BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi *blockchain* telah mendorong transformasi signifikan dalam sistem keuangan digital, yang memiliki mekanisme transaksi yang terdesentralisasi, transparan, dan tidak bergantung pada perantara institusional. Namun di balik hal tersebut, peningkatan aktivitas pada ekosistem ini menyebabkan volume dan kompleksitas transaksi meningkat secara signifikan (FATF, 2020). Ekosistem yang paling berkembang adalah Ethereum, karena tidak hanya berfungsi sebagai jaringan aset kripto tetapi juga sebagai infrastruktur bagi berbagai aplikasi keuangan terdesentralisasi, token digital, dan kontrak pintar (Eurojust and Europol, 2024). Temuan *red flag indicators* menunjukkan bahwa fitur teknologi *blockchain* meningkatkan anonimitas yang digunakan oleh penjahat untuk menyembunyikan sumber dana dan menyamarkan jejak transaksi (FATF, 2020). Oleh karena itu, analisis transaksi *blockchain* semakin banyak digunakan untuk mengidentifikasi alamat yang berpotensi terkait aktivitas illicit sebagai indikator awal dalam upaya pengawasan *anti money laundering* pada ekosistem kripto.

Kegagalan dalam mendeteksi alamat berisiko pada transaksi *blockchain* Ethereum memiliki implikasi signifikan dalam konteks *Anti Money Laundering* (AML), karena dapat memungkinkan aktivitas ilegal tetap berlangsung tanpa terdeteksi, sehingga meningkatkan risiko pencucian uang, pendanaan ilegal, serta penyalahgunaan ekosistem *blockchain*, sekaligus menambah beban operasional melalui keterlambatan investigasi dan kebutuhan pengawasan manual. Kesalahan deteksi berupa *false positive* juga berdampak pada terganggunya aktivitas pengguna yang sah dan menurunnya kepercayaan terhadap sistem. Oleh karena itu, pengembangan model deteksi tidak hanya perlu berfokus pada akurasi, tetapi juga pada keseimbangan dalam mengelola kesalahan klasifikasi, sehingga dapat mendukung mitigasi risiko secara efektif dalam konteks AML.

Penelitian berbasis *machine learning* untuk mendeteksi alamat berisiko pada transaksi Ethereum dilakukan oleh (Farrugia dkk., 2020) dan (Pahuja & Kamal,

2023), menggunakan paradigma pelatihan *centralized learning*. Paradigma pelatihan ini mengumpulkan seluruh data transaksi ke dalam satu lingkungan komputasi untuk melatih model, sehingga memungkinkan pemanfaatan informasi global dan menghasilkan performa deteksi yang tinggi, dengan akurasi mencapai lebih dari 96%. Namun, pendekatan *centralized* memiliki keterbatasan ketika diterapkan dalam konteks dunia nyata, khususnya pada skenario kolaborasi lintas *exchange* atau *platform* pertukaran kripto. Data transaksi pada masing-masing *exchange* umumnya bersifat sensitif dan tidak dapat dibagikan secara bebas karena pertimbangan privasi, regulasi, serta risiko keamanan. Pemusatan data dalam satu sistem dapat menciptakan ketergantungan pada satu otoritas dan mengurangi fleksibilitas dalam mendukung analisis kolaboratif.

Federated Learning menawarkan solusi yaitu paradigma pelatihan pembelajaran terdistribusi yang memungkinkan beberapa *exchange* untuk berkolaborasi dalam melatih model tanpa harus berbagi data mentah. Paradigma pelatihan ini berpotensi mengatasi keterbatasan *centralized learning* dalam hal privasi dan distribusi data (Garst dkk., 2025). Namun, penerapan *federated learning* dalam konteks deteksi alamat berisiko pada aktivitas keuangan ilegal masih menghadapi sejumlah tantangan. Karena sebagian besar penelitian *federated learning* dalam domain AML masih berfokus pada data keuangan tradisional seperti perbankan (Guembe dkk., 2022), sehingga belum sepenuhnya merepresentasikan karakteristik unik transaksi *blockchain*. Selain itu, penerapan *federated learning* juga menimbulkan konsekuensi operasional, seperti peningkatan *overhead* komunikasi dan kompleksitas proses pelatihan, yang berpotensi memengaruhi efisiensi sistem.

Pemilihan algoritma XGBoost dalam penelitian ini didasarkan pada kemampuannya dalam menangani data tabular dengan kompleksitas tinggi serta performanya yang telah terbukti unggul dalam beberapa penelitian deteksi fraud dan anomali, termasuk pada konteks transaksi blockchain (Farrugia dkk., 2020); (Pahuja & Kamal, 2023). XGBoost memiliki karakteristik efisiensi komputasi dan stabilitas pelatihan yang menjadikannya baseline yang kuat untuk dibandingkan dalam skenario centralized maupun federated (Garst dkk., 2025). Penggunaan

XGBoost sebagai dasar model bertujuan untuk mengevaluasi pengaruh arsitektur pembelajaran terhadap kinerja model, tanpa dipengaruhi oleh pemilihan algoritma yang belum teruji pada konteks penelitian ini.

