 ISSN:2302-8149
e-ISSN:2540-9719

Klasifikasi Transaksi Keuangan Mencurigakan dengan Metode *Light Gradient Boosting Machine* (LGBM) berdasarkan Indikator *Social Network Analysis* (SNA)

Classification of Suspicious Financial Transactions using Light Gradient Boosting Machine Method (LGBM) based on Social Network Analysis (SNA) Indicators

[1]Ayu Fara Paramitha*, [2] Yuti Dewita Arimbi, [3]Slamet Riyanto, [4]Niken Fitria Apriani, [5]Al Hafiz Akbar Maulana Siagian
[1]Perangkat Lunak dan Sistem Informasi, Manajemen Sistem Informasi, Universitas Gunadarma
[2]Teknik Informatika, Teknologi Industri, Universitas Gunadarma
[3,4,5]Badan Riset dan Inovasi Nasional Jakarta, Indonesia
*e-mail: ayufaraparamitha01@gmail.com

(***received*:** 01 August 2023, ***revised*:** 11 December 2023, ***accepted*:** 16 January 2024)

Abstrak

Pencucian uang merupakan perbuatan yang dilakukan oleh perorangan maupun suatu kelompok untuk menyembunyikan atau menyamarkan asal usul harta kekayaan yang didapat dari tindakan ilegal menjadi harta kekayaan yang seolah-olah didapat dari tindakan legal. Umumnya terdapat tiga proses pencucian uang, diantaranya *placement*, *layering*, dan *integration*. Kompleksitas proses yang terjadi pada pencucian uang yang telah dijelaskan diatas mengakibatkan sulitnya melakukan penelusuran terkait transaksi keuangan yang mencurigakan serta menemukan pihak-pihak yang terlibat dan transaksi mana saja yang terlibat ke dalam jaringan terduga tindak pencucian uang. Untuk menyelesaikan permasalahan ini, metode *Social Network Analysis* (SNA) diterapkan untuk mendapatkan data berupa fitur-fitur SNA. Pada tahap selanjutnya, fitur-fitur SNA tersebut digunakan sebagai indikator pengidentifikasian transaksi keuangan yang mencurigakan. Data indikator yang telah diperoleh digunakan untuk membuat model pengklasifikasian dengan menggunakan metode *Light Gradient-Boosting Machine* (LGBM). Hasil dari penelitian ini adalah model yang dibuat dengan menggunakan metode SNA dan LGBM menghasilkan akurasi sebesar 97%. Nilai *precision*, *recall*, dan F1-*Score* untuk data transaksi yang tidak mencurigakan masing-masing sebesar 98%, 97%, dan 97%, sedangkan untuk data transaksi yang mencurigakan masing-masing sebesar 97%, 98%, dan 97%. Hasil akurasi yang didapat cukup tinggi dan menunjukkan bahwa metode yang digunakan mampu menyelesaikan masalah klasifikasi transaksi keuangan yang mencurigakan dengan baik. Hasil studi ini diharapkan bisa menjadi alternatif dalam mendeteksi transaksi keuangan yang mencurigakan untuk mencegah praktik pencucian uang.

Kata kunci: Pencucian Uang, Transaksi Keuangan Mencurigakan, Social Network Analysis, Light Gradient-Boosting Machine.

Abstract

Money laundering is an act committed by individuals or a group to conceal or disguise the origin of wealth obtained from illegal activities into assets that appear to have been acquired through legal means. Generally, there are three money laundering processes: placement, layering, and integration. The complexity of these money laundering processes described above makes it difficult to trace suspicious financial transactions and identify the parties involved and which transactions are connected to the suspected money laundering network. To address this issue, Social Network Analysis (SNA) is implemented to generate SNA features. In the following stage, these SNA features are employed as indicators to detect suspicious financial activities. The gathered indicator data is utilized to build a classification model using the Light Gradient-Boosting Machine (LGBM) approach. The

 ISSN:2302-8149
e-ISSN:2540-9719

results of this study show that the model created using SNA and LGBM methods achieved an accuracy of 97%. The precision, recall, and F1-Score values for non-suspicious transaction data were 98%, 97%, and 97%, respectively, while for suspicious transaction data, they were 97%, 98%, and 97%, respectively. The achieved accuracy values were quite high indicating that the used approach was capable of effectively classifying suspicious financial activities. We believe that the findings of this study could be an alternative method for detecting suspicious financial transactions in order to avoid money laundering operations.

