BAB IV
HASIL DAN PEMBAHASAN

4.1 Gambaran Umum Implementasi Penelitian

Bab ini menyajikan hasil implementasi dan pembahasan terhadap lima konfigurasi model, yaitu SVM, GCN, GCN + Class Weight, GAT, dan GAT + Class Weight, dalam mendeteksi transaksi ilegal pada jaringan Bitcoin menggunakan Elliptic Bitcoin Dataset. Tahapan yang dilakukan pada bab ini mengikuti metodologi penelitian yang telah dijelaskan pada Bab III, dimulai dari eksplorasi dataset, prapemrosesan data, pembentukan *graph*, pelatihan model, hingga evaluasi hasil klasifikasi.

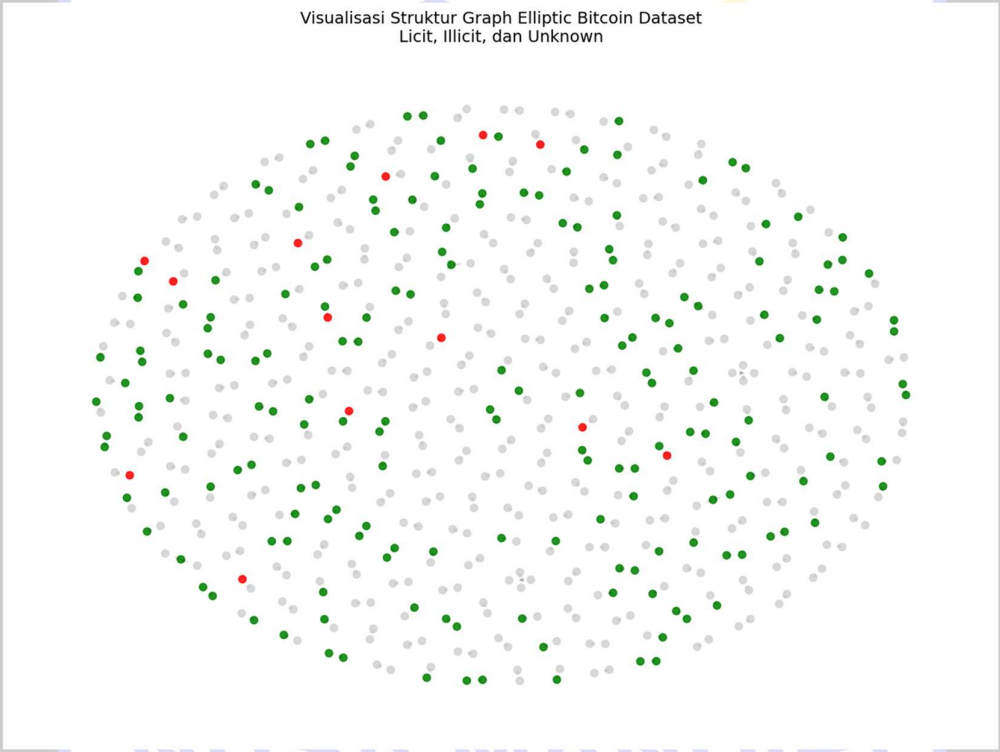
Pada penelitian ini, setiap transaksi *Bitcoin* direpresentasikan sebagai *node*, sedangkan hubungan antartransaksidirepresentasikan sebagai *edge* sehingga membentuk sebuah *graph*. Representasi tersebut memungkinkan model berbasis *graph*, khususnya *Graph Attention Network (GAT),* memanfaatkan tidak hanya informasi dari masing-masing transaksi, tetapi juga hubungan antartransaksi yang terdapat pada jaringan *Bitcoin*. Pendekatan ini diharapkan mampu mengenali pola transaksi ilegal dengan lebih baik dibandingkan metode klasifikasi yang hanya menggunakan fitur transaksi secara individual. Selain menyajikan hasil implementasi model, bab ini juga membahas karakteristik dataset yang digunakan, proses pembentukan *graph*, hasil pelatihan model, serta evaluasi performa menggunakan beberapa metrik klasifikasi. Selanjutnya dilakukan analisis terhadap transaksi *fraud* yang berhasil maupun yang belum berhasil dideteksi oleh model, sehingga dapat diketahui bagaimana struktur *graph* berkontribusi dalam proses deteksi transaksi ilegal.

Sebagai bagian dari pembahasan, hasil penelitian juga dianalisis dalam konteks *digital forensik* untuk melihat bagaimana model *Graph Attention Network (GAT)* dapat membantu proses identifikasi transaksi yang mencurigakan serta

memberikan informasi pendukung bagi proses investigasi pada jaringan *Blockchain*.

4.1.1 Visualisasi Struktur *Graph*

Untuk memperoleh gambaran mengenai karakteristik jaringan transaksi *Bitcoin*, dilakukan visualisasi struktur *graph* berdasarkan hubungan antartransaksi pada Elliptic *Bitcoin* Dataset. Visualisasi ini bertujuan untuk memperlihatkan pola keterhubungan antar*node* sehingga karakteristik jaringan dapat diamati sebelum dilakukan proses pelatihan model berbasis graph, yaitu GCN dan GAT



Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Gambar 4.1 Visualisasi Struktur *Graph* pada Elliptic *Bitcoin* Dataset

Berdasarkan Gambar 4.1, struktur graph pada Elliptic Bitcoin Dataset divisualisasikan dengan tetap mempertahankan tiga kategori transaksi, yaitu Licit, Illicit, dan Unknown. Setiap node merepresentasikan satu transaksi Bitcoin, sedangkan edge menunjukkan adanya hubungan antartransaksi. Dataset secara keseluruhan terdiri atas 203.769 node dan 234.355 edge. Namun, karena jumlah

node dan edge cukup besar, visualisasi dilakukan menggunakan sampel sebanyak 300 edge agar pola hubungan antartransaksi dapat terlihat dengan lebih jelas.

Hasil visualisasi sampel menghasilkan 595 node yang terdiri atas 399 transaksi Unknown, 183 transaksi Licit, dan 13 transaksi Illicit. Jumlah node Unknown terlihat lebih dominan dibandingkan dua kategori lainnya. Hal ini sesuai dengan distribusi dataset, karena sebagian besar transaksi pada Elliptic Bitcoin Dataset memang memiliki label Unknown. Pada tahap pelatihan dan evaluasi model dalam penelitian ini, transaksi Unknown tidak digunakan sebagai target dalam proses pelatihan model karena tidak memiliki label pasti sebagai Licit atau Illicit. Namun, pada tahap analisis struktur graph, transaksi Unknown tetap penting untuk dipertahankan karena transaksi tersebut masih memiliki hubungan dengan node lain di dalam jaringan. Dengan demikian, transaksi Unknown tidak hanya dipandang sebagai data tanpa label, tetapi juga sebagai bagian dari struktur jaringan yang dapat memberikan informasi tambahan mengenai pola keterhubungan transaksi Bitcoin

Tabel 4.1 Ringkasan Hasil Visualisasi Struktur *Graph*

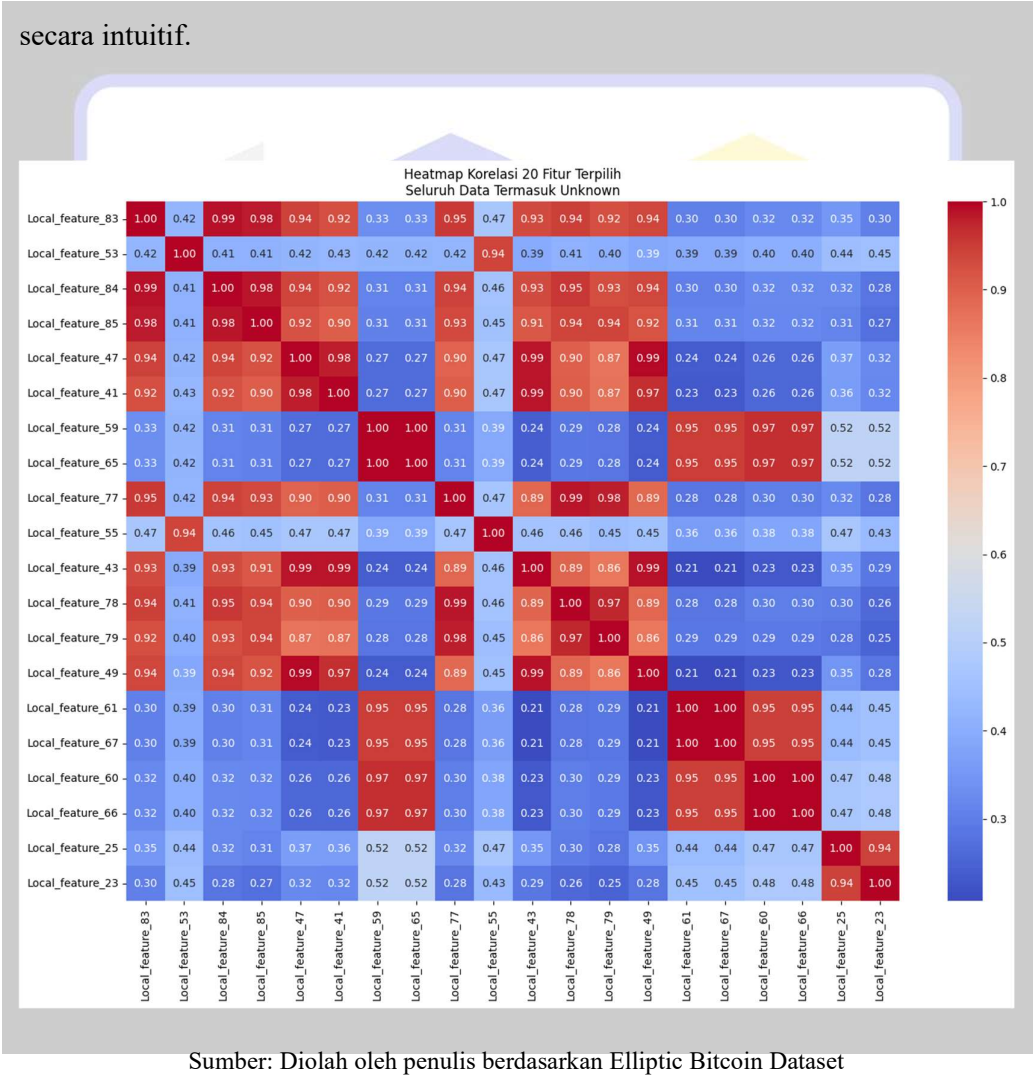
Parameter	Nilai
Jumlah *Node* yang divisualisasikan	595
Jumlah *Edge* yang divisualisasikan	300
Metode Visualisasi	*NetworkX*
Bentuk Data	*Graph*
Jenis Visualisasi	Sampel Sub*graph*

Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Berdasarkan Tabel 4.1, visualisasi *graph* menggunakan 595 *node* dan 300 *edge* sebagai representasi dari struktur *graph* pada Elliptic *Bitcoin* Dataset. Jumlah tersebut merupakan sampel (sub*graph*) dari keseluruhan dataset yang terdiri atas 203.769 *node* dan 234.355 *edge*. Penggunaan sampel dilakukan agar pola hubungan antartransaksi dapat diamati dengan jelas, namun tetap mampu memberikan gambaran mengenai karakteristik struktur graph yang digunakan dalam implementasi model berbasis graph, yaitu GCN dan GAT.

4.1.2 Analisis Korelasi Fitur

Analisis korelasi fitur dilakukan untuk mengetahui tingkat hubungan antar fitur numerik yang terdapat pada Elliptic *Bitcoin* Dataset. Mengingat dataset memiliki 165 fitur, analisis difokuskan pada 20 fitur yang memiliki nilai korelasi rata-rata tertinggi agar pola hubungan antar fitur dapat diamati dengan lebih jelas. Visualisasi dilakukan menggunakan heatmap sehingga tingkat korelasi antar fitur dapat terlihat secara intuitif.



Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Gambar 4.2 Heatmap Korelasi Fitur

Berdasarkan Gambar 4.2, analisis korelasi dilakukan terhadap 165 fitur yang terdapat pada seluruh 203.769 transaksi di dalam Elliptic Bitcoin Dataset. Pada tahap ini, transaksi Unknown tetap disertakan karena analisis korelasi tidak

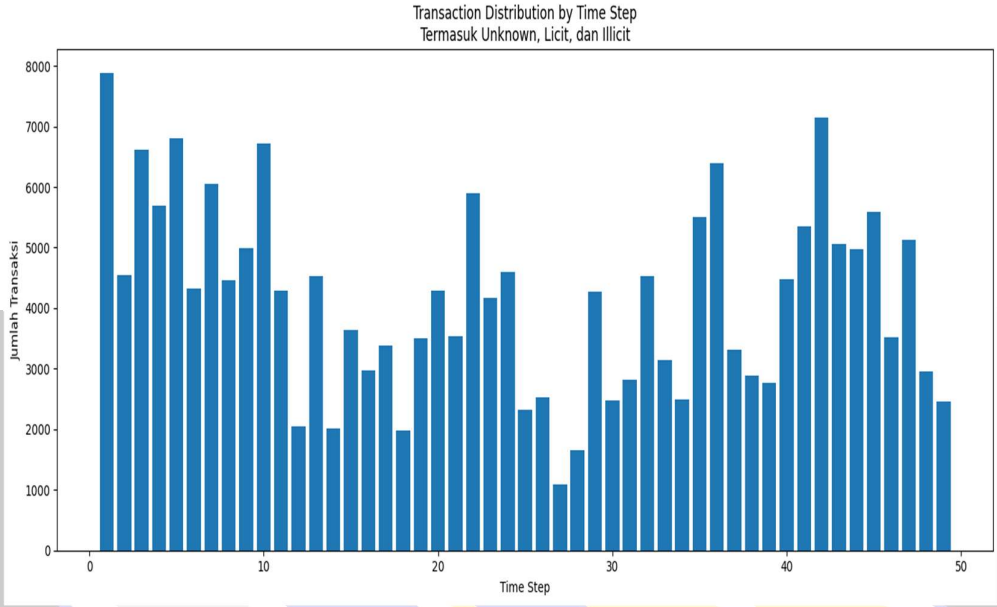
membutuhkan label kelas sebagai target, melainkan hanya menggunakan nilai numerik dari masing-masing fitur transaksi.