Penggunaan skenario simulasi *multi exchange* bertujuan untuk merepresentasikan kondisi dunia nyata di mana data transaksi tersebar di berbagai *exchange* atau *platform* pertukaran kripto dan tidak dapat dipusatkan secara langsung (Guembe dkk., 2022). Setiap *exchange* memiliki data pengguna dan transaksi yang bersifat privat serta terpisah secara sistem, sehingga pendekatan *federated learning* menjadi relevan untuk mendukung kolaborasi tanpa pelanggaran privasi (Garst dkk., 2025). Simulasi ini dilakukan melalui partisi data secara horizontal ke dalam beberapa entitas yang berperan sebagai *exchange* berbeda. Distribusi data antar *exchange* diasumsikan mengikuti skenario *independent and identically distributed* (IID), dengan tujuan untuk menciptakan kondisi eksperimental yang terkontrol dan meminimalkan pengaruh heterogenitas data terhadap hasil pelatihan model (Jimenez-Gutierrez dkk., 2026). Pendekatan IID dipilih sebagai baseline awal untuk mengevaluasi kinerja *federated* XGBoost dalam kondisi ideal, sehingga perbedaan performa yang diamati dapat lebih jelas dikaitkan dengan perbedaan arsitektur pembelajaran, bukan pada variasi distribusi data. Asumsi distribusi IID tidak sepenuhnya merepresentasikan kondisi dunia nyata yang cenderung non IID (Jimenez-Gutierrez dkk., 2026), sehingga hasil penelitian ini perlu diinterpretasikan sebagai *baseline* empiris sebelum diterapkan pada skenario yang lebih kompleks.

Berdasarkan kesenjangan dari penelitian sebelumnya, penelitian ini berfokus pada penerapan paradigma pelatihan *federated* XGBoost berbasis *sequential federated boosting* untuk deteksi alamat berisiko pada transaksi *blockchain* Ethereum dalam skenario simulasi *multi exchange*, menggunakan pendekatan *centralized learning* sebagai *baseline* untuk menyediakan acuan kinerja optimal. Untuk merepresentasikan kolaborasi lintas *exchange*, data dipartisi secara horizontal ke dalam beberapa entitas dengan asumsi distribusi *independent and identically distributed* (IID) untuk menciptakan lingkungan simulasi yang terkontrol, sehingga perbedaan performa dijelaskan oleh perbedaan arsitektur

pembelajaran. Evaluasi dilakukan secara empiris dan terstruktur dengan dua fokus utama, yaitu membandingkan kinerja model *Federated* XGBoost terhadap *centralized* sebagai baseline, serta menganalisis karakteristik pembelajaran dan implikasi operasionalnya, termasuk dinamika konvergensi dan trade off antara utilitas model dan overhead komunikasi dalam skenario pembelajaran terdistribusi.

1.2 Rumusan Masalah

1. Bagaimana kinerja penerapan *federated* XGBoost dalam mendeteksi alamat berisiko pada transaksi blockchain Ethereum dalam skenario simulasi *multi exchange,* jika dibandingkan dengan pendekatan *centralized* sebagai *baseline*?
2. Bagaimana karakteristik dan implikasi operasional dari penerapan *federated* XGBoost, termasuk *trade off* antara utilitas model dan overhead komunikasi jika dibandingkan dengan pendekatan *centralized* sebagai *baseline*?

1.3 Tujuan Penelitian

1. Melakukan penerapan dan evaluasi kinerja *federated* XGBoost dalam mendeteksi alamat berisiko pada transaksi blockchain Ethereum dalam skenario simulasi *multi exchange*, dengan menggunakan pendekatan *centralized* sebagai baseline pembanding.
2. Melakukan analisis karakteristik pembelajaran dan implikasi operasional dari *federated* XGBoost, termasuk *trade off* antara utilitas model dan overhead komunikasi serta membandingkannya dengan pendekatan *centralized* sebagai *baseline.*

1.4 Manfaat Penelitian

Manfaat Akademik

1. Memberikan kontribusi empiris dalam kajian penerapan *federated learning,* khususnya *federated* XGBoost terhadap analisis transaksi *blockchain* Ethereum untuk deteksi alamat berisiko dalam konteks *Anti Money Laundering* (AML).

2. Memperkaya literatur dengan menyediakan evaluasi terstruktur mengenai kinerja dan karakteristik pembelajaran *federated learning*, termasuk dinamika konvergensi dan implikasi distribusi data dalam skenario simulasi *multi exchange*.
3. Menyediakan *baseline* empiris yang dapat digunakan sebagai acuan dalam penelitian selanjutnya untuk mengkaji kelayakan penerapan pembelajaran terdistribusi pada analisis data *blockchain*, khususnya dalam konteks deteksi alamat berisiko.