Keywords: *Money Laundering, Suspicious Financial Transaction, Social Network Analysis, Light Gradient-Boosting Machine*

1 Pendahuluan

Pencucian uang adalah perbuatan yang dilakukan oleh perorangan maupun kelompok dengan tujuan menyembunyikan sumber perolehan harta kekayaan secara ilegal menjadi legal [1]. Menurut undang-undang Nomor 15 tahun 2002 tentang tindak pidana pencucian uang pasal 1 ayat 1, pencucian uang merupakan proses atau perbuatan menempatkan, mentransfer, membayarkan, membelanjakan, menghibahkan, menyumbangkan, menitipkan, membawa keluar negeri, menukarkan, atau perbuatan lainnya atas Harta Kekayaan yang patut diduga atau diketahui berasal dari hasil tindak pidana yang bertujuan untuk menyembunyikan atau menyamarkan asal usul Harta Kekayaan tersebut sehingga seolah-olah merupakan Harta Kekayaan yang sah [2]. Transaksi ini sangat sulit untuk ditelusuri karena uang "kotor" hasil tindak kejahatan yang didapat akan bercampur dengan uang bersih yang telah diperoleh.

Proses pencucian uang yang difokuskan pada penelitian ini adalah pencucian uang melalui transaksi keuangan pada bank yang dilakukan oleh seseorang atau sekelompok orang. Seiring berkembangnya modus dan teknik pencucian uang, seringkali membuat penyedia jasa keuangan dan aparat penegak hukum tidak menyadari adanya proses tindak pidana pencucian uang melalui transaksi pada bank, sehingga transaksi yang dilakukan oleh seseorang atau sekelompok orang tidak lagi menimbulkan kecurigaan karena transaksi yang dilakukan menyerupai transaksi yang normal [3]. Umumnya terdapat tiga proses pencucian uang, diantaranya *placement, layering,* dan *integration. Placement* merupakan tahap pembagian sejumlah besar uang ke dalam jumlah yang lebih kecil atau jumlah yang tidak dicurigai oleh pihak bank kepada sejumlah rekening yang berbeda [4]. *Layering* merupakan tahap menyembunyikan uang yang telah dibagi dengan mentransfernya ke sejumlah rekening berbeda agar sumber dari uang tersebut tersembunyi atau tidak diketahui [4]. *Integration* merupakan tahap mengumpulkan uang yang telah dibagi menjadi sejumlah kecil kepada seseorang [4]. Berdasarkan proses pencucian uang yang telah dijabarkan diatas dapat diketahui bahwa terdapat pola transaksi keuangan mencurigakan yang dapat ditelusuri untuk mengetahui pihak-pihak yang terlibat pada sebuah kasus pencucian uang. Pihak-pihak yang terlibat dapat dibagi menjadi dua pihak, yaitu adalah pelaku utama dan komplotan yang membantu pelaku utama untuk menyembunyikan uang "kotor" yang dimiliki.

Kompleksitas proses yang terjadi pada pencucian uang yang telah dijelaskan diatas mengakibatkan sulitnya melakukan penelusuran terkait transaksi keuangan yang mencurigakan serta menemukan pihak-pihak yang terlibat dan transaksi mana saja yang terlibat ke dalam jaringan terduga tindak pencucian uang. Untuk menyelesaikan masalah ini maka diperlukan sebuah metode yang dapat menemukan pihak-pihak yang diduga melakukan tindak pidana pencucian uang. Metode yang dapat digunakan adalah metode graf yang dapat menelusuri keterhubungan antar rekening. Salah satu metode graf yang dapat digunakan adalah metode *Social Network Analysis* (SNA).

Penggunaan metode graf berupa metode SNA untuk menganalisis data transaksi keuangan dan mengidentifikasikan pelaku telah dilakukan sebelumnya oleh [4]. Penelitian tersebut juga membandingkan metode baru yang dibuat dengan tiga metode supervised learning seperti, *Support Vector Machine* (SVM), *Decision Tree* (DT), dan *Deep Learning*. Hasil dari penelitian ini menunjukkan bahwa metode yang diajukan memiliki nilai *precision, recall,* dan *F-measure* tertinggi dibandingkan dengan ketiga metode *supervised learning* yang telah disebutkan. Pada penelitian tersebut, model SNA yang telah dibuat dikombinasikan dengan metode baru sehingga menghasilkan nilai yang cukup tinggi, namun hal ini dapat dikembangkan kembali dengan mengkombinasikan model SNA dengan model *Machine Learning* untuk meningkatkan akurasi dari model yang dibuat.

 ISSN:2302-8149
e-ISSN:2540-9719

Salah satu metode *machine learning* yang dapat digunakan adalah metode *Light Gradient-Boosting Machine* (LGBM). [5] telah melakukan klasifikasi aktivitas penipuan pada data Ethereum dengan menggunakan metode LGBM. Metode yang digunakan [5] untuk mengklasifikasikan aktivitas penipuan diantaranya *Random Forest*, *Multi-Layer Perceptron*, *Logistic Regression*, KNN, XGBoost, SVC, dan ADAboost. Hasil yang diperoleh adalah LGBM dan *Random Forest* memiliki performa yang terbaik dengan akurasi yang tinggi, dimana LGBM memiliki akurasi sebesar 99.17% dan *Random Forest* sebesar 98.26%. LGBM sendiri merupakan model *ensemble boosting* yang dapat menggabungkan model pembelajaran yang lemah untuk membentuk suatu model pembelajar yang lebih kuat dan akurat [6]. LGBM juga dapat mempercepat proses pelatihan dan mengurangi penggunaan memori.

