BAB I

PENDAHULUAN

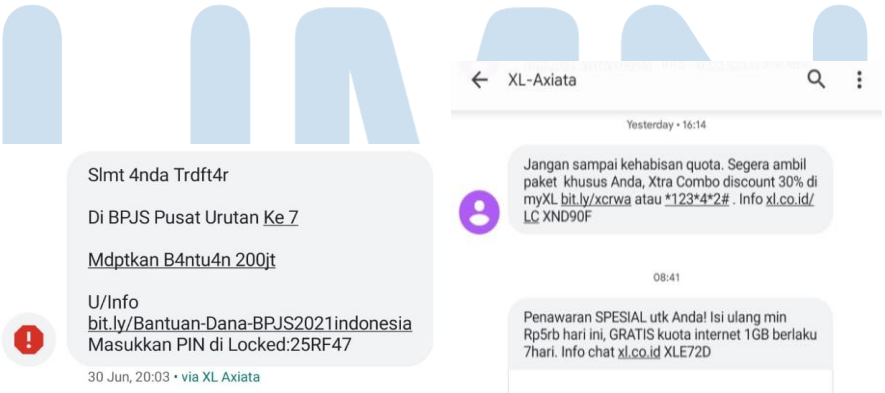
1.1 Latar Belakang

Data pribadi merupakan informasi penting yang tidak boleh disebarkan oleh sembarangan orang. Terdapat keamanan informasi/ *information security* yang dibuat untuk memenuhi prinsip *confidentiality, integrity,* dan *availability* (CIA) yang melindungi aset informasi terhadap ancaman yang tidak diinginkan. Dalam hal ini, ada yang disebut sebagai insiden keamanan informasi yaitu potensi pelanggaran maupun ancaman terhadap keamanan informasi yang diakses melalui cara tidak sah atau diretas oleh pihak yang tidak seharusnya memiliki akses informasi tersebut. *Data breach* atau biasa disebut dengan pelanggaran data merupakan bagian dari insiden keamanan informasi.

Menurut informasi dari Badan Siber dan Sandi Negara (BSSN) [1], *data breach* terjadi akibat pencurian atau hilangnya perangkat penyimpanan data, akses ilegal terhadap sistem atau informasi, adanya keterlibatan dari orang dalam, dan adanya kelalaian dalam penerapan protokol pengamanan informasi. Penyebab paling umum terjadinya *data breach* yaitu: 1) Adanya peretasan/ *hacking* pada *password* melalui serangan *brute force*; 2) *error* yang terjadi pada sistem maupun *human error;* 3) adanya *social attacks* melalui *phishing* yaitu membuat tautan palsu yang memancing target melalui penyamaran yang paling umum terjadi yaitu dengan *clickbait,* sms palsu, *e-mail* palsu, maupun telepon palsu dengan kata-kata yang menarik; 4) *Malware (Malicous software)*, program yang dirancang untuk mengeksploitasi data-data penting pada sistem dan *database.*

Menurut Katadata yang mengutip situs *patrolisiber.id*, terdapat 15.152 aduan kejahatan siber yang dilaporkan melalui situs tersebut dengan kasus penipuan paling banyak dilaporkan sebanyak 4.601 kasus sepanjang Januari hingga September 2021. Menurut *platform* tersebut, kejahatan siber paling banyak menggunakan aplikasi berbasis teks seperti pesan WhatsApp sebanyak 8.357 kasus, Instagram sebanyak 2.621 kasus, dan Telepon/ SMS sebanyak 2.324 kasus [2].

Phishing memiliki kaitan yang erat dengan teknik penipuan digital melalui *social engineering*. Menurut laporan ISACA yang berjudul *Global State of Cybersecurity 2020: Threat Landscape and Security Practices*, *social engineering* merupakan ancaman terbesar selama pandemi dengan jumlah persentase mencapai lima belas persen [3]. Salah satu bentuk penipuan yang dilakukan adalah melalui pesan SMS (*Short Messaging Service*) atau sebutannya *SMShing.* Oleh karena itu diperlukannya sebuah sistem yang dapat mengklasifikasi pesan SMS yang merupakan penipuan atau *scam* dan bukan penipuan atau *non-scam*. Tugas klasifikasi teks umumnya dilakukan dengan pelatihan model AI (Artificial Intelligence). Menurut artikel yang ditulis oleh Verizon, pesan yang terklasifikasi sebagai *scam* adalah pesan dengan intensi untuk memperdaya *user* dengan memberikan informasi palsu yang dapat menyebabkan korban memberikan kembali informasi sensitif tentang dirinya ataupun memberikan sejumlah uang atau materi ke penipu. Pola yang ada pada pesan *scam* adalah kata yang disingkat secara berlebih, mengalami perubahan *character* dari *alphabet* menjadi angka, dan umumnya memiliki pesan yang menarik perhatian korban dengan menjanjikan hadiah berupa uang ataupun bantuan. Sementara, pesan *non-scam* adalah yang berisi teks percakapan manusia pada umumnya dan promosi dari sebuah perusahaan. [4]



Gambar 1. 1 Cuplikan SMS Scam dan Non-scam

Gambar 1.1 merupakan cuplikan perbedaan SMS yang merupakan *scam* (kiri) dan *non-scam* (kanan). Data berupa pesan teks SMS milik para *user* merupakan sebuah data yang tidak tersentralisasi (*decentralized data)* dengan jumlah total yang

sangat banyak. Artinya, setiap orang memiliki perangkat *smartphone* yang menyimpan pesan SMS dan jika dijumlahkan dengan jumlah pesan SMS *user* lain, pesan-pesan tersebut membentuk sebuah *dataset* besar yang dapat digunakan untuk melatih model pengklasifikasi. Namun, pengkoleksian dan pengiriman data pesan SMS milik seorang *user* ke suatu pihak merupakan sebuah *privacy concern*.