Manfaat Praktis

1. Memberikan gambaran mengenai kelayakan penerapan *federated learning* dalam lingkungan analitik *blockchain* yang melibatkan data terdistribusi lintas *exchange* tanpa memerlukan sentralisasi data mentah.
2. Memberikan wawasan bagi pengembang sistem analitik *blockchain* dan platform pertukaran kripto terkait *trade off* antara utilitas model dan *overhead* operasional pada aspek komunikasi dan proses pelatihan sistem *federated*.
3. Menjadi referensi awal dalam memahami implikasi implementasi pembelajaran kolaboratif pada sistem deteksi alamat berisiko, termasuk keterbatasan dan tantangan yang perlu diperhatikan dalam penerapan di lingkungan nyata.

1.5 Batasan Penelitian

Batasan penelitian ini disusun untuk memperjelas aspek-aspek yang tidak dilakukan dalam penelitian, sehingga ruang lingkup penelitian tetap terarah dan sesuai dengan tujuan yang telah ditetapkan. Batasan penelitian ini adalah sebagai berikut.

1. Penelitian ini tidak melakukan pembuktian forensik atau investigasi hukum terhadap aktivitas pencucian uang, melainkan hanya mendeteksi alamat berisiko berdasarkan pola transaksi pada dataset Ethereum.

2. Penelitian ini tidak menggunakan data langsung dari sistem *exchange* nyata, melainkan menggunakan dataset *benchmark* Ethereum yang disimulasikan ke dalam skenario *multi exchange* melalui partisi data.
3. Penelitian ini tidak mengevaluasi skenario distribusi data non IID, karena eksperimen dibatasi pada distribusi IID untuk memperoleh kondisi awal yang terkontrol dalam membandingkan *centralized learning* dan *federated learning*.
4. Penelitian ini tidak membandingkan berbagai algoritma *machine learning*, karena model yang digunakan dibatasi pada XGBoost dan adaptasinya dalam skenario *federated* melalui mekanisme *sequential federated boosting.*
5. Penelitian ini tidak menerapkan mekanisme *privacy preserving* tambahan seperti *differential privacy, homomorphic encryption*, atau *secure aggregation*, karena evaluasi difokuskan pada utilitas model, konvergensi, *runtime*, dan *communication cost.*
6. Penelitian ini tidak mengimplementasikan sistem *federated learning* pada lingkungan produksi, sehingga hasil penelitian diposisikan sebagai evaluasi empiris dalam skenario simulasi terkontrol.

1.6 Sistematika Penulisan

1. Bab I Pendahuluan memaparkan latar belakang penelitian terkait urgensi deteksi alamat berisiko dalam konteks aktivitas ilegal pada *blockchain* Ethereum serta tantangan distribusi data pada skenario *multi exchange*. Bab ini juga memuat rumusan masalah, tujuan, ruang lingkup, manfaat penelitian, batasan penelitian, serta sistematika penulisan.
2. Bab II Landasan Teori menyajikan dasar teoritis yang meliputi konsep deteksi aktivitas ilegal pada Ethereum dalam konteks AML, karakteristik data transaksi *blockchain*, metode klasifikasi berbasis XGBoost, serta paradigma *federated learning* beserta tantangan utamanya seperti heterogenitas data, konvergensi, dan overhead komunikasi. Bab ini juga

mengkaji penelitian terdahulu untuk mengidentifikasi gap dan memperkuat posisi penelitian.

3. Bab III Metodologi Penelitian menjelaskan rancangan eksperimen, meliputi pra-pemrosesan data, partisi horizontal untuk simulasi *multi exchange* berbasis IID, serta implementasi *centralized* XGBoost sebagai *baseline* dan *federated* XGBoost berbasis *sequential boosting*. Bab ini juga mencakup konfigurasi model, mekanisme pelatihan federated, serta skema evaluasi yang mencakup metrik utilitas dan aspek operasional.
4. Bab IV Hasil dan Pembahasan menyajikan hasil eksperimen dan analisis komparatif antara *centralized* dan *federated learning*. Pembahasan difokuskan pada evaluasi utilitas model, stabilitas konvergensi, serta *trade off* antara performa dan *overhead* komunikasi, termasuk analisis implikasi operasional dalam skenario pembelajaran terdistribusi.
5. Bab V Kesimpulan dan Saran berisi kesimpulan yang menjawab rumusan masalah berdasarkan hasil evaluasi utilitas dan operasional, serta saran pengembangan penelitian, seperti eksplorasi skenario non IID, peningkatan skala federasi, dan optimasi efisiensi komunikasi.