Penelitian ini menggunakan metode *Social Network Analysis* (SNA) untuk menemukan keterhubungan antar rekening serta pola-pola transaksi keuangan yang mencurigakan dan menemukan pelaku utama yang menjadi pusat dari pola-pola yang ditemukan. Metode ini juga dipakai untuk mendapatkan indikator-indikator yang dapat digunakan untuk melakukan klasifikasi. Penelitian ini juga menggunakan metode *Light Gradient-Boosting Machine* (LGBM) untuk mengklasifikasi data yang termasuk kedalam transaksi keuangan yang mencurigakan. Pada metode SNA terdapat beberapa indikator yang digunakan, seperti *degree centrality, betweenness centrality, closeness centrality, eccentricity*, dan *modularity*. Indikator ini akan digunakan untuk melakukan klasifikasi terhadap data transaksi keuangan yang dimiliki. Penelitian ini diharapkan dapat membantu berbagai pihak dalam melakukan pendeteksian transaksi keuangan yang mencurigakan pada data transaksi keuangan.

2 Tinjauan Literatur

Penelitian mengenai pendeteksian pencucian uang terus berkembang seiring dengan perkembangan teknologi yang ada. Penelitian terkait mengenai keterhubungan antar rekening yang terdapat pada data transaksi keuangan dengan menggunakan metode SNA telah dilakukan oleh beberapa peneliti terdahulu. Penelitian yang dilakukan oleh [7] menghasilkan sebuah desain model relasional untuk mengidentifikasi hubungan seperti hubungan keluarga dan bisnis serta transaksi antar pelanggan yang mencurigakan berdasarkan data profil pelanggan. Data yang digunakan merupakan data yang dibuat dengan menggunakan *software tool* dengan bahasa C#. Berbeda dengan penelitian [7], [4] melakukan analisis data transaksi keuangan dan membuat metode baru untuk mengidentifikasikan pelaku utama dan komplotan tindak pidana pencucian uang berdasarkan dataset transaksi keuangan berdasarkan hasil analisis SNA. Penelitian tersebut juga membandingkan metode yang diajukan dengan tiga metode *supervised learning* seperti, *Support Vector Machine* (SVM), *Decision Tree* (DT), dan *Deep Learning*. Hasil dari penelitian ini menunjukkan bahwa metode yang diajukan memiliki nilai *precision, recall*, dan *F-measure* tertinggi dibandingkan dengan ketiga metode *supervised learning* yang telah disebutkan. Penelitian yang dilakukan oleh [4] dapat dikembangkan kembali dengan menambahkan metode *Machine Learning* (ML) kedalam model SNA yang telah dibuat. Salah satu cara yang dapat dilakukan adalah dengan melakukan *node classification* untuk membuat model klasifikasi pencucian uang berdasarkan indikator yang ada pada SNA.

Penelitian mengenai *node classification* dengan menggunakan indikator yang ada pada metode graf telah dilakukan oleh beberapa peneliti terdahulu. [8] melakukan penelitian mengenai node classification dengan menggunakan *Graph Neural Networks* (GNNs). Penelitian ini menyoroti ketidakseimbangan pada sampel yang dapat mengakibatkan tidak optimalnya hasil klasifikasi, sehingga [8] membangun sebuah kerangka kerja bernama GraphSMOTE yang menggabungkan algoritma *oversampling* minoritas sintesis dengan informasi terkait graf untuk meningkatkan performa klasifikasi.

Algoritma *machine learning* yang akan diterapkan pada penelitian ini adalah metode *Light Gradient-Boosting Machine* (LGBM). Penelitian mengenai penggunaan metode LGBM untuk melakukan klasifikasi pada data keuangan telah dilakukan sebelumnya oleh para peneliti terdahulu. [5] telah melakukan klasifikasi aktivitas penipuan pada data Ethereum dengan menggunakan metode LGBM. Proses optimalisasi pada parameter LGBM dapat dimodifikasi dengan menggunakan pendekatan estimasi terstruktur jarak Euclidean. Performa yang digunakan untuk mengklasifikasikan aktivitas penipuan seperti *Random Forest*, *Multi-Layer Perceptron*, *Logistic Regression*, KNN, XGBoost, SVC, dan ADAboost. Hasil yang diperoleh adalah LGBM dan *Random Forest* memiliki

 ISSN:2302-8149
e-ISSN:2540-9719

performa yang terbaik dengan akurasi yang tinggi, dimana LGBM memiliki akurasi sebesar 99.17% dan *Random Forest* sebesar 98.26%. Penelitian yang dilakukan oleh [9] mengenai pendeteksian penipuan dalam transaksi kartu kredit menggunakan LGBM yang telah dioptimalkan dengan menggunakan algoritma hiperparameter berbasis Bayesian diintegrasikan untuk menyetel parameter dari *Light Gradient-Boosting Machine* (LightGBM). Penelitian ini menggunakan dua dataset dan membandingkan hasil deteksi dengan beberapa pendekatan lain. Hasil pendekatan yang diusulkan mengungguli pendekatan lain dan mencapai kinerja tertinggi dalam hal akurasi (98,40%), *Area under receiver operating characteristic curve* (AUC) (92,88%), Presisi (97,34%), dan F1-*score* (56,95%).

