifikasi dan *distance correlation* pada *feature buffer*.

1.6 Sistematika Penulisan

- Bab 1 PENDAHULUAN
  Bab ini memuat latar belakang penelitian, rumusan masalah, batasan permasalahan, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

- Bab 2 LANDASAN TEORI
  Bab ini berisi kajian pustaka yang mendasari penelitian, meliputi penelitian terdahulu, *Federated Learning*, FedAvg, *data scarcity*, *label skew*, FLea, *Differential Privacy*, *Gaussian Mechanism*, klasifikasi aktivitas manusia, HARNet, serta metrik evaluasi.

- Bab 3 METODOLOGI PENELITIAN
  Bab ini menjelaskan alur penelitian, persiapan *dataset*, pembagian data *client*, rancangan skema *Federated Learning*, implementasi model, penerapan *Gaussian Mechanism* pada *feature buffer*, tahapan pengujian parameter, serta rancangan evaluasi.

- Bab 4 HASIL DAN DISKUSI
  Bab ini menyajikan hasil pemilihan batas *clipping C*, nilai δ, dan *privacy budget* (ε), serta membahas akurasi klasifikasi, *utility loss*, *distance correlation*, dan *privacy-utility trade-off*.

- Bab 5 SIMPULAN DAN SARAN
  Bab ini memuat simpulan berdasarkan hasil penelitian serta saran untuk pengembangan penelitian selanjutnya, baik dari sisi peningkatan performa model maupun penguatan mekanisme *privacy*.