Dari keseluruhan 165 fitur, dipilih 20 fitur yang memiliki rata-rata korelasi absolut tertinggi terhadap fitur lainnya. Beberapa fitur yang berada pada urutan tertinggi antara lain Local_feature_83, Local_feature_53, Local_feature_84, Local_feature_85, Local_feature_47, dan Local_feature_41. Pemilihan 20 fitur dilakukan agar hubungan antarfitur dapat ditampilkan lebih jelas dalam bentuk heatmap dan tidak terlalu padat apabila seluruh 165 fitur ditampilkan sekaligus.

Hasil heatmap menunjukkan bahwa terdapat beberapa kelompok fitur yang memiliki hubungan cukup kuat, tetapi tidak seluruh fitur mempunyai korelasi yang tinggi. Kondisi ini menunjukkan bahwa informasi pada dataset tidak hanya bergantung pada satu kelompok fitur tertentu, tetapi berasal dari kombinasi berbagai karakteristik transaksi. Dengan memasukkan transaksi Unknown dalam analisis ini, pola fitur yang dianalisis menjadi lebih mewakili keseluruhan dataset dan tidak hanya terbatas pada transaksi yang telah memiliki label Licit dan Illicit.

4.1.3 Analisis *Time step*

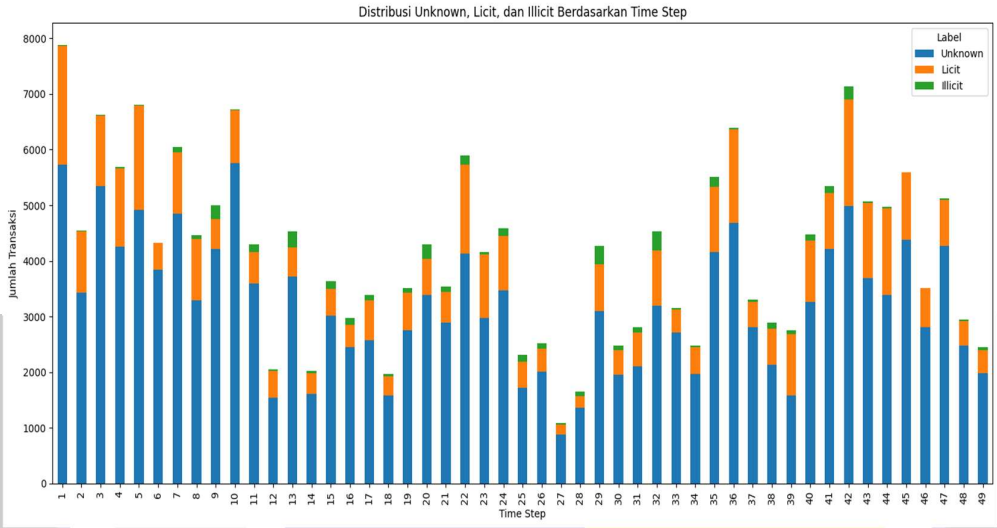
Analisis *time step* dilakukan untuk memahami bagaimana distribusi transaksi *Bitcoin* pada setiap periode pengamatan dalam Elliptic *Bitcoin* Dataset. Dataset ini membagi seluruh transaksi ke dalam 49 *time step* yang merepresentasikan urutan waktu transaksi pada jaringan *Blockchain*. Analisis ini bertujuan untuk mengetahui pola distribusi transaksi secara temporal serta mengidentifikasi apakah terdapat periode tertentu yang memiliki aktivitas transaksi lebih tinggi dibandingkan periode lainnya.



Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Gambar 4.3 Transaction Distribution by *Time step*

Berdasarkan Gambar 4.3, distribusi seluruh transaksi pada Elliptic Bitcoin Dataset tersebar dalam 49 time step. Jumlah transaksi pada setiap time step terlihat berbeda-beda, sehingga aktivitas transaksi pada jaringan Bitcoin tidak terjadi secara merata pada setiap periode pengamatan. Pada analisis terbaru ini, transaksi Unknown tetap disertakan bersama transaksi Licit dan Illicit. Hal tersebut dilakukan karena transaksi *Unknown* juga merupakan bagian dari aktivitas transaksi pada setiap time step, walaupun belum mempunyai label pasti. Dengan memasukkan transaksi *Unknown*, distribusi temporal yang diperoleh dapat menggambarkan keseluruhan aktivitas transaksi yang terdapat pada dataset, bukan hanya transaksi yang telah mempunyai label. Hasil analisis time step menunjukkan bahwa distribusi transaksi, termasuk transaksi Illicit, berubah pada setiap periode pengamatan. Analisis ini digunakan untuk memahami karakteristik temporal dataset, tetapi model GCN dan GAT pada penelitian ini tidak secara khusus memodelkan dependensi temporal. Model utama tetap memanfaatkan fitur transaksi dan struktur hubungan antarnode pada graph

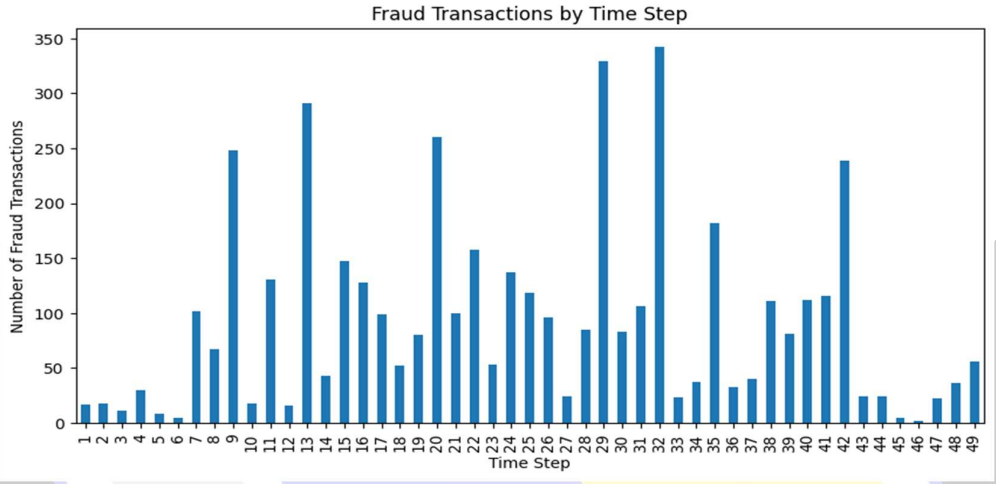


Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Gambar 4.4 Distribusi Label Unknown, Licit, dan Illicit Berdasarkan Time Step

Berdasarkan Gambar 4.4, transaksi Unknown terlihat mendominasi hampir seluruh time step dibandingkan transaksi Licit maupun Illicit. Sebagai contoh, pada time step 1 terdapat 5.733 transaksi Unknown, 2.130 transaksi Licit, dan 17 transaksi Illicit. Pada time step 32 terdapat 3.202 transaksi Unknown, 981 transaksi Licit, dan 342 transaksi Illicit. Dominasi transaksi Unknown menunjukkan bahwa sebagian besar transaksi pada dataset belum mempunyai ground truth yang pasti. Meskipun demikian, transaksi tersebut tetap memiliki fitur dan hubungan dengan transaksi lain pada struktur graph.

Oleh karena itu, transaksi Unknown perlu dianalisis lebih lanjut untuk mengetahui apakah pola fitur dan keterhubungannya lebih mendekati transaksi Licit atau Illicit. Distribusi ini juga menunjukkan bahwa jumlah transaksi Illicit tidak sama pada setiap periode. Pada beberapa time step jumlah transaksi Illicit meningkat, sedangkan pada periode lainnya relatif rendah. Hal tersebut menunjukkan bahwa aktivitas transaksi ilegal pada jaringan Bitcoin memiliki pola temporal yang berubah-ubah



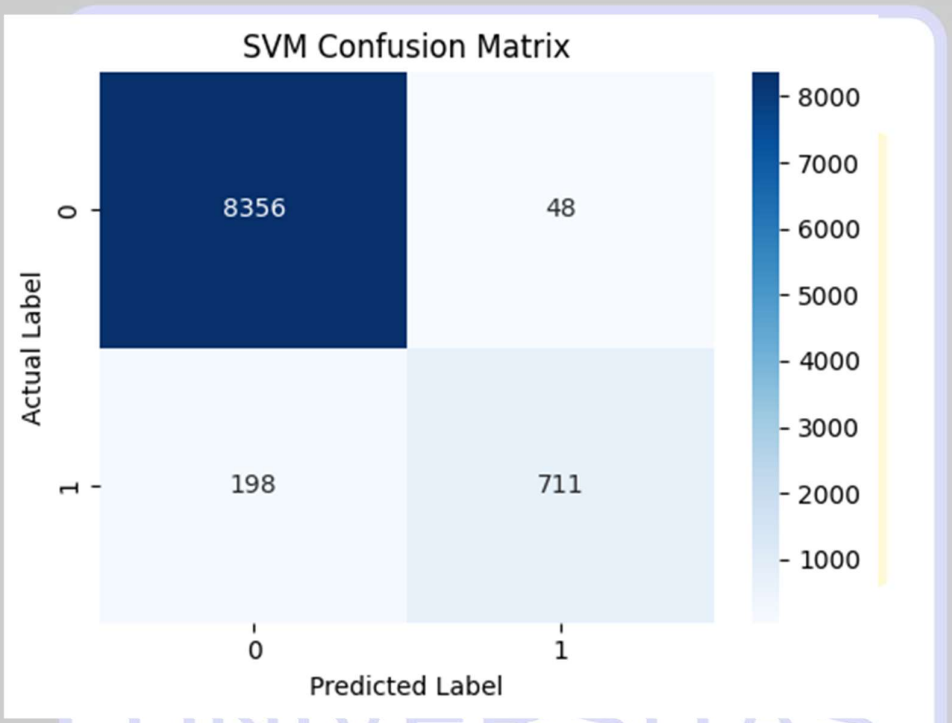
Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Gambar 4.5 *Fraud* Transaction by *Time step*

Berdasarkan Gambar 4.5, distribusi transaksi *fraud* juga tidak merata pada seluruh *time step*. Terlihat bahwa beberapa periode memiliki jumlah transaksi ilegal yang lebih tinggi dibandingkan periode lainnya, sedangkan pada beberapa periode awal maupun akhir jumlah transaksi *fraud* relatif lebih sedikit. Hal ini menunjukkan bahwa aktivitas transaksi ilegal tidak terjadi secara konstan, melainkan cenderung terkonsentrasi pada periode-periode tertentu.

Hasil analisis *time step* menunjukkan bahwa baik distribusi transaksi secara keseluruhan maupun distribusi transaksi *fraud* memiliki pola yang bersifat dinamis. Variasi jumlah transaksi pada setiap periode menunjukkan bahwa karakteristik jaringan *Bitcoin* berubah seiring waktu. Selain itu, adanya konsentrasi transaksi *fraud* pada beberapa *time step* mengindikasikan bahwa aktivitas ilegal memiliki kecenderungan muncul pada periode tertentu. Informasi temporal tersebut menjadi salah satu alasan penggunaan *Graph Convolutional Network (GCN)*, karena model tidak hanya memanfaatkan atribut transaksi, tetapi juga hubungan antartransaksi dalam struktur *graph* sehingga mampu menangkap pola yang sulit dikenali oleh metode klasifikasi berbasis fitur saja.

Setelah itu tahap berikutnya adalah *Confusion Matrix* yang digunakan untuk melihat kemampuan model dalam mengklasifikasikan transaksi *Bitcoin* ke dalam kelas licit dan illicit**.** Dari hasil pengujian dapat diketahui jumlah prediksi yang benar (*True Positive* dan *True Negative*) maupun prediksi yang salah (*False Positive* dan *False Negative*). Semakin tinggi nilai *True Positive* dan *True Negative*, serta semakin rendah nilai *False Positive* dan *False Negative*, maka semakin baik kemampuan model dalam melakukan klasifikasi



Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Gambar 4.6 *Confusion Matrix* SVM

Berdasarkan Gambar 4.6, hasil pengujian model *Support Vector Machine (SVM)* menunjukkan bahwa model mampu mengklasifikasikan sebagian besar transaksi *Bitcoin* dengan baik. Dari hasil *Confusion Matrix* terlihat bahwa terdapat 8.356 transaksi legal yang berhasil diprediksi dengan benar sebagai transaksi legal **(***True Negative*/TN). Hasil ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali transaksi normal sehingga hanya sedikit transaksi legal yang mengalami kesalahan klasifikasi.