3 Metode Penelitian

Tahapan yang dilakukan pada penelitian ini dimulai dengan mengolah data transaksi keuangan yang dimiliki dengan menggunakan metode SNA. Data yang diperoleh dari Kaggle akan diolah dengan menggunakan metode SNA untuk mendapatkan data berupa fitur-fitur SNA yang selanjutnya akan digunakan sebagai indikator untuk mengidentifikasikan transaksi keuangan yang mencurigakan. Visualisasi SNA dilakukan dengan menggunakan *library* python yang bernama networkx, namun untuk melakukan perhitungan indikator SNA dilakukan dengan menggunakan aplikasi Gephi.

Gambar 1 Gambaran Umum Penelitian

Tahapan - tahapan yang dilakukan pada penelitian ini terlihat pada Gambar 1. Berikut merupakan penjabaran tahapan penelitian yang dilakukan:

1. Input Data Transaksi Keuangan
   Penelusuran pola-pola pencucian uang dapat dilakukan dengan menganalisis transaksi keuangan yang dimiliki oleh bank. Data yang dimiliki oleh bank bersifat rahasia sehingga penelitian ini menggunakan data sintesis yang dapat diakses melalui *website* Kaggle (https://www.kaggle.com/datasets/maryam1212/money-laundering-data). Dataset yang disediakan berupa data transaksi keuangan dengan tipe *cash-in* dan *transfer* dengan jumlah data secara keseluruhan sebesar 2340 data dengan detail data yang tertuang pada Tabel 1. Data transaksi keuangan yang digunakan dapat dilihat pada Gambar 2.

	typeofaction	sourceid	destinationid	amountofmoney	date	isfraud	typeoffraud
0	cash-in	30105	28942	494528	2019-07-19 14:40:00	1	type1
1	cash-in	30105	8692	494528	2019-05-17 14:57:00	1	type1
2	cash-in	30105	60094	494528	2019-07-20 13:20:00	1	type1
3	cash-in	30105	20575	494528	2019-07-03 14:15:00	1	type1
4	cash-in	30105	45938	494528	2019-05-26 10:40:00	1	type1

Gambar 2 Data Transaksi Keuangan

Tabel 1 Distribusi Dataset Yang Digunakan

Transaksi Keuangan	*isfraud*	Jumlah Data
Mencurigakan	1	1399
Tidak Mencurigakan	0	941
Total		**2340**

2. Seleksi Fitur
   Data yang telah didapat kemudian diseleksi sesuai dengan kebutuhan penelitian ini. Data utama yang dibutuhkan antara lain *sourceid*, *destinationid*, dan *amountofmoney*. Penamaan

 ISSN:2302-8149
e-ISSN:2540-9719

data tersebut diubah untuk menyederhanakan penyebutan sehingga masing-masing namanya menjadi *source*, *target*, dan *weight*.

3. Implementasi *Social Network Analysis* (SNA)
Social Network Analysis merupakan sebuah proses dalam mendefinisikan komunitas atau kumpulan sosial berdasarkan jaringan dan teori graf. Metode ini digunakan karena metode ini dapat menganalisis hubungan antar node untuk menemukan struktur dan keterhubungan antar individu atau bahkan organisasi [10]. Pada metode SNA terdapat beberapa istilah yang akan sering digunakan pada penelitian ini, diantaranya adalah *node* dan *edge*. *Node* merupakan properti pada metode SNA yang merepresentasikan posisi aktor dalam jaringan, sedangkan *edge* merupakan properti yang merepresentasikan interaksi antara dua aktor atau lebih [11]. Pada metode SNA terdapat beberapa Penelitian ini menggunakan lima indikator SNA. Indikator yang dipilih didasari oleh penelitian yang dilakukan oleh [4]. Indikator yang digunakan diantaranya *degree centrality*, *closeness centrality*, *eccentricity*, *modularity*, dan *betweenness centrality*. Karena metode graf yang digunakan adalah *Directed Graph* maka untuk indikator *degree centrality* akan menggunakan *in-degree centrality* dan *out-degree centrality*. Berikut merupakan fungsi dari masing-masing indikator yang digunakan [12].
   a. *In-degree Centrality*
   In-degree berfungsi untuk menunjukkan berapa banyak sebuah *node* dihubungi oleh *node* lain. Rumus umum dari *In-degree Centrality* dapat dilihat seperti dibawah ini.
   $$C_{indeg}(v) = \frac{d\ {}_{v}^{in}}{|N|-1} \quad (1)$$
   Dimana,
   v = *Node*
   N = Banyaknya *node* pada jaringan
   $d\ {}_{v}^{in}$ = *In-degree* dari *node* v