Satu cara yang dapat dilakukan untuk melakukan *training* menggunakan *user data* yang sensitif dengan menjaga *user privacy* adalah menggunakan teknik Federated Learning. Berbeda dengan pelatihan Classic ML yang pada umumnya dilakukan pada satu mesin, Federated Learning memiliki sebuah algoritma bernama FedAvg (Federated Averaging) yang dapat mengagregasi lebih dari satu model dari berbagai macam perangkat menjadi sebuah *global model* bisa diperbaharui dan disebar ulang ke berbagai perangkat *user* untuk mengulangi kembali tahap pelatihan. Data yang dikomunikasikan dari *client* ke *server* adalah *learning parameter* saja sehingga tidak ada perpindahan *user data* seperti yang dilakukan pada pelatihan Classic ML. Berdasarkan eksperimen yang dilakukan jurnal [5] yaitu perbandingan antara Classical ML dengan Federated Learning, Federated Learning dapat mengalahkan performa Classical ML yang terdiri dari Centralized ML dan Distributed ML dikarenakan memberi opsi *distributed and collaborative training* dan juga menyediakan berbagai macam fitur *security*. Konklusi pada jurnal menyatakan bahwa mempertimbangkan performa yang dicapai dan aktivitas *malicious* yang terjadi di era ini, Federated Learning merupakan cara yang cocok untuk diterapkan pada aplikasi AI zaman sekarang, terutama pada *platform* yang melibatkan data pribadi *user* yang berjumlah banyak seperti pelatihan model pengklasifikasi SMS.

Pelatihan model dengan Federated Learning merupakan bidang yang aktif diteliti, sehingga terdapat berbagai macam modifikasi metode Federated Learning seperti *standard* Federated Averaging (FedAvg) [6], FedAvg dengan komunikasi secara Daisy-chain (FedDC) yang mengutamakan akurasi pada *dataset* kecil [7], dan *simple daisy-chaining* tanpa agregasi yang dari percobaan jurnal [7] dapat membandingi FedAvg menggunakan *dataset* kecil. Data yang digunakan untuk melakukan Federated Learning umumnya bersifat *non-iid* (*non independent and*

identically distributed) di mana data antara satu *user* dengan lainnya sangat berbeda karakteristik atau *feature space*-nya [5]. *Dataset* pesan SMS berbahasa Indonesia yang tersedia juga memiliki jumlah yang sangat terbatas.

Penelitian ini membandingkan metode Federated Learning menggunakan agregasi (Common Federated Learning) dengan metode Federated Learning tanpa agregasi (Daisy-chain Federated Learning) untuk mengklasifikasi pesan menjadi dua kelas yaitu *scam* dan *non-scam*. Perbandingan ini menguji metode mana yang memiliki performa lebih baik pada jumlah *dataset* yang sedikit dengan ada atau tidaknya keterlibatan algoritma agregasi pada pelatihan secara Federated Learning. Karena *dataset* pesan SMS berbahasa Indonesia yang tersedia memiliki jumlah sangat terbatas, hal ini menjadi alasan mengapa metode *simple daisy-chaining* dipilih untuk digunakan dalam penelitian ini. Menurut jurnal [7] metode ini (Daisy-chain Federated Learning) efektif menghadapi *dataset* kecil dan juga satu-satunya metode tanpa agregasi.

1.2 Identifikasi Masalah

Berikut identifikasi masalah yang diteliti pada tugas akhir ini, yaitu:

1) Bagaimana performa akurasi metode Federated Learning dengan agregasi (Common Federated Learning) dibanding metode Federated Learning tanpa agregasi (Daisy-chain Federated Learning) dalam mengklasifikasi pesan SMS yang berupa *scam* dan *non-scam*?

1.3 Batasan Penelitian

Batasan masalah yang ada pada penelitian ini adalah sebagai berikut.

1) Bahasa yang digunakan untuk pesan SMS adalah bahasa Indonesia di mana *dataset* SMS berbahasa Indonesia relatif berjumlah sedikit.
2) *Pre-training global model* tidak dilakukan terlebih dahulu, melainkan data langsung digunakan untuk Federated Learning.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah melatih sebuah model yang mengklasifikasi pesan SMS menjadi dua kelas yaitu *scam* dan *non-scam* menggunakan teknik Federated Learning dengan agregasi dan membandingkan performanya dengan pelatihan Federated Learning tanpa agregasi.

1.5 Manfaat Penelitian

Manfaat pada penelitian ini adalah sebagai berikut.

1) Memberikan acuan dasar sebelum melakukan implementasi nyata teknik Federated Learning pada perangkat *user* yang sesungguhnya.
2) Dapat menjadi referensi untuk penelitian lebih lanjut terkait topik Natural Language Processing secara Federated Learning terutama yang menggunakan bahasa Indonesia.