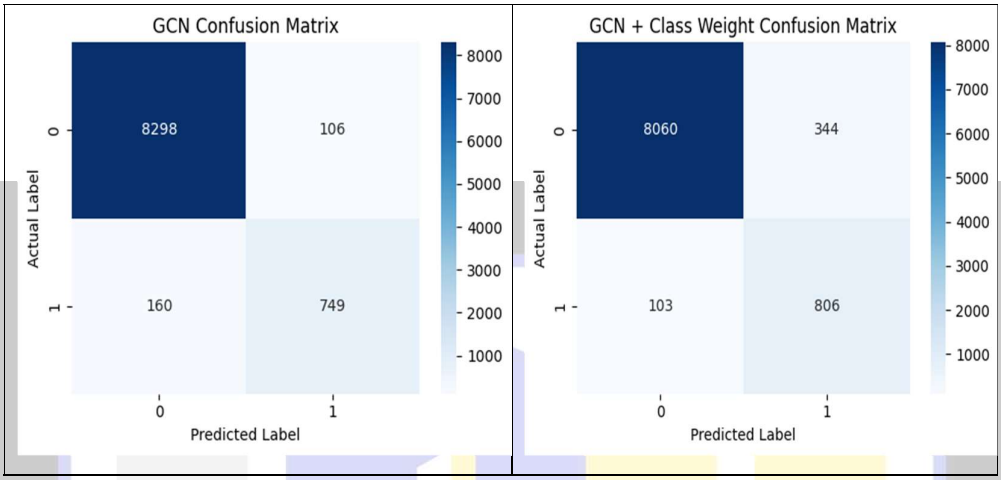
Selain itu, model juga berhasil mendeteksi 711 transaksi ilegal dengan benar (*True Positive*/TP). Hal ini menunjukkan bahwa SVM mampu mengenali pola pada sebagian besar transaksi ilegal yang terdapat pada dataset. Namun demikian, masih terdapat 198 transaksi ilegal yang diprediksi sebagai transaksi legal (*False Negative*/FN). Kesalahan ini menjadi perhatian utama karena transaksi yang seharusnya teridentifikasi justru tidak terdeteksi oleh model.

Di sisi lain, terdapat 48 transaksi legal yang diprediksi sebagai transaksi ilegal (*False Positive*/FP). Jumlah tersebut relatif kecil jika dibandingkan dengan jumlah transaksi legal yang berhasil diklasifikasikan dengan benar. Kesalahan seperti ini dapat menyebabkan adanya transaksi yang sebenarnya tidak bermasalah ikut diperiksa lebih lanjut. Meskipun demikian, dampaknya masih lebih ringan dibandingkan apabila transaksi ilegal tidak berhasil terdeteksi.

Apabila dikaitkan dengan hasil evaluasi model, nilai *Accuracy* sebesar 97,36% menunjukkan bahwa secara keseluruhan SVM memiliki tingkat klasifikasi yang tinggi. Akan tetapi, nilai *Recall* sebesar 78,22% mengindikasikan bahwa model masih belum mampu mendeteksi seluruh transaksi ilegal yang ada pada data pengujian. Hal ini terlihat dari masih adanya 198 transaksi ilegal yang lolos dari proses klasifikasi.

Dalam penelitian ini, *False Negative* merupakan jenis kesalahan yang paling penting untuk diperhatikan. Hal tersebut karena tujuan penelitian adalah mendeteksi transaksi ilegal pada jaringan *Bitcoin* untuk mendukung kebutuhan *digital forensik*. Semakin sedikit nilai *False Negative*, semakin kecil kemungkinan transaksi ilegal lolos dari proses identifikasi sehingga proses investigasi dapat dilakukan dengan lebih efektif. Oleh karena itu, meskipun model SVM menghasilkan akurasi yang tinggi, kemampuannya dalam mendeteksi seluruh transaksi ilegal masih perlu ditingkatkan. Kondisi inilah yang menjadi alasan penggunaan *Graph Attention Network (GAT)* sebagai metode utama dalam penelitian ini, karena GAT mampu memanfaatkan hubungan antartransaksi pada struktur *graph* melalui mekanisme attention sehingga diharapkan dapat

meningkatkan kemampuan deteksi terhadap transaksi ilegal dibandingkan model berbasis *graph* konvensional seperti GCN.



Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Gambar 4.7 *Confusion Matrix* GCN dan GCN + Class Weight

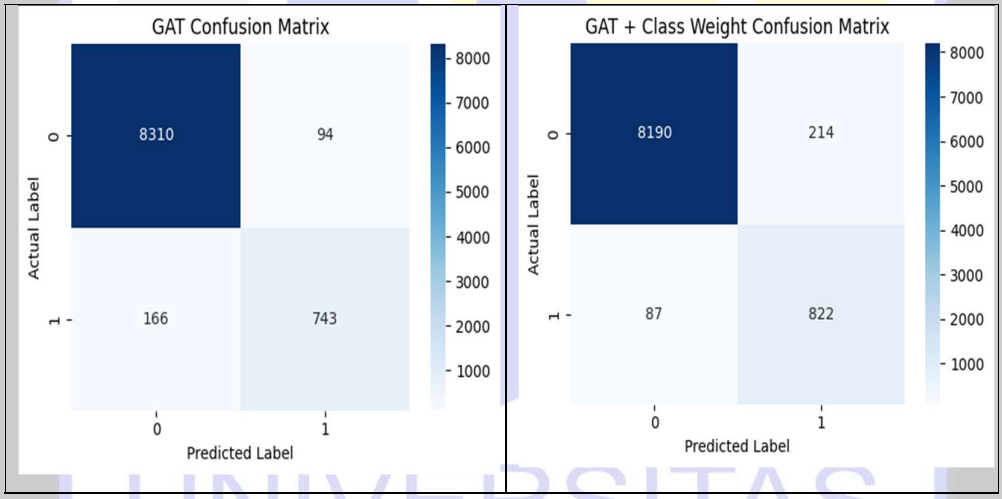
Gambar 4.7 menunjukkan hasil *Confusion Matrix* dari model *Graph Convolutional Network (GCN)* dan GCN + *Class Weight* yang digunakan untuk mendeteksi transaksi ilegal pada jaringan *Bitcoin*. Perbandingan kedua model ini bertujuan untuk mengetahui pengaruh penerapan *Class Weight* dalam mengatasi ketidakseimbangan kelas (*class imbalance*) yang terdapat pada dataset Elliptic, di mana jumlah transaksi ilegal jauh lebih sedikit dibandingkan transaksi legal.

Berdasarkan hasil pengujian, model GCN menghasilkan *True Negative* (TN) sebanyak 8.298, yang berarti sebanyak 8.298 transaksi legal berhasil diklasifikasikan dengan benar sebagai transaksi legal. Selain itu, model juga mampu mendeteksi 749 transaksi ilegal dengan benar atau *True Positive* (TP). Namun, masih terdapat 160 transaksi ilegal yang diprediksi sebagai transaksi legal (*False Negative*/FN) dan 106 transaksi legal yang diprediksi sebagai transaksi ilegal (*False Positive*/FP). Nilai *False Negative* tersebut menunjukkan bahwa masih ada sejumlah transaksi ilegal yang tidak berhasil dikenali oleh model sehingga berpotensi lolos dari proses identifikasi.

Setelah diterapkan *Class Weight*, terlihat adanya perubahan pada hasil klasifikasi. Model GCN + *Class Weight* berhasil meningkatkan jumlah *True Positive* menjadi 806 transaksi, sementara jumlah *False Negative* menurun menjadi 103 transaksi. Hasil ini menunjukkan bahwa pemberian bobot pada kelas minoritas membuat model menjadi lebih sensitif dalam mengenali transaksi ilegal. Dengan demikian, jumlah transaksi ilegal yang berhasil dideteksi menjadi lebih banyak dibandingkan model GCN tanpa *Class Weight*. Di sisi lain, penerapan *Class Weight* juga menyebabkan jumlah *False Positive* meningkat dari 106 menjadi 344 transaksi. Hal ini menunjukkan bahwa model menjadi lebih sering mengklasifikasikan transaksi legal sebagai transaksi ilegal. Kondisi tersebut merupakan konsekuensi dari pemberian bobot yang lebih besar pada kelas minoritas sehingga model cenderung lebih berhati-hati dalam melakukan klasifikasi terhadap transaksi yang memiliki karakteristik mirip dengan transaksi ilegal.

Apabila kedua model dibandingkan secara keseluruhan, terlihat bahwa penggunaan *Class Weight* memberikan peningkatan yang cukup baik terhadap kemampuan model dalam mendeteksi transaksi ilegal. Hal tersebut ditunjukkan dengan bertambahnya jumlah *True Positive* dari 749 menjadi 806 transaksi, serta menurunnya jumlah *False Negative* dari 160 menjadi 103 transaksi. Meskipun jumlah *False Positive* meningkat, perubahan tersebut masih dapat diterima karena tujuan utama penelitian ini adalah meminimalkan jumlah transaksi ilegal yang tidak terdeteksi. Dalam konteks *digital forensik*, *False Negative* merupakan jenis kesalahan yang paling perlu diperhatikan. Transaksi ilegal yang tidak berhasil diidentifikasi berpotensi menyebabkan aktivitas kriminal, seperti pencucian uang atau pendanaan kegiatan ilegal, lolos dari proses investigasi. Sebaliknya, *False Positive* hanya menyebabkan transaksi legal ikut ditandai sebagai transaksi yang mencurigakan dan masih dapat diverifikasi kembali oleh penyidik melalui proses pemeriksaan lanjutan. Oleh karena itu, mengurangi jumlah *False Negative* menjadi prioritas yang lebih penting dibandingkan mempertahankan jumlah *False Positive* yang rendah.

Berdasarkan hasil tersebut, dapat disimpulkan bahwa penerapan *Class Weight* mampu meningkatkan kemampuan model GCN dalam mendeteksi transaksi ilegal pada dataset yang memiliki distribusi kelas tidak seimbang. Meskipun jumlah *False Positive* mengalami peningkatan, penurunan *False Negative* menunjukkan bahwa model menjadi lebih efektif dalam mengenali transaksi ilegal. Hasil ini sejalan dengan tujuan penelitian, yaitu membangun model yang mampu mendeteksi transaksi ilegal pada jaringan *Bitcoin* secara lebih akurat untuk mendukung kebutuhan *digital forensik*. Oleh karena itu, model GCN + *Class Weight* dapat dianggap lebih sesuai digunakan dibandingkan model GCN tanpa *Class Weight* karena mampu mengidentifikasi lebih banyak transaksi ilegal yang berpotensi menjadi objek investigasi.



Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Gambar 4.8 *Confusion Matrix* GAT dan GAT + Class Weight

Gambar 4.8 menunjukkan hasil *Confusion Matrix* dari model *Graph Attention Network (GAT)* dan *GAT + Class Weight* dalam mendeteksi transaksi ilegal pada jaringan *Bitcoin*. Perbandingan kedua model ini dilakukan untuk mengetahui pengaruh penerapan *Class Weight* terhadap kemampuan model dalam mengatasi ketidakseimbangan kelas (*class imbalance*) pada dataset Elliptic. Karena jumlah transaksi ilegal jauh lebih sedikit dibandingkan transaksi legal, pemberian bobot pada kelas minoritas diharapkan dapat meningkatkan kemampuan model dalam mengenali transaksi ilegal.

Berdasarkan hasil pengujian, model GAT menghasilkan *True Negative* (TN) sebanyak 8.310, yang berarti sebanyak 8.310 transaksi legal berhasil diprediksi dengan benar sebagai transaksi legal. Selain itu, model juga mampu mendeteksi 743 transaksi ilegal dengan benar atau *True Positive* (TP). Namun demikian, masih terdapat 166 transaksi ilegal yang diprediksi sebagai transaksi legal (*False Negative*/FN) serta 94 transaksi legal yang diprediksi sebagai transaksi ilegal (*False Positive*/FP). Hasil ini menunjukkan bahwa model GAT sudah memiliki kemampuan klasifikasi yang cukup baik, tetapi masih terdapat sejumlah transaksi ilegal yang belum berhasil diidentifikasi. Setelah diterapkan *Class Weight*, hasil klasifikasi mengalami perubahan. Model GAT + *Class Weight* mampu meningkatkan jumlah *True Positive* menjadi 822 transaksi, sedangkan jumlah *False Negative* menurun menjadi 87 transaksi. Penurunan *False Negative* ini menunjukkan bahwa pemberian bobot pada kelas minoritas membuat model lebih efektif dalam mengenali transaksi ilegal. Dengan demikian, jumlah transaksi ilegal yang berhasil dideteksi meningkat dibandingkan model GAT tanpa *Class Weight*. Di sisi lain, penerapan *Class Weight* juga menyebabkan jumlah *False Positive* meningkat dari 94 menjadi 214 transaksi. Hal ini menunjukkan bahwa model menjadi lebih sensitif terhadap pola transaksi yang dianggap mencurigakan sehingga lebih banyak transaksi legal yang diprediksi sebagai transaksi ilegal. Kondisi tersebut merupakan konsekuensi dari penerapan *Class Weight*, di mana model lebih memprioritaskan pendeteksian kelas minoritas sehingga sebagian transaksi legal ikut terklasifikasi sebagai transaksi yang perlu mendapat perhatian.