   b. *Out-degree Centrality*
   Out-degree Centrality berfungsi untuk menunjukkan berapa banyak sebuah *node* menghubungi *node* lainnya. Rumus umum dari *Out-degree Centrality* dapat dilihat seperti dibawah ini.
   $$C_{outdeg}(v) = \frac{d\ {}_{v}^{out}}{|N|-1} \quad (2)$$
   Dimana,
   v = *Node*
   N = Banyaknya *node* pada jaringan
   $d\ {}_{v}^{out}$ = *Out-degree* dari *node* v

   c. *Closeness Centrality*
   Closeness Centrality berfungsi untuk mengukur seberapa dekat sebuah *node* dengan *node* lainnya. Rumus umum *Closeness Centrality* terdapat pada persamaan 3 dibawah ini.
   $$C_{close}(v) = \frac{|N|-1}{\sum_{u \in N \setminus \{v\}} d(v,u)} \quad (3)$$
   Dimana,
   v = *Node*
   N = Banyaknya *node* pada jaringan
   d(v, u) = Panjang dari jalur terpendek dari v ke u

 ISSN:2302-8149
e-ISSN:2540-9719

d. *Betweenness Centrality*
Betweenness Centrality berfungsi untuk menunjukkan seberapa besar sebuah *node* menjadi penghubung *node* lainnya. Berikut merupakan rumus umum dari *Betweenness Centrality*.

$$C_{btw}(v) = \sum_{s,t \in N} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}} \quad (4)$$

Dimana,
v = *Node*
$\sigma_{s,t}(v)$ = Angka jalur terpendek antara *node* s dan t yang melewati *node* v
$\sigma_{s,t}$ = Angka jalur terpendek antara *node* s dan t

e. *Eccentricity*
Eccentricity berfungsi untuk menunjukkan jarak terbesar dari semua jarak antar *node*.
f. *Modularity*
Berfungsi untuk memetakan kelompok atau *cluster* yang ada pada sebuah jaringan. Sebuah jaringan yang awalnya termasuk kedalam satu kelompok besar akan membelah atau terpecah hingga terbentuk sebuah klaster atau kelompok kecil pada jaringan. Pemecahan ini dilakukan berdasarkan nilai *betweenness* yang dimiliki sebuah simpul. Simpul dengan nilai *betweenness* yang tinggi menandakan bahwa simpul tersebut menjadi penghubung diantara dua kelompok.

Indikator-indikator ini akan digunakan sebagai fitur untuk melakukan klasifikasi dengan menggunakan metode LGBM. Indikator diatas akan dihitung menggunakan aplikasi Gephi (https://gephi.org/users/download/). Gephi yang digunakan adalah Gephi versi 0.10.1 untuk pengguna windows. Aplikasi Gephi sendiri merupakan aplikasi *open source* yang dapat memvisualisasi dan menganalisis jaringan untuk menemukan tren dan latar belakang dari sebuah data [13].

4. Implementasi *Light Gradient-Boosting Machine* (LGBM)
LGBM adalah suatu model data yang menggunakan *Gradient Boosting Decision Trees* (GBDT) yang dikembangkan oleh Microsoft pada tahun 2017 [14]. LGBM adalah suatu kerangka kerja pembelajaran mesin yang menggunakan gradien dan memanfaatkan pohon keputusan serta teknik *boosting* [15]. LGBM merupakan model *ensemble boosting* yang dapat menggabungkan model pembelajaran yang lemah untuk membentuk suatu model pembelajar yang lebih kuat dan akurat [6]. Perbedaan LGBM dan model XGBoost terletak pada penggunaan algoritma berbasis histogram yang bertujuan untuk mempercepat proses pelatihan, mengurangi penggunaan memori, dan mengadopsi strategi pertumbuhan *leaf-wise* dengan batasan kedalaman pohon [15]. Algoritma ini dipilih karena meskipun LGBM tidak mengurangi akurasi prediksi, algoritma ini secara signifikan meningkatkan kecepatan peramalan dan mengurangi penggunaan memori [14].