Jika kedua model dibandingkan, terlihat bahwa penerapan *Class Weight* memberikan peningkatan yang cukup signifikan terhadap kemampuan model dalam mendeteksi transaksi ilegal. Hal ini ditunjukkan dengan meningkatnya jumlah *True Positive* dari 743 menjadi 822 transaksi, serta berkurangnya jumlah *False Negative* dari 166 menjadi 87 transaksi. Meskipun jumlah *False Positive* mengalami peningkatan, hasil tersebut masih dapat diterima karena tujuan utama penelitian ini adalah mengurangi kemungkinan transaksi ilegal tidak terdeteksi.

Dalam konteks *digital forensik*, *False Negative* merupakan jenis kesalahan yang paling perlu diperhatikan karena transaksi ilegal yang tidak berhasil diidentifikasi

dapat menyebabkan aktivitas kriminal, seperti pencucian uang, pendanaan ilegal, atau tindak kejahatan lainnya, lolos dari proses investigasi. Sebaliknya, *False Positive* hanya menyebabkan transaksi legal ikut ditandai sebagai transaksi yang mencurigakan dan masih dapat diverifikasi kembali melalui proses investigasi lanjutan. Oleh karena itu, penurunan jumlah *False Negative* menjadi indikator yang lebih penting dibandingkan peningkatan *False Positive*.

Berdasarkan hasil tersebut, dapat disimpulkan bahwa penerapan *Class Weight* berhasil meningkatkan kemampuan model GAT dalam mendeteksi transaksi ilegal pada dataset yang memiliki distribusi kelas tidak seimbang. Meskipun jumlah *False Positive* meningkat, penurunan *False Negative* menunjukkan bahwa model menjadi lebih efektif dalam mengenali transaksi ilegal dibandingkan model GAT tanpa *Class Weight*. Hasil ini menunjukkan bahwa penggunaan *Class Weight* mampu meningkatkan sensitivitas model terhadap kelas minoritas sehingga lebih sesuai digunakan untuk mendukung kebutuhan *digital forensik*, di mana kemampuan mendeteksi transaksi ilegal menjadi aspek yang lebih penting daripada meminimalkan kesalahan pada transaksi legal.

Berdasarkan hasil yang diperoleh, model GAT + *Class Weight* bahkan menunjukkan peningkatan yang lebih besar dibandingkan GCN + *Class Weight*, yang terlihat dari penurunan *False Negative* pada GAT dari 166 menjadi 87 transaksi setelah penerapan *Class Weight*, sementara pada GCN penurunannya hanya dari 160 menjadi 103 transaksi. Hal ini menunjukkan bahwa mekanisme attention pada GAT mampu memanfaatkan informasi hubungan antartransaksi dengan lebih baik ketika dikombinasikan dengan *Class Weight*. Dengan demikian, model ini memiliki potensi yang lebih tinggi dalam mendukung proses identifikasi transaksi ilegal pada jaringan *Bitcoin* untuk kebutuhan digital forensik.

Hasil pengujian menunjukkan bahwa penerapan *Class Weight* memberikan pengaruh terhadap kemampuan model dalam mengenali transaksi *Illicit*. Pada GAT + Class Weight, kelas *Illicit* diberikan bobot sebesar 5,1225, jauh lebih besar dibandingkan bobot kelas *Licit* sebesar 0,5541. Pemberian bobot tersebut menyebabkan kesalahan klasifikasi terhadap transaksi *Illicit* menghasilkan penalti

loss yang lebih besar selama proses pelatihan. Akibatnya, model menjadi lebih sensitif dalam mengenali kelas *Illicit*. Pengaruh tersebut terlihat dari peningkatan kemampuan deteksi kelas minoritas. Model GAT + Class Weight memperoleh *recall* sebesar 90,43%, lebih tinggi dibandingkan GAT tanpa Class Weight. Selain itu, GAT + Class Weight menghasilkan *Macro Recall* sebesar 0,9394. Hasil tersebut menunjukkan bahwa penerapan *Class Weight* membantu model mengurangi kecenderungan untuk terlalu dominan memprediksi kelas mayoritas.

Namun, peningkatan sensitivitas terhadap kelas *Illicit* juga menimbulkan *trade-off*. Peningkatan *recall* diikuti dengan penurunan *precision* karena lebih banyak transaksi *Licit* yang ikut diprediksi sebagai *Illicit*. Dalam konteks penelitian ini, kondisi tersebut masih dapat diterima karena tujuan utama model adalah mengurangi *False Negative*, yaitu transaksi ilegal yang tidak berhasil terdeteksi. Dalam proses digital forensik, transaksi yang terindikasi ilegal masih dapat diperiksa lebih lanjut oleh investigator, sedangkan transaksi ilegal yang tidak terdeteksi berpotensi tidak masuk ke dalam proses investigasi.

Tabel 4.2 Perbandingan Hasil *Confusion Matrix* Seluruh Model

Model	TN	FP	FN	TP
SVM	8356	48	198	711
GCN	8298	106	160	749
GCN + *Class Weight*	8060	344	103	806
GAT	8310	94	166	743
GAT + *Class Weight*	8190	214	**87**	**822**

Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Tabel 4.2 menyajikan perbandingan hasil *Confusion Matrix* dari lima model yang digunakan dalam penelitian, yaitu *Support Vector Machine (SVM)*, *Graph Convolutional Network (GCN)*, GCN + *Class Weight*, *Graph Attention Network* (GAT), dan GAT + *Class Weight*. Perbandingan ini dilakukan untuk melihat kemampuan masing-masing model dalam mengklasifikasikan transaksi legal dan transaksi ilegal berdasarkan nilai *True Negative* (TN), *False Positive* (FP), *False Negative* (FN), dan *True Positive* (TP).

Berdasarkan hasil pengujian, model SVM menghasilkan nilai *True Negative* sebanyak 8.356, *False Positive* sebanyak 48, *False Negative* sebanyak 198, dan *True Positive* sebanyak 711. Hasil tersebut menunjukkan bahwa SVM memiliki kemampuan yang sangat baik dalam mengenali transaksi legal karena menghasilkan nilai *False Positive* yang paling rendah. Namun, model ini masih memiliki kelemahan dalam mendeteksi transaksi ilegal, yang terlihat dari nilai *False Negative* sebesar 198. Artinya, masih terdapat cukup banyak transaksi ilegal yang tidak berhasil dikenali oleh model.

Model GCN menunjukkan peningkatan dibandingkan SVM. Jumlah *False Negative* berhasil diturunkan menjadi 160 transaksi, sedangkan jumlah *True Positive* meningkat menjadi 749 transaksi. Hasil ini menunjukkan bahwa pemanfaatan struktur *graph* mampu membantu model mengenali hubungan antartransaksi sehingga kemampuan mendeteksi transaksi ilegal menjadi lebih baik dibandingkan pendekatan berbasis data tabular seperti SVM.

Penerapan *Class Weight* pada model GCN memberikan perubahan yang cukup signifikan. Nilai *False Negative* kembali menurun menjadi 103 transaksi, sedangkan *True Positive* meningkat menjadi 806 transaksi. Hasil tersebut menunjukkan bahwa pemberian bobot pada kelas minoritas berhasil meningkatkan sensitivitas model terhadap transaksi ilegal. Akan tetapi, peningkatan tersebut juga menyebabkan nilai *False Positive* bertambah menjadi 344 transaksi. Hal ini menunjukkan bahwa model menjadi lebih sering mengklasifikasikan transaksi legal sebagai transaksi ilegal karena lebih memprioritaskan pendeteksian kelas minoritas.

Pada model GAT, diperoleh nilai *True Negative* sebesar 8.310, *False Positive* sebanyak 94, *False Negative* sebanyak 166, dan *True Positive* sebanyak 743. Secara umum, performa GAT hampir setara dengan GCN, meskipun jumlah transaksi ilegal yang berhasil dideteksi masih sedikit lebih rendah dibandingkan GCN. Hasil ini menunjukkan bahwa mekanisme attention pada GAT sudah mampu memanfaatkan hubungan antar*node*, tetapi belum memberikan peningkatan yang signifikan ketika belum dikombinasikan dengan teknik penanganan ketidakseimbangan kelas.

Performa terbaik diperoleh oleh model *GAT + Class Weight*. Model ini menghasilkan *True Positive* sebanyak 822 transaksi dan *False Negative* hanya sebesar 87 transaksi, yang merupakan nilai terendah di antara seluruh model yang diuji. Hal tersebut menunjukkan bahwa GAT + *Class Weight* memiliki kemampuan paling baik dalam mengenali transaksi ilegal. Walaupun nilai *False Positive* meningkat menjadi 214 transaksi, jumlah tersebut masih lebih rendah dibandingkan GCN + *Class Weight* yang menghasilkan 344 *False Positive*. Dengan demikian, model GAT + *Class Weight* mampu memberikan keseimbangan yang lebih baik antara peningkatan deteksi transaksi ilegal dan jumlah kesalahan klasifikasi pada transaksi legal.

Apabila dibandingkan secara keseluruhan, terlihat bahwa penggunaan metode berbasis *graph* memberikan hasil yang lebih baik dibandingkan model SVM. Selain itu, penerapan *Class Weight* pada model GCN maupun GAT terbukti mampu mengurangi jumlah *False Negative* secara signifikan. Penurunan *False Negative* merupakan hasil yang sangat penting karena menunjukkan semakin sedikit transaksi ilegal yang lolos dari proses klasifikasi.

Dalam konteks *digital forensik*, *False Negative* merupakan jenis kesalahan yang paling perlu diminimalkan. Transaksi ilegal yang tidak berhasil dideteksi berpotensi menyebabkan aktivitas kriminal, seperti pencucian uang, pendanaan ilegal, maupun tindak kejahatan lainnya, tidak teridentifikasi selama proses investigasi. Sebaliknya, *False Positive* hanya menyebabkan transaksi legal ikut ditandai sebagai transaksi yang mencurigakan dan masih dapat diverifikasi kembali melalui proses pemeriksaan lanjutan oleh investigator. Oleh karena itu, model dengan nilai *False Negative* yang lebih rendah dinilai lebih sesuai untuk mendukung kebutuhan *digital forensik*.

Berdasarkan hasil perbandingan pada Tabel 4.2, dapat disimpulkan bahwa model GAT + *Class Weight* merupakan model dengan performa terbaik dalam penelitian ini. Model tersebut berhasil menghasilkan jumlah *True Positive* tertinggi dan *False Negative* terendah dibandingkan model lainnya. Hasil ini menunjukkan bahwa kombinasi mekanisme *Graph Attention Network* dengan *Class Weight* mampu meningkatkan kemampuan model dalam mendeteksi transaksi ilegal pada

dataset yang memiliki distribusi kelas tidak seimbang. Dengan demikian, model GAT + *Class Weight* dinilai paling efektif untuk mendukung proses identifikasi transaksi ilegal pada jaringan *Bitcoin* dalam kebutuhan *digital forensik*.

Tabel 4.3 Hasil Evaluasi Model

Model	*Accuracy* (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM	**97,36**	**93,68**	78,22	**85,25**
GCN	97,14	87,60	82,40	84,92
GCN + *Class Weight*	95,20	70,09	88,67	78,29
GAT	97,21	88,77	81,74	85,11
GAT + *Class Weight*	96,77	79,34	**90,43**	84,52

Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Berdasarkan Tabel 4.3, dapat diketahui bahwa setiap model memiliki keunggulan pada metrik evaluasi yang berbeda. Hal ini menunjukkan bahwa tidak ada satu model yang unggul pada seluruh aspek, sehingga pemilihan model perlu disesuaikan dengan tujuan penelitian, yaitu mendeteksi transaksi ilegal pada jaringan *Bitcoin*. Model SVM memperoleh akurasi tertinggi sebesar 97,36%, dengan *Precision* 93,68% dan *F1-score* 85,25%, yang juga merupakan nilai tertinggi dibandingkan model lainnya. Hasil ini menunjukkan bahwa SVM sangat baik dalam mengklasifikasikan data secara keseluruhan dan menghasilkan prediksi transaksi ilegal dengan tingkat ketepatan yang tinggi. Namun, nilai *recall* sebesar 78,22% merupakan yang terendah di antara seluruh model, sehingga masih terdapat cukup banyak transaksi ilegal yang tidak berhasil terdeteksi (*False Negative*).

Model GCN menghasilkan akurasi sebesar 97,14%, yang hanya sedikit lebih rendah dibandingkan SVM. Meskipun *Precision* GCN menurun menjadi 87,60%, nilai *recall* meningkat menjadi 82,40%. Peningkatan *recall* ini menunjukkan bahwa GCN mampu menemukan lebih banyak transaksi ilegal dibandingkan SVM. Hal tersebut mengindikasikan bahwa pemanfaatan informasi hubungan antartransaksi pada struktur *graph* memberikan kontribusi dalam meningkatkan kemampuan deteksi transaksi ilegal.

Pada model GCN + *Class Weight*, terjadi perubahan karakteristik performa. Nilai *recall* meningkat cukup signifikan menjadi 88,67%, namun diikuti penurunan akurasi menjadi 95,20%, *Precision* menjadi 70,09%, dan *F1-score* menjadi 78,29%. Kondisi ini menunjukkan bahwa pemberian bobot kelas berhasil meningkatkan kemampuan model dalam mendeteksi transaksi ilegal, tetapi juga menyebabkan jumlah prediksi positif yang salah (*False Positive*) menjadi lebih banyak. Model GAT memperoleh akurasi sebesar 97,21%, dengan *Precision* 88,77%, *recall* 81,74%, dan *F1-score* 85,11%. Secara keseluruhan, performa GAT cukup seimbang dan mendekati GCN. Mekanisme attention pada GAT mampu memberikan hasil klasifikasi yang baik, meskipun peningkatan *recall* dibandingkan GCN tidak terlalu signifikan.

Sementara itu, GAT + *Class Weight* menghasilkan *recall* tertinggi sebesar 90,43%, yang berarti model ini paling baik dalam mendeteksi transaksi ilegal dan paling sedikit melewatkan transaksi ilegal dibandingkan model lainnya. Meskipun demikian, peningkatan *recall* tersebut diikuti penurunan *Precision* menjadi 79,34% dan akurasi menjadi 96,77%. Hal ini menunjukkan adanya *trade-off* antara kemampuan mendeteksi transaksi ilegal dan meningkatnya jumlah *False Positive*.

Berdasarkan keseluruhan hasil evaluasi, apabila akurasi dan *Precision* dijadikan sebagai acuan utama, maka SVM merupakan model dengan performa terbaik. Namun, apabila penelitian lebih memprioritaskan kemampuan mendeteksi sebanyak mungkin transaksi ilegal, maka GAT + *Class Weight* menjadi model yang lebih sesuai karena memiliki nilai *recall* tertinggi sebesar 90,43%. Mengingat penelitian ini berfokus pada deteksi transaksi ilegal dalam konteks *digital forensik*, metrik *recall* menjadi lebih penting dibandingkan akurasi, karena kegagalan mendeteksi transaksi ilegal (*False Negative*) dapat menyebabkan aktivitas kriminal pada jaringan *Blockchain* tidak teridentifikasi. Dengan demikian, dapat disimpulkan bahwa penambahan *Class Weight* pada model berbasis *Graph Neural Network* mampu meningkatkan kemampuan model dalam mendeteksi transaksi ilegal yang ditunjukkan oleh peningkatan nilai *recall*. Meskipun terjadi penurunan pada nilai akurasi dan *Precision*, peningkatan *recall* menjadi aspek yang lebih penting dalam penelitian ini karena tujuan utama adalah meminimalkan jumlah

False Negative, sehingga semakin sedikit transaksi ilegal yang lolos dari proses deteksi.

Pemilihan Recall sebagai metrik yang lebih diprioritaskan juga berkaitan dengan tujuan penggunaan model dalam mendukung digital forensik. Pada hasil pengujian, GAT + Class Weight memperoleh Recall tertinggi sebesar 90,43%. Nilai tersebut menunjukkan bahwa model mampu mendeteksi sebagian besar transaksi Illicit yang terdapat pada data pengujian. Meskipun Precision yang dihasilkan lebih rendah dibandingkan beberapa model lainnya, peningkatan Recall menunjukkan bahwa jumlah transaksi Illicit yang terlewat atau False Negative dapat ditekan. Dalam proses investigasi, kondisi ini lebih diutamakan karena transaksi yang terdeteksi sebagai mencurigakan masih dapat diperiksa kembali oleh investigator, sedangkan transaksi Illicit yang tidak terdeteksi berpotensi tidak masuk ke proses pemeriksaan lebih lanjut. Pada penelitian ini, Recall menjadi metrik yang lebih diprioritaskan karena tujuan utama penelitian adalah mendeteksi sebanyak mungkin transaksi Illicit. Recall menunjukkan kemampuan model dalam menemukan transaksi Illicit dari seluruh transaksi Illicit yang sebenarnya terdapat pada data pengujian. Nilai Recall yang tinggi menunjukkan bahwa jumlah transaksi Illicit yang tidak berhasil terdeteksi atau False Negative (FN) semakin sedikit. Dalam konteks digital forensik, False Negative menjadi perhatian karena transaksi Illicit yang tidak terdeteksi berpotensi tidak masuk ke tahap pemeriksaan lebih lanjut. Sementara itu, False Positive masih dapat dilakukan verifikasi kembali pada tahap analisis oleh investigator. Oleh karena itu, penelitian ini lebih memprioritaskan Recall untuk menekan False Negative, tetapi tetap mempertimbangkan Precision, F1-score, Accuracy, dan AUC dalam menilai performa model secara keseluruhan.

Tabel 4.4 Hasil Evaluasi Model Menggunakan *Macro Average* dan *Micro Average*

Model	Macro Precision	Macro Recall	Macro F1	Micro Precision	Micro Recall	Micro F1
SVM	0,9568	0,8882	0,9190	0,9736	0,9736	0,9736
GCN	0,9286	0,9057	0,9167	0,9714	0,9714	0,9714
GCN + Class Weight	0,8441	0,9229	0,8780	0,9520	0,9520	0,9520
GAT	0,9341	0,9031	0,9178	0,9721	0,9721	0,9721
GAT + Class Weight	0,8915	0,9394	0,9136	0,9677	0,9677	0,9677

Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Tabel 4.4 menyajikan hasil evaluasi kelima model menggunakan pendekatan *Macro Average* dan *Micro Average*. Evaluasi ini digunakan untuk memberikan gambaran performa model secara lebih menyeluruh, terutama karena dataset yang digunakan memiliki distribusi kelas yang tidak seimbang antara transaksi *Licit* dan *Illicit*. *Macro Average* menghitung nilai *Precision*, *Recall*, dan *F1-score* pada setiap kelas secara terpisah, kemudian menghitung rata-ratanya tanpa mempertimbangkan jumlah data pada masing-masing kelas. Oleh karena itu, *Macro Average* dapat memberikan gambaran mengenai kemampuan model dalam mengenali kedua kelas secara lebih seimbang, termasuk kelas minoritas yaitu transaksi *Illicit*.

Sementara itu, *Micro Average* menghitung seluruh hasil klasifikasi dari semua kelas secara bersama-sama sebelum menghitung nilai *Precision*, *Recall*, dan *F1-score*. Pada klasifikasi *binary single-label* yang digunakan dalam penelitian ini, nilai *Micro Precision*, *Micro Recall*, dan *Micro F1-score* memiliki nilai yang sama dengan *Accuracy*. Oleh karena itu, nilai *Micro Average* lebih menggambarkan performa klasifikasi secara keseluruhan dan dapat lebih dipengaruhi oleh jumlah data pada kelas mayoritas.Berdasarkan hasil evaluasi *Macro Average*, SVM memperoleh nilai *Macro Precision* tertinggi sebesar 0,9568 dan *Macro F1-score* tertinggi sebesar 0,9190. Hasil tersebut menunjukkan bahwa SVM memiliki keseimbangan yang baik antara *Precision* dan *Recall* ketika performa kedua kelas dihitung secara rata-rata. Model GAT memperoleh *Macro Precision* sebesar 0,9341, *Macro Recall* sebesar 0,9031, dan *Macro F1-score* sebesar 0,9178, sehingga memiliki nilai *Macro F1-score* yang mendekati SVM. Sementara itu,

GCN memperoleh *Macro Precision* sebesar 0,9286, *Macro Recall* sebesar 0,9057, dan *Macro F1-score* sebesar 0,9167.

Penerapan *Class Weight* memberikan perubahan terhadap karakteristik performa model. Pada GCN + *Class Weight*, nilai *Macro Recall* meningkat menjadi 0,9229, sedangkan *Macro Precision* menjadi 0,8441 dan *Macro F1-score* menjadi 0,8780. Pada GAT + *Class Weight*, nilai *Macro Recall* meningkat menjadi **0,9394**, yang merupakan nilai tertinggi di antara seluruh model. Namun, *Macro Precision* model tersebut menjadi 0,8915 dengan *Macro F1-score* sebesar 0,9136. Hasil ini menunjukkan adanya *trade-off*, yaitu peningkatan kemampuan model dalam mengenali kelas secara lebih sensitif diikuti dengan penurunan ketepatan prediksi pada sebagian kasus.

Pada evaluasi *Micro Average*, SVM memperoleh nilai tertinggi sebesar 0,9736, diikuti oleh GAT sebesar 0,9721, GCN sebesar 0,9714, GAT + *Class Weight* sebesar 0,9677, dan GCN + *Class Weight* sebesar 0,9520. Nilai tersebut menunjukkan bahwa SVM memiliki performa klasifikasi keseluruhan tertinggi berdasarkan *Micro Average*. Meskipun demikian, *Micro Average* tidak menjadi satu-satunya dasar dalam menentukan model utama karena penelitian ini berfokus pada deteksi transaksi *Illicit* yang merupakan kelas minoritas.

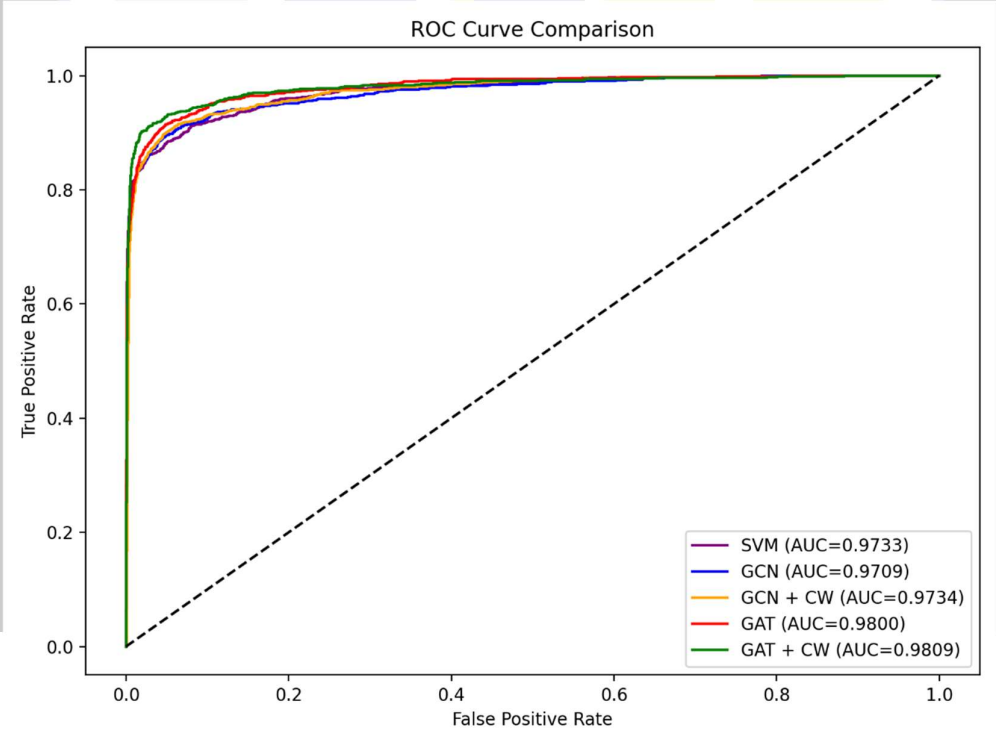
Secara keseluruhan, hasil evaluasi menunjukkan bahwa setiap model memiliki keunggulan pada metrik yang berbeda. SVM memperoleh nilai *Macro Precision*, *Macro F1-score*, dan *Micro F1-score* tertinggi, sedangkan GAT + *Class Weight* memperoleh *Macro Recall* tertinggi sebesar 0,9394**.** Hasil ini menunjukkan bahwa penggunaan *Class Weight* meningkatkan sensitivitas model terhadap kelas minoritas, meskipun diikuti oleh penurunan beberapa metrik lainnya. Dalam konteks penelitian ini, GAT + *Class Weight* tetap dipilih sebagai model utama karena tujuan penelitian lebih memprioritaskan kemampuan mendeteksi transaksi *Illicit* dan mengurangi jumlah *False Negative*, bukan hanya memperoleh nilai akurasi atau *Micro Average* tertinggi.

Tabel 4.5 Perkembangan Nilai *Training* Loss per *Epoch*

Model	Epoch 0	Epoch 50	Epoch 100	Epoch 150
GCN	0,5638	0,1126	0,0891	0,0748
GCN + Class Weight	0,7824	0,1926	0,1290	0,0943
GAT	0,8777	0,1369	0,0965	0,0706
GAT + Class Weight	0,7985	0,2460	0,1074	0,0581

Sumber: Diolah oleh Penulis (2026), berdasarkan output eksekusi kode pelatihan model pada Lampiran tesis (variabel losses, dicetak tiap 50 epoch).

Selain menggunakan nilai *training loss*, performa model juga dievaluasi menggunakan *Receiver Operating Characteristic (ROC) Curve*. Kurva ROC digunakan untuk menggambarkan kemampuan model dalam membedakan transaksi legal dan transaksi ilegal pada berbagai nilai ambang (*threshold*). Perbandingan kurva ROC antar model memberikan gambaran mengenai model yang memiliki kemampuan diskriminasi paling baik.



Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Gambar 4.9 Perbandingan Kurva ROC

Berdasarkan Gambar 4.9, kurva ROC digunakan untuk membandingkan kemampuan masing-masing model dalam membedakan transaksi legal dan transaksi ilegal. Semakin mendekati sudut kiri atas, maka semakin baik kemampuan model dalam melakukan klasifikasi. Garis putus-putus diagonal menunjukkan performa klasifikasi secara acak (*random classifier*), sehingga kurva yang berada jauh di atas garis tersebut menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik.

Hasil pengujian menunjukkan bahwa seluruh model memiliki nilai ROC-AUC di atas 0,97. Nilai ROC-AUC tertinggi diperoleh oleh GAT + Class Weight sebesar 0,9809, diikuti oleh GAT yang memiliki nilai sekitar 0,9800. Perbedaan keduanya relatif kecil, tetapi hasil ini menunjukkan bahwa penambahan Class Weight tetap mampu mempertahankan kemampuan diskriminasi model sekaligus meningkatkan Recall terhadap kelas Illicit.

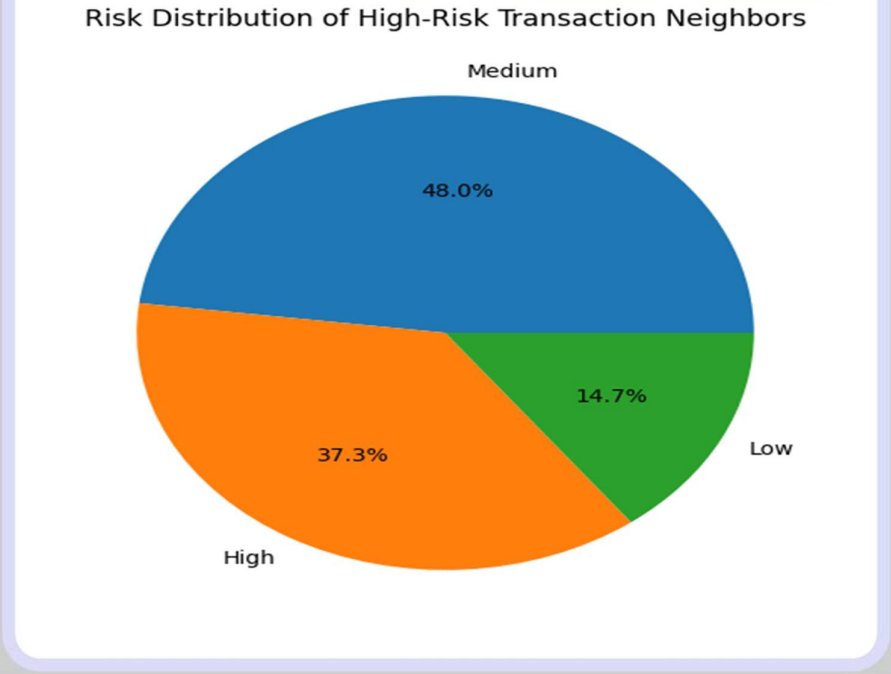
Sementara itu, model SVM memperoleh nilai AUC sebesar 0,9733, sedangkan GCN memperoleh nilai AUC sebesar 0,9709. Kedua model tersebut tetap menunjukkan performa yang sangat baik karena memiliki nilai AUC di atas 0,97. Namun, dibandingkan model berbasis *Graph Neural Network*, kemampuan keduanya dalam membedakan transaksi legal dan ilegal masih sedikit lebih rendah.

Secara keseluruhan, perbedaan nilai AUC antar model relatif kecil, sehingga seluruh model dapat dikatakan memiliki kemampuan klasifikasi yang sangat baik. Akan tetapi, model berbasis *Graph Neural Network*, khususnya GAT dan GAT + *Class Weight*, menunjukkan performa yang lebih unggul karena mampu memanfaatkan informasi struktur *graph* pada jaringan *Bitcoin* secara lebih efektif. Hasil ini mendukung temuan pada evaluasi sebelumnya bahwa penggunaan *Graph Neural Network*, terutama dengan penambahan *Class Weight*, tidak hanya memberikan kemampuan klasifikasi yang baik, tetapi juga meningkatkan keberhasilan dalam mendeteksi transaksi ilegal, yang merupakan tujuan utama penelitian ini.

4.2 Analisis Hasil Deteksi Transaksi *Illicit*

4.2.1 Analisis Karakteristik dan Pola Transaksi *Illicit*

Selain mengevaluasi performa model, penelitian ini juga menganalisis distribusi tingkat risiko *neighbor* pada transaksi *Bitcoin* menggunakan metode yang telah dijelaskan pada subbab 3.10. Analisis ini dilakukan terhadap 5.054 transaksi *neighbor* yang terhubung langsung dengan 4.545 transaksi illicit melalui 8.145 *edge*, untuk mengetahui proporsi transaksi yang termasuk dalam kategori risiko rendah (*Low*), sedang (*Medium*), dan tinggi (*High*). Informasi tersebut memberikan gambaran mengenai tingkat risiko transaksi berdasarkan hubungan dengan transaksi lain dalam jaringan *Bitcoin*. Hasil visualisasi distribusi tingkat risiko *neighbor* dapat dilihat pada Gambar 4.10.



Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Gambar 4.10 Distribusi Tingkat Risiko *Neighbor*

Gambar 4.10 menunjukkan distribusi tingkat risiko *neighbor* pada transaksi *Bitcoin* berdasarkan hasil analisis terhadap hubungan antar*node* dalam jaringan *Blockchain*. Setiap transaksi dikelompokkan ke dalam tiga kategori risiko, yaitu *High*, *Medium*, dan *Low*, berdasarkan karakteristik *neighbor* yang dimilikinya.

Berdasarkan hasil visualisasi, kategori *Medium* memiliki proporsi terbesar, yaitu sebesar 48,0%. Hasil ini menunjukkan bahwa sebagian besar transaksi berada pada tingkat risiko menengah, sehingga masih memerlukan perhatian lebih lanjut karena berpotensi memiliki hubungan dengan transaksi yang mencurigakan. Kategori *High* menempati urutan kedua dengan persentase 37,3%, yang menunjukkan bahwa lebih dari sepertiga transaksi memiliki tingkat risiko tinggi dan berpotensi terlibat dalam aktivitas ilegal. Sementara itu, kategori *Low* hanya sebesar 14,7%, yang mengindikasikan bahwa jumlah transaksi dengan risiko rendah relatif lebih sedikit dibandingkan dua kategori lainnya.

Distribusi tersebut menunjukkan bahwa sebagian besar transaksi dalam jaringan *Bitcoin* memiliki tingkat risiko menengah hingga tinggi. Kondisi ini mengindikasikan bahwa informasi hubungan antartransaksi (*neighbor*) dapat memberikan kontribusi penting dalam proses deteksi transaksi ilegal. Dengan memanfaatkan struktur *graph*, model tidak hanya menganalisis karakteristik setiap transaksi secara individual, tetapi juga mempertimbangkan keterkaitannya dengan transaksi lain yang berada di sekitarnya. Hasil ini mendukung penggunaan *Graph Neural Network* dalam penelitian ini, karena mampu memanfaatkan informasi struktur jaringan untuk meningkatkan efektivitas deteksi transaksi yang berpotensi bersifat *fraud*.

Tabel 4.6 Transaksi Unknown dengan Jumlah Hubungan Tertinggi terhadap Transaksi Illicit

txId	Time Step	Illicit → Unknown	Total Koneksi
355174807	32	68	68
339502544	13	63	63
139273997	28	59	59
355238405	32	58	58
73160850	42	49	49
246979389	15	36	36
246979279	15	29	29
179039621	9	27	27
73282397	42	25	25
94654151	7	22	22

Berdasarkan hasil analisis hubungan antartransaksi, ditemukan sebanyak 1.458 edge dengan arah dari transaksi Illicit menuju transaksi Unknown dan 3.993 edge dengan arah dari transaksi Unknown menuju transaksi Illicit. Setelah dilakukan penggabungan berdasarkan txId, diperoleh sebanyak 4.027 transaksi Unknown yang memiliki hubungan langsung dengan setidaknya satu transaksi Illicit.

Jumlah hubungan setiap transaksi Unknown dengan transaksi Illicit bervariasi. Hasil statistik menunjukkan bahwa rata-rata transaksi Unknown memiliki sekitar 1,35 hubungan dengan transaksi Illicit, dengan nilai median sebesar 1 hubungan. Hal ini menunjukkan bahwa sebagian besar transaksi Unknown hanya memiliki satu keterhubungan langsung dengan transaksi Illicit. Namun, terdapat beberapa transaksi yang memiliki jumlah hubungan jauh lebih tinggi, bahkan mencapai 68 hubungan. Sebagai contoh, transaksi Unknown dengan txId 355174807 pada time step 32 memiliki 68 hubungan langsung dengan transaksi Illicit. Selain itu, txId 339502544 memiliki 63 hubungan dan txId 139273997 memiliki 59 hubungan. Jumlah hubungan yang tinggi tidak dapat digunakan untuk langsung menyimpulkan bahwa transaksi Unknown tersebut merupakan transaksi ilegal. Namun, pola keterhubungan ini dapat digunakan sebagai indikator struktural untuk menentukan transaksi yang perlu mendapatkan perhatian lebih lanjut dalam proses investigasi digital forensik.

4.2.2 Ringkasan Hasil Deteksi *Fraud*

Setelah seluruh proses pelatihan, pengujian, dan analisis hasil dilakukan, diperoleh beberapa temuan utama yang menjadi ringkasan dari penelitian ini. Ringkasan tersebut disusun berdasarkan hasil evaluasi model menggunakan berbagai metrik, seperti *Accuracy*, *Recall*, *ROC-AUC*, serta hasil *Confusion Matrix*. Ringkasan hasil penelitian disajikan pada Tabel 4.7.

Tabel 4.7 Ringkasan Hasil Deteksi *Fraud*

Aspek Analisis	Hasil
Model Terbaik berdasarkan *Recall*	GAT + *Class Weight*
Model Terbaik berdasarkan *Accuracy*	SVM
Nilai *Recall* Tertinggi	90,43%
Nilai AUC Tertinggi	0,9809
False Negative Terendah	87
True Positive Tertinggi	822

Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Berdasarkan Tabel 4.7, dapat disimpulkan bahwa setiap model memiliki keunggulan pada aspek evaluasi yang berbeda. Model SVM memperoleh nilai *Accuracy* tertinggi sebesar 97,36%, yang menunjukkan bahwa model ini memiliki kemampuan terbaik dalam mengklasifikasikan transaksi secara keseluruhan. Namun, nilai *Accuracy* yang tinggi tidak selalu menunjukkan kemampuan terbaik dalam mendeteksi transaksi ilegal, terutama pada dataset yang memiliki distribusi kelas tidak seimbang. Model GAT + *Class Weight* menunjukkan performa terbaik dalam mendeteksi transaksi ilegal dengan memperoleh nilai *recall* tertinggi sebesar 90,43% dan nilai AUC tertinggi sebesar 0,9809. Selain itu, model ini juga menghasilkan jumlah *False Negative* terendah, yaitu sebanyak 87 transaksi, serta *True Positive* tertinggi, yaitu sebanyak 822 transaksi. Hasil tersebut menunjukkan bahwa model GAT + *Class Weight* mampu mengidentifikasi lebih banyak transaksi ilegal dan meminimalkan jumlah transaksi ilegal yang tidak terdeteksi dibandingkan model lainnya.

Secara keseluruhan, hasil ringkasan ini menunjukkan bahwa penggunaan *Graph Attention Network* yang dikombinasikan dengan *Class Weight* memberikan peningkatan kemampuan dalam mendeteksi transaksi *fraud* pada jaringan *Bitcoin*. Meskipun model SVM memiliki akurasi yang sedikit lebih tinggi, model GAT + *Class Weight* lebih sesuai dengan tujuan penelitian karena mampu meningkatkan keberhasilan deteksi transaksi ilegal sekaligus mengurangi jumlah *False Negative*. Oleh karena itu, *GAT + Class Weight* dipilih sebagai model utama dalam penelitian ini karena paling sesuai dengan tujuan penelitian, yaitu meningkatkan kemampuan

deteksi transaksi Illicit. Pemilihan tersebut didasarkan terutama pada Recall tertinggi dan jumlah *False Negative* terendah, bukan karena model tersebut unggul pada seluruh metrik evaluasi..

Tabel 4.8 Hasil Prediksi GAT + Class Weight terhadap Transaksi Unknown

Prediksi	Jumlah	Persentase
Licit	119.100	75,76%
Illicit	38.105	24,24%
Total	157.205	100%

Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Setelah model GAT + Class Weight selesai dilatih menggunakan transaksi yang memiliki label Licit dan Illicit, model selanjutnya digunakan untuk melakukan inferensi terhadap 157.205 transaksi Unknown. Berdasarkan hasil prediksi, sebanyak 119.100 transaksi atau 75,76% memiliki pola yang lebih mendekati kelas Licit. Sementara itu, sebanyak 38.105 transaksi atau 24,24% diprediksi memiliki pola yang lebih mendekati kelas Illicit.