5. Evaluasi
Tahap ini dilakukan untuk menguji model yang telah dibuat dengan menggunakan LGBM. Pengujian dilakukan dengan mengukur keakuratan metode yang digunakan dalam mengklasifikasikan transaksi keuangan yang mencurigakan atau tidak. Parameter pengujian yang digunakan adalah *accuracy*, *precision*, *recall*, dan *f1-score*. Pengujian ini dilakukan pada data yang tidak termasuk kedalam transaksi keuangan yang tidak mencurigakan yang kemudian akan disebut sebagai class 0. Pengujian ini juga dilakukan pada data yang termasuk kedalam transaksi keuangan yang mencurigakan yang kemudian akan disebut sebagai class 1. Jumlah rekening yang termasuk kedalam class 0 adalah sebanyak 290 data rekening, sedangkan jumlah rekening yang termasuk kedalam class 1 adalah sebanyak 274 data rekening.

 ISSN:2302-8149
e-ISSN:2540-9719

4 Hasil dan Pembahasan

Penelitian ini dilakukan untuk melakukan klasifikasi transaksi yang mencurigakan atau tidak pada data transaksi keuangan. Proses penelitian yang akan dilakukan dapat dilihat pada Gambar 1 diatas. Metode yang diterapkan pada penelitian ini adalah metode SNA dan metode LGBM.

1. Dataset

Dataset yang digunakan pada penelitian ini merupakan data transaksi keuangan yang terdiri dari *typeofaction*, *sourceid*, *destinationid*, *amountofmoney*, *date*, *isfraud*, dan *typeoffraud*. Dataset tersebut mula-mula akan diseleksi terlebih dahulu untuk mendapatkan data-data yang dibutuhkan yaitu berupa data *sourceid*, *destinationid*, dan *amountofmoney*. Data tersebut disederhanakan penamaannya dengan mengubah masing-masing data diatas menjadi *source*, *target*, dan *weight*. Gambar 3 menunjukkan bentuk dataset setelah dilakukan seleksi dan penyederhanaan.

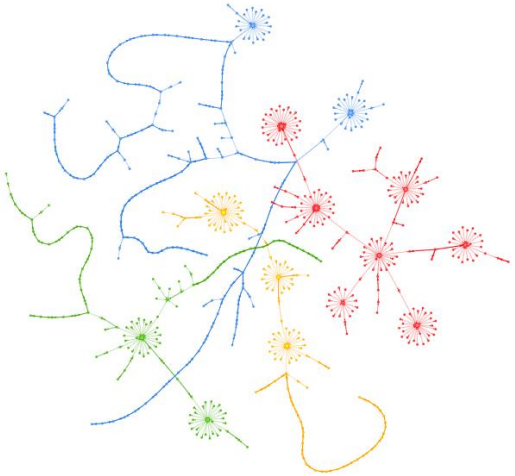
	source	target	weight
0	30105	28942	494528
1	30105	8692	494528
2	30105	60094	494528
3	30105	20575	494528
4	30105	45938	494528

Gambar 3 Data Transaksi Keuangan Setelah Transformasi

Pada Gambar 3, *data source* merupakan data yang berisikan nomor-nomor rekening dari pihak pengirim, sedangkan target merupakan data yang berisikan nomor-nomor rekening dari pihak penerima. *Data weight* pada gambar diatas merupakan jumlah uang yang terlibat pada transaksi yang terjadi. Dataset ini akan digunakan untuk memperoleh data yang berisikan indikator-indikator yang digunakan untuk melakukan klasifikasi dengan menggunakan metode SNA.

2. *Social Network Analysis* (SNA)

Metode SNA digunakan untuk mendapatkan indikator yang digunakan untuk melakukan klasifikasi. Metode ini juga dapat digunakan untuk menemukan pola-pola transaksi keuangan yang mencurigakan. Pola-pola tersebut didapat dari keterhubungan antar rekening yang ada pada data transaksi keuangan. Pada dasarnya SNA dapat digunakan untuk membuat visualisasi dari data yang ada. Bentuk visualisasi SNA mengenai pola transaksi keuangan yang mencurigakan dapat terlihat seperti di Gambar 4.



Gambar 4 Pola Transaksi Keuangan yang Mencurigakan

 ISSN:2302-8149
e-ISSN:2540-9719

Pola yang terdapat pada Gambar 4 memiliki warna yang berbeda. Warna-warna ini menandakan bahwa setiap warna yang ada mewakili komunitas atau komplotan dari transaksi keuangan yang mencurigakan. Jika dilihat dari tahapan pencucian uang yang telah disebutkan, maka titik tengah dengan warna yang tebal dapat diketahui sebagai terduga pelaku utama dari tindak pencucian uang karena titik atau *node* tersebut merupakan pusat dari dilakukannya transaksi keuangan yang mencurigakan. Pelaku utama tersebut berperan pada tahapan placement dan integration. Sedangkan alur yang memanjang seperti tali pada gambar diatas merupakan terduga komplotan dan termasuk kedalam tahapan *layering*.

Perhitungan indikator SNA pada data transaksi keuangan ini dilakukan dengan menggunakan aplikasi Gephi. Hasil perhitungan indikator SNA terhadap data yang digunakan dapat dilihat pada Gambar 5.

Id	indegree	outdegree	Degree	weighted indegree	weighted outdegree	Weighted Degree	Eccentricity	closnesscentrality	modularity_class	betweenesscentrality
30105	1	20	21	6306849	9890569	16197418	1	1.0	133	0.000003
28942	1	0	1	494528	0	494528	0	0.0	133	0.000000
8692	1	0	1	494528	0	494528	0	0.0	133	0.000000
60094	2	0	2	1719336	0	1719336	0	0.0	133	0.000000
20575	1	0	1	494528	0	494528	0	0.0	133	0.000000

Gambar 5 Data Indikator SNA

3. Light Gradient-Boosting Machine

Pada Tahap ini dilakukan pengklasifikasian dengan menggunakan metode LGBM. Metode ini akan menggunakan indikator SNA yang telah didapat dan membuat model pengklasifikasian dengan indikator tersebut. Langkah pertama yang dilakukan adalah menambahkan *class* yang ada pada dataset utama ke dalam data indikator SNA sehingga menghasilkan dataset baru seperti pada Gambar 6.

	Id	indegree	outdegree	Degree	weighted indegree	weighted outdegree	Weighted Degree	Eccentricity	closnesscentrality	modularity_class	betweenesscentrality	isfraud
0	30105	1	20	21	6306849	9890569	16197418	1	1.0	133	0.000003	1
1	28942	1	0	1	494528	0	494528	0	0.0	133	0.000000	1
2	8692	1	0	1	494528	0	494528	0	0.0	133	0.000000	1
3	60094	2	0	2	1719336	0	1719336	0	0.0	133	0.000000	1
4	20575	1	0	1	494528	0	494528	0	0.0	133	0.000000	1
...	...	...	...	...	...	...	...	...	...	...	...	...
2813	14945	0	1	1	0	106907	106907	1	1.0	131	0.000000	0
2814	9532	0	1	1	0	106907	106907	1	1.0	131	0.000000	0
2815	27332	0	1	1	0	106907	106907	1	1.0	131	0.000000	0
2816	32685	0	1	1	0	106907	106907	1	1.0	131	0.000000	0
2817	26390	0	1	1	0	106907	106907	1	1.0	131	0.000000	0

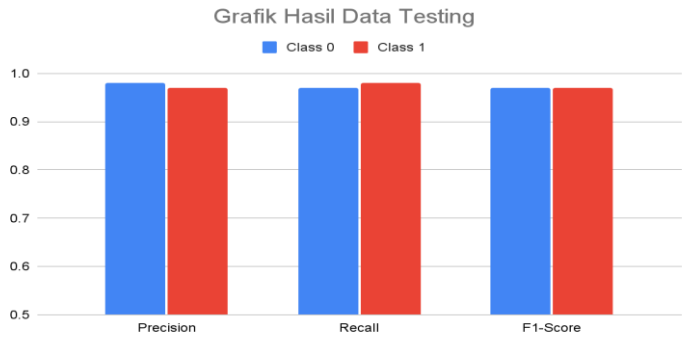
Gambar 6 Dataset Baru

Setelah dibentuk dataset baru, langkah selanjutnya adalah membagi dataset tersebut menjadi dataset *training* dan dataset *testing*. Jumlah persentase dari dataset *training* adalah sebesar 80% dan dataset *testing* adalah sebesar 20%. Hasil akurasi dari penelitian ini dapat dilihat seperti pada Gambar 7 dan Gambar 8.

	precision	recall	f1-score	support
class 0	0.98	0.97	0.97	290
class 1	0.97	0.98	0.97	274
accuracy			0.97	564
macro avg	0.97	0.97	0.97	564
weighted avg	0.97	0.97	0.97	564

Gambar 7 Hasil Akurasi, *Precision, Recall* dan *F1-Score* Data Testing

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 8 Visualisasi Grafis Hasil Data Testing

Berdasarkan Gambar 7, terlihat bahwa model yang dilatih untuk data pengujian transaksi keuangan menghasilkan akurasi sebesar 97%. Hasil ini menunjukkan bahwa metode yang digunakan mampu menyelesaikan masalah klasifikasi transaksi keuangan yang mencurigakan dengan baik. Tingkat akurasi yang tinggi ini dapat disebabkan oeh beberapa faktor seperti kualitas data yang baik, pengolahan data yang tepat, dan pemilihan parameter yang optimal. Hasil evaluasi juga menunjukkan bahwa nilai *precision, recall,* dan F1-*Score* masing-masing menghasilkan nilai 98%, 97%, dan 97% untuk data yang tidak mencurigakan, serta 97%, 98%, dan 97% untuk data yang mencurigakan. Tingginya nilai *precision, recall,* dan F1-*Score* dalam penelitian ini menunjukkan kemampuan model dalam menyelesaikan masalah klasifikasi. Secara keseluruhan, hasil ini mengindikasikan kinerja model yang baik dalam menghadapi tantangan klasifikasi transaksi keuangan yang mencurigakan. Nilai *precision* sebesar 98% menekankan kemampuan model dalam mengidentifikasi data positif dengan sangat akurat. Selain akurasi dan precision yang tinggi, hasil evaluasi juga menunjukkan nilai *recall* dan F1-*Score* yang baik dalam penelitian ini. *Recall* sebesar 97% menunjukkan kemampuan model dalam mengidentifikasi sebagian besar data positif yang sebenarnya. F1-*Score* sebesar 97% mencerminkan keseimbangan antara *precision* dan *recall*, mengindikasikan bahwa model ini dapat memberikan kinerja yang baik dalam memprediksi dan mengklasifikasikan data dengan benar. Hasil evaluasi ini secara keseluruhan menggambarkan performa model yang sangat baik dalam menangani tantangan klasifikasi transaksi keuangan yang mencurigakan, dan menegaskan bahwa model tersebut cocok untuk digunakan dalam skenario pengawasan dan deteksi aktivitas keuangan yang mencurigakan dengan tingkat keberhasilan yang tinggi.