Hasil ini menunjukkan bahwa sebagian besar transaksi Unknown memiliki karakteristik yang lebih mendekati transaksi Licit. Namun, terdapat sekitar seperempat dari keseluruhan transaksi Unknown yang memperoleh probabilitas lebih tinggi pada kelas Illicit. Kelompok ini dapat digunakan sebagai hasil penyaringan awal untuk menentukan transaksi yang perlu diperiksa lebih lanjut. Perlu ditegaskan bahwa transaksi Unknown yang diprediksi sebagai Illicit tidak dapat langsung dinyatakan sebagai transaksi ilegal, karena data Unknown tidak memiliki ground truth yang dapat digunakan untuk memastikan label sebenarnya. Prediksi model hanya menunjukkan kecenderungan pola berdasarkan fitur dan hubungan graph yang telah dipelajari dari transaksi Licit dan Illicit.

Tabel 4.9 Ringkasan Analisis Transaksi Unknown untuk Digital Forensik

Hasil Analisis	Jumlah
Total Unknown	157.205
Diprediksi Illicit	38.105
Terhubung langsung dengan Illicit	4.027
Diprediksi Illicit dan terhubung dengan Illicit	**1.942**

Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Hasil prediksi model kemudian dikombinasikan dengan analisis struktur graph. Dari 157.205 transaksi Unknown, ditemukan sebanyak 4.027 transaksi yang memiliki hubungan langsung dengan transaksi Illicit. Ketika informasi tersebut digabungkan dengan hasil prediksi GAT + Class Weight, diperoleh sebanyak 1.942 transaksi Unknown yang diprediksi sebagai Illicit dan pada saat yang sama memiliki hubungan langsung dengan transaksi Illicit.

Kelompok sebanyak 1.942 transaksi tersebut dapat dipertimbangkan sebagai kandidat yang memiliki tingkat prioritas lebih tinggi untuk diperiksa lebih lanjut. Hal ini disebabkan karena transaksi tersebut tidak hanya memiliki pola fitur yang menurut model lebih mendekati kelas Illicit, tetapi juga mempunyai keterhubungan langsung dengan transaksi yang telah memiliki label Illicit di dalam graph.

Meskipun demikian, hasil ini tetap bersifat sebagai indikator awal dan bukan bukti final bahwa transaksi tersebut merupakan transaksi ilegal. Dalam konteks digital forensik, hasil model dapat digunakan untuk mempersempit jumlah transaksi yang perlu dianalisis oleh investigator sehingga pemeriksaan dapat dilakukan secara lebih terarah.

Tabel 4.10 Transaksi Unknown dengan Probabilitas Illicit Tertinggi

No.	txId	Time Step	Prob. Licit	Prob. Illicit	Prediksi
1	100832039	35	8,814314E-10	1,000000	Illicit
2	99997463	42	3,701203E-09	1,000000	Illicit
3	386161499	31	5,057577E-08	1,000000	Illicit
4	66665998	38	1,733060E-08	1,000000	Illicit
5	386134758	31	1,105979E-08	1,000000	Illicit
6	277797469	24	1,186307E-08	1,000000	Illicit
7	139227905	28	1,396797E-09	1,000000	Illicit
8	379046510	23	5,008599E-08	1,000000	Illicit
9	139227935	28	5,384553E-08	1,000000	Illicit
10	166521556	29	8,140619E-08	1,000000	Illicit

Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Berdasarkan Tabel 4.10, sepuluh transaksi Unknown dengan nilai probabilitas Illicit tertinggi seluruhnya memperoleh nilai probabilitas yang sangat mendekati 1. Nilai tersebut menunjukkan bahwa berdasarkan pola fitur transaksi yang telah dipelajari oleh model GAT + Class Weight, transaksi-transaksi tersebut memiliki karakteristik yang sangat mendekati pola transaksi Illicit. Sebaliknya, nilai probabilitas pada kelas Licit sangat kecil, yaitu berada pada kisaran 10^{-10} sampai 10^{-8}

Sebagai contoh, transaksi dengan txId 100832039 pada time step 35 memperoleh probabilitas Licit sebesar $8{,}814314 \times 10^{-10}$ dan probabilitas Illicit sebesar 1, sehingga model memprediksi transaksi tersebut sebagai Illicit. Pola yang sama juga terlihat pada sembilan transaksi lainnya yang memiliki probabilitas Illicit sangat tinggi. Meskipun nilai probabilitas model sangat tinggi, hasil tersebut tidak dapat langsung digunakan untuk menyatakan bahwa transaksi Unknown tersebut benar-benar merupakan transaksi ilegal. Hal ini karena transaksi Unknown tidak memiliki label aktual atau ground truth yang dapat digunakan untuk memastikan kelas sebenarnya. Oleh karena itu, hasil prediksi ini lebih tepat digunakan sebagai indikator awal atau skor prioritas untuk membantu menentukan transaksi yang perlu diperiksa lebih lanjut dalam proses digital forensik.

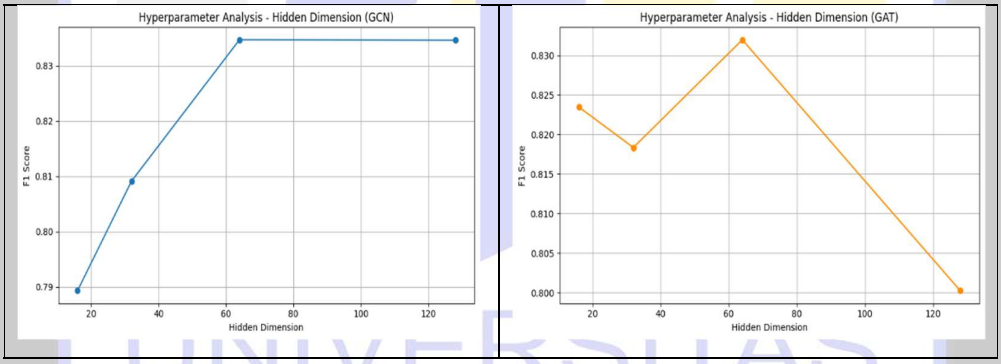
Dalam penelitian ini, hasil probabilitas tersebut kemudian dapat dikombinasikan dengan informasi struktur graph, khususnya hubungan transaksi Unknown dengan

transaksi yang telah berlabel Illicit. Dengan demikian, transaksi yang memiliki probabilitas Illicit tinggi sekaligus mempunyai keterhubungan langsung dengan transaksi Illicit dapat diprioritaskan untuk proses analisis lebih lanjut oleh investigator.

4.3 Analisis Struktur *Graph*

4.3.1 Analisis *Hyperparameter* dan *Embedding* GCN & GAT

Selain mengevaluasi performa model menggunakan berbagai metrik evaluasi, penelitian ini juga melakukan analisis terhadap salah satu *hyperparameter* penting pada model berbasis *graph*, yaitu *hidden dimension*, yang diuji pada model *Graph Convolutional Network (GCN)* dan *Graph Attention Network* (GAT). Analisis ini bertujuan untuk mengetahui pengaruh jumlah *hidden dimension* terhadap kemampuan model dalam mendeteksi transaksi ilegal pada jaringan *Bitcoin*. Hasil analisis ditunjukkan pada Gambar 4.11.



Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

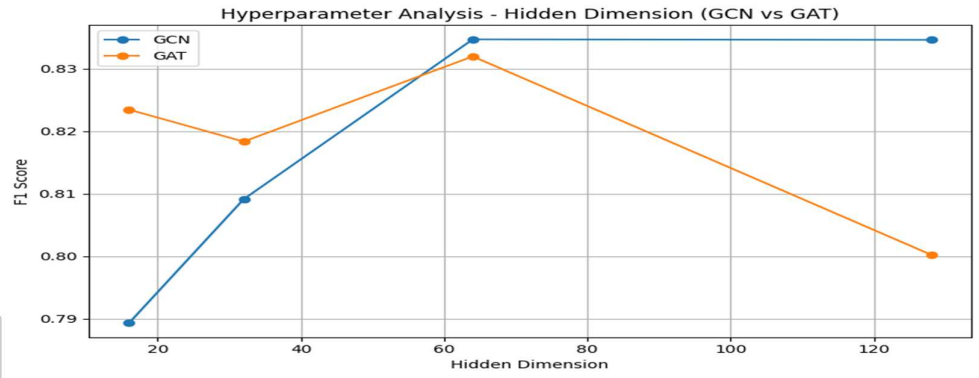
Gambar 4.11 Analisis Pengaruh *Hidden dimension* GCN dan GAT

Berdasarkan Gambar 4.11, dilakukan analisis pengaruh jumlah *hidden dimension* terhadap performa model GCN dan GAT menggunakan metrik *F1-score*. Pengujian dilakukan pada beberapa nilai *hidden dimension*, yaitu 16, 32, 64, dan 128, untuk mengetahui konfigurasi yang memberikan hasil terbaik pada proses deteksi transaksi *fraud*. Pada model GCN, terlihat bahwa nilai *F1-score* mengalami peningkatan seiring bertambahnya jumlah *hidden dimension*. Saat *hidden dimension* sebesar 16, model menghasilkan *F1-score* sekitar 0,79. Nilai tersebut kemudian meningkat menjadi sekitar 0,81 pada *hidden dimension* 32 dan mencapai

sekitar 0,835 pada *hidden dimension* 64. Ketika *hidden dimension* ditingkatkan menjadi 128, nilai *F1-score* tidak mengalami peningkatan yang berarti dan cenderung tetap. Hal ini menunjukkan bahwa penggunaan *hidden dimension* sebesar 64 sudah mampu menghasilkan representasi fitur yang optimal sehingga penambahan dimensi yang lebih besar tidak memberikan peningkatan performa yang signifikan. Sementara itu, pada model GAT terlihat pola yang sedikit berbeda. Nilai *F1-score* mengalami penurunan dari *hidden dimension* 16 ke 32, kemudian meningkat dan mencapai nilai tertinggi sekitar 0,832 pada *hidden dimension* 64. Namun, ketika *hidden dimension* dinaikkan menjadi 128, nilai *F1-score* justru mengalami penurunan yang cukup besar hingga mendekati 0,80. Kondisi ini menunjukkan bahwa penggunaan *hidden dimension* yang terlalu besar pada GAT dapat menyebabkan model menjadi kurang optimal dalam melakukan generalisasi terhadap data pengujian.

Secara keseluruhan, kedua model menunjukkan bahwa *hidden dimension* sebesar 64 merupakan konfigurasi yang memberikan performa terbaik. Hasil ini mengindikasikan bahwa jumlah *hidden dimension* yang terlalu kecil belum mampu menangkap informasi penting dari struktur *graph*, sedangkan penggunaan *hidden dimension* yang terlalu besar tidak selalu meningkatkan performa model dan bahkan dapat menyebabkan penurunan akurasi klasifikasi akibat kompleksitas model yang semakin tinggi. Oleh karena itu, *hidden dimension* sebesar 64 dipilih sebagai konfigurasi terbaik dan digunakan pada proses pelatihan model dalam penelitian ini.

UNIVERSITAS NUSA MANDIRI



Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Gambar 4.12 Analisis Pengaruh *Hidden dimension* GCN vs GAT

Berdasarkan Gambar 4.12, dilakukan analisis pengaruh nilai *hidden dimension* terhadap performa model GCN dan GAT menggunakan metrik *F1-score*. *Hidden dimension* merupakan salah satu *hyperparameter* yang menentukan jumlah neuron pada lapisan tersembunyi (hidden layer) dalam proses pembelajaran *Graph Neural Network*. Pengujian dilakukan menggunakan empat nilai *hidden dimension*, yaitu 16, 32, 64, dan 128, untuk mengetahui konfigurasi yang menghasilkan performa terbaik dalam mendeteksi transaksi *fraud* pada jaringan *Bitcoin*. Pada model GCN, terlihat bahwa nilai *F1-score* mengalami peningkatan seiring bertambahnya *hidden dimension*. Saat *hidden dimension* sebesar 16, model menghasilkan *F1-score* sekitar 0,79. Nilai tersebut meningkat menjadi sekitar 0,81 pada *hidden dimension* 32, kemudian mencapai nilai tertinggi sekitar 0,835 pada *hidden dimension* 64. Ketika *hidden dimension* dinaikkan menjadi 128, nilai *F1-score* tidak mengalami peningkatan lagi dan cenderung tetap. Hal ini menunjukkan bahwa *hidden dimension* sebesar 64 sudah mampu menghasilkan representasi fitur yang optimal sehingga penambahan dimensi yang lebih besar tidak memberikan peningkatan performa yang signifikan. Berbeda dengan GCN, model GAT menunjukkan pola yang lebih fluktuatif. Pada *hidden dimension* 16, model memperoleh *F1-score* sekitar 0,823. Nilai tersebut sedikit menurun pada *hidden dimension* 32, kemudian meningkat dan mencapai nilai terbaik sekitar 0,832 pada *hidden dimension* 64. Namun, ketika *hidden dimension* ditingkatkan menjadi 128, nilai *F1-score* mengalami penurunan hingga sekitar 0,80. Hasil ini menunjukkan

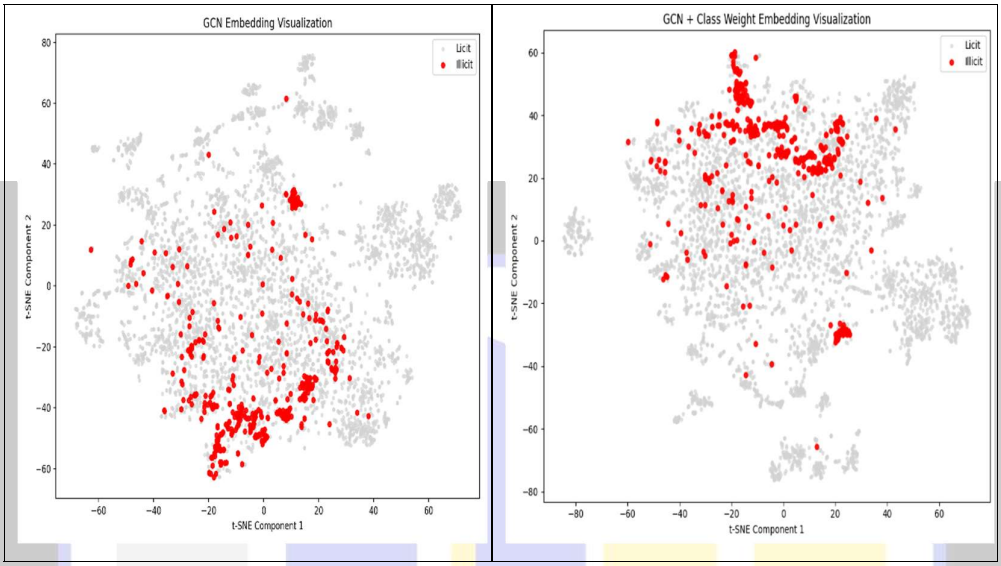
bahwa penggunaan *hidden dimension* yang terlalu besar pada GAT dapat menyebabkan penurunan kemampuan model dalam melakukan generalisasi terhadap data pengujian.

Jika dibandingkan secara keseluruhan, kedua model sama-sama mencapai performa terbaik pada *hidden dimension* 64. Namun, GCN menunjukkan performa yang lebih stabil setelah mencapai nilai tersebut, sedangkan GAT mengalami penurunan performa ketika *hidden dimension* diperbesar menjadi 128. Hal ini menunjukkan bahwa GCN lebih konsisten terhadap perubahan nilai *hidden dimension* dibandingkan GAT. Berdasarkan hasil analisis tersebut, dapat disimpulkan bahwa *hidden dimension* sebesar 64 merupakan konfigurasi yang paling optimal untuk kedua model. Nilai tersebut mampu menghasilkan keseimbangan antara kemampuan model dalam mempelajari representasi fitur pada *graph* dan kompleksitas model, sehingga memberikan performa deteksi transaksi *fraud* yang lebih baik dibandingkan konfigurasi *hidden dimension* lainnya. Oleh karena itu, *hidden dimension* sebesar 64 dipilih sebagai parameter yang digunakan pada proses pelatihan model dalam penelitian ini.

Analisis *hidden dimension* dilakukan pada model dasar GCN dan GAT, karena *Class Weight* hanya memengaruhi bobot pada fungsi loss dan tidak mengubah arsitektur maupun kapasitas representasi model, sehingga hasil *hidden dimension* optimal pada GCN dan GAT juga berlaku untuk varian GCN + *Class Weight* dan GAT + *Class Weight*.

Selain melakukan evaluasi menggunakan metrik *Accuracy*, *Precision*, *recall*, dan *F1-score*, penelitian ini juga melakukan visualisasi *embedding* untuk melihat bagaimana model memetakan representasi setiap transaksi ke dalam ruang fitur *(feature space).* Visualisasi *embedding* bertujuan untuk mengetahui sejauh mana model mampu membedakan transaksi legal (licit) dan transaksi ilegal (illicit) berdasarkan representasi yang dipelajari selama proses pelatihan. Pada visualisasi ini digunakan metode *t-Distributed Stochastic Neighbor Embedding (t-SNE)* untuk mereduksi dimensi fitur sehingga hasil pembelajaran model dapat ditampilkan

dalam bentuk dua dimensi. Perbandingan hasil visualisasi *embedding* antara model GCN dan GCN + *Class Weight* ditunjukkan pada Gambar 4.13.



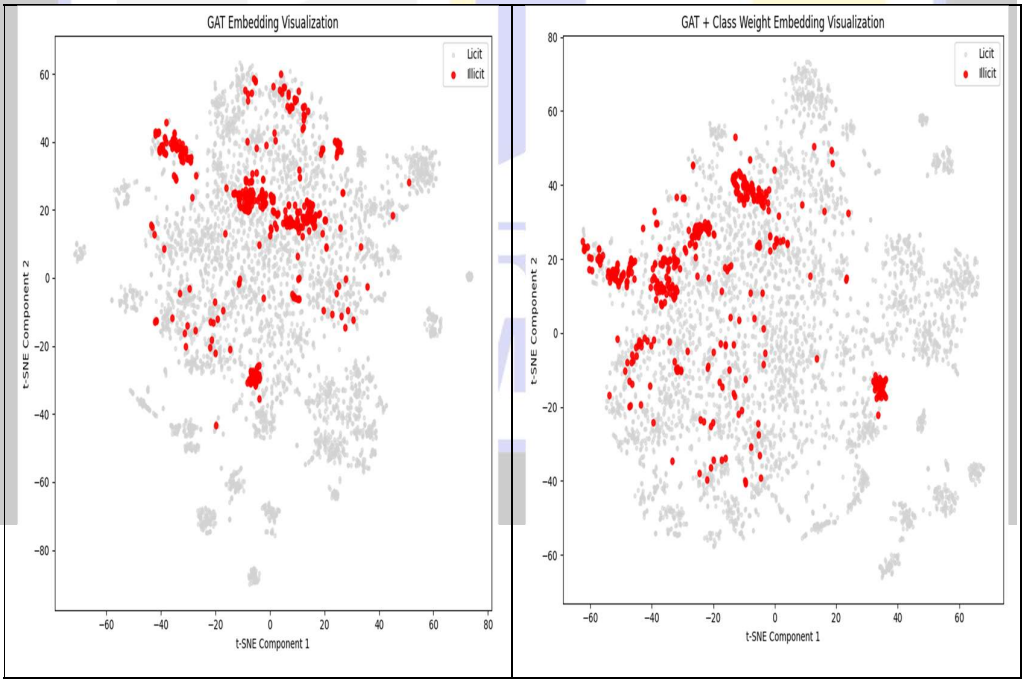
Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Gambar 4.13 Perbandingan Visualisasi *Embedding* Transaksi Menggunakan GCN dan GCN + Class Weight

Berdasarkan Gambar 4.13, terlihat bahwa model GCN dan GCN + *Class Weight* menghasilkan pola distribusi *embedding* yang berbeda. Pada model GCN, titik-titik yang merepresentasikan transaksi ilegal (ditunjukkan dengan warna merah) masih banyak tersebar di antara transaksi legal (warna abu-abu). Kondisi ini menunjukkan bahwa model GCN masih mengalami kesulitan dalam memisahkan karakteristik transaksi ilegal secara lebih jelas, sehingga masih terdapat kemungkinan terjadinya kesalahan klasifikasi. Sementara itu, pada model GCN + *Class Weight* terlihat bahwa transaksi ilegal membentuk kelompok (*cluster*) yang lebih jelas dibandingkan model GCN. Penyebaran titik berwarna merah menjadi lebih terfokus pada beberapa area tertentu, sehingga pemisahan antara transaksi legal dan transaksi ilegal menjadi lebih mudah diamati. Hasil ini menunjukkan bahwa penambahan *Class Weight* membantu model memberikan perhatian yang lebih besar terhadap kelas minoritas, sehingga representasi fitur yang dihasilkan menjadi lebih baik. Meskipun masih terdapat beberapa titik transaksi ilegal yang berada di sekitar kelompok transaksi legal, secara keseluruhan model GCN + *Class Weight*

mampu menghasilkan pemisahan kelas yang lebih baik dibandingkan GCN. Hasil visualisasi ini sejalan dengan hasil evaluasi sebelumnya yang menunjukkan bahwa penambahan *Class Weight* meningkatkan nilai *recall* dan mengurangi jumlah *False Negative*. Dengan demikian, representasi *embedding* yang lebih terstruktur menunjukkan bahwa model lebih mampu mengenali pola transaksi ilegal pada jaringan *Bitcoin*.

Setelah dilakukan analisis pada model GCN, penelitian ini juga membandingkan hasil visualisasi *embedding* yang dihasilkan oleh model GAT dan GAT + *Class Weight*. Visualisasi *embedding* digunakan untuk melihat bagaimana kedua model memetakan representasi setiap transaksi ke dalam ruang fitur setelah proses pelatihan selesai. Dengan menggunakan metode *t-Distributed Stochastic Neighbor Embedding* (t-SNE), representasi berdimensi tinggi direduksi menjadi dua dimensi sehingga pola penyebaran transaksi legal (licit) dan transaksi ilegal (illicit) dapat diamati secara lebih jelas. Perbandingan hasil visualisasi *embedding* pada kedua model ditunjukkan pada Gambar 4.14.



Sumber: Diolah oleh penulis berdasarkan Elliptic Bitcoin Dataset

Gambar 4.14 Perbandingan Visualisasi *Embedding* Transaksi Menggunakan GAT dan GAT + Class Weight

Berdasarkan Gambar 4.14, terlihat bahwa model GAT dan GAT + *Class Weight* mampu menghasilkan representasi *embedding* yang cukup baik dalam membedakan transaksi legal dan transaksi ilegal. Pada model GAT, sebagian besar transaksi ilegal yang ditunjukkan dengan warna merah mulai membentuk beberapa kelompok (*cluster*), meskipun masih terdapat sejumlah titik yang tersebar di antara transaksi legal. Kondisi ini menunjukkan bahwa model telah mampu mengenali pola hubungan antar transaksi, namun pemisahan antarkelas belum sepenuhnya optimal. Sementara itu, pada model GAT + *Class Weight* terlihat bahwa transaksi ilegal membentuk kelompok yang lebih jelas dan penyebarannya lebih terpusat dibandingkan model GAT. Selain itu, jumlah transaksi ilegal yang berada di dalam kelompok transaksi legal juga tampak lebih sedikit. Hal ini menunjukkan bahwa penambahan *Class Weight* membantu model memberikan perhatian yang lebih besar terhadap kelas minoritas sehingga representasi fitur transaksi ilegal menjadi lebih mudah dibedakan dari transaksi legal. Secara keseluruhan, hasil visualisasi *embedding* menunjukkan bahwa model GAT + *Class Weight* menghasilkan pemisahan kelas yang lebih baik dibandingkan model GAT. Temuan ini sejalan dengan hasil evaluasi sebelumnya, di mana GAT + *Class Weight* memperoleh nilai *recall* tertinggi sebesar 90,43% dan jumlah *False Negative* terendah sebanyak 87 transaksi. Dengan demikian, visualisasi *embedding* ini memperkuat hasil evaluasi kuantitatif bahwa penambahan *Class Weight* mampu meningkatkan kemampuan model dalam mengenali pola transaksi *fraud* pada jaringan *Bitcoin*.

4.4 Pemetaan Hasil Penelitian ke dalam Kerangka Investigasi *Digital forensik*

Sebagaimana telah dijelaskan pada subbab 3.11, penelitian ini mengadopsi kerangka kerja forensik digital dari NIST yang terdiri atas empat tahap, yaitu *collection*, *examination*, *analysis*, dan *reporting*. Pada tahap collection, data yang digunakan berasal dari Elliptic Bitcoin Dataset yang terdiri atas 203.769 transaksi, termasuk 157.205 transaksi dengan label Unknown. Seluruh transaksi tetap dipertahankan pada tahap eksplorasi agar struktur jaringan yang dianalisis tetap menggambarkan kondisi dataset secara keseluruhan.

Pada tahap examination, dilakukan pemeriksaan struktur graph, distribusi time step, korelasi fitur, serta hubungan transaksi Unknown dengan transaksi Illicit. Hasil examination menunjukkan bahwa terdapat 4.027 transaksi Unknown yang memiliki hubungan langsung dengan transaksi Illicit. Pada tahap analysis, model GAT + Class Weight digunakan untuk melakukan inferensi terhadap transaksi Unknown. Hasil prediksi menunjukkan bahwa 119.100 transaksi Unknown diprediksi memiliki pola lebih mendekati Licit, sedangkan 38.105 transaksi diprediksi lebih mendekati Illicit. Hasil prediksi tersebut kemudian digabungkan dengan analisis hubungan graph. Pada tahap reporting, ditemukan sebanyak 1.942 transaksi Unknown yang memenuhi dua indikator, yaitu diprediksi sebagai Illicit oleh model dan memiliki hubungan langsung dengan transaksi Illicit. Transaksi tersebut dapat dimasukkan dalam daftar kandidat untuk pemeriksaan lanjutan oleh investigator. Namun, hasil ini tidak digunakan sebagai bukti bahwa transaksi tersebut pasti ilegal, melainkan sebagai informasi pendukung dalam proses penentuan prioritas investigasi.

UNIVERSITAS
NUSA MANDIRI