5 Kesimpulan

Penelitian ini berfokus pada tindak pidana pencucian uang melalui transaksi keuangan. Pada penelitian ini dilakukan klasifikasi untuk menentukan transaksi keuangan yang mencurigakan atau tidak. Metode yang digunakan pada penelitian ini untuk mengklasifikasikan transaksi keuangan yang mencurigakan pada data transaksi keuangan adalah metode LGBM. Metode ini menggunakan indikator-indikator yang dihasilkan melalui pengolahan data dengan menggunakan metode SNA. Hasil dari penelitian ini adalah model yang dibuat dengan menggunakan metode SNA dan LGBM menghasilkan akurasi sebesar 97%. Nilai *precision*, *recall*, dan F1-*Score* untuk data transaksi yang tidak mencurigakan masing-masing sebesar 98%, 97%, dan 97%, sedangkan untuk data transaksi yang mencurigakan masing-masing sebesar 97%, 98%, dan 97%. Meskipun hasil dari penelitian ini masih jauh dari sempurna, model yang telah dibuat dapat diterapkan ke dalam studi kasus data asli dengan menyesuaikan data dan model yang telah ada. Pada kasus nyata ketidakseimbangan data sangat mungkin terjadi sehingga akan menghasilkan akurasi yang berbeda. Untuk kedepannya, penelitian ini dapat dikembangkan dengan menemukan kombinasi indikator-indikator SNA lain yang dapat digunakan sebagai dasar klasifikasi. Penelitian ini juga dapat dikembangkan dengan mencoba metode graf lain selain metode SNA, misalnya *Graph Convolutional Networks* [16], dan menggunakan metode *machine learning* lainnya, misalnya *Deep Neural Network* (DNN) dan *random forest* [17], untuk menemukan metode klasifikasi yang lebih baik.

ISSN:2302-8149
e-ISSN:2540-9719

Referensi

[1] N. Nugroho, S. Sunarmi, M. Siregar, and R. Munthe, "Analisis terhadap Pencegahan Tindak Pidana Pencucian Uang oleh Bank Negara Indonesia," *ARBITER: Jurnal Ilmiah Magister Hukum*, vol. 2, no. 1, pp. 100–110, May 2020, doi: 10.31289/arbiter.v2i1.126.

[2] Indonesia, Pemerintah Pusat, "Undang-undang (UU) Nomor 15 Tahun 2002 tentang Tindak Pidana Pencucian Uang." Apr. 17, 2002. Accessed: Jul. 18, 2023. [Online]. Available: http://peraturan.bpk.go.id/Details/44452/uu-no-15-tahun-2002

[3] A. R. Handoko, "Perancangan Sistem Pakar Analisa Transaksi Keuangan Mencurigakan Menggunakan Metode Forward Chaining," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 10, no. 2, pp. 701–712, Nov. 2019, doi: 10.24176/simet.v10i2.3523.

[4] M. Mahootiha, A. H. Golpayegani, and B. Sadeghian, "Designing a New Method for Detecting Money Laundering based on Social Network Analysis," in *2021 26th International Computer Conference, Computer Society of Iran (CSICC)*, Mar. 2021, pp. 1–7. doi: 10.1109/CSICC52343.2021.9420621.

[5] R. M. Aziz, M. F. Baluch, S. Patel, and P. Kumar, "A Machine Learning based Approach to Detect the Ethereum Fraud Transactions with Limited Attributes," *Karbala International Journal of Modern Science*, vol. 8, no. 2, pp. 139–151, May 2022, doi: 10.33640/2405-609X.3229.

[6] M. Massaoudi, S. S. Refaat, I. Chihi, M. Trabelsi, F. S. Oueslati, and H. Abu-Rub, "A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting," *Energy*, vol. 214, p. 118874, Jan. 2021, doi: 10.1016/j.energy.2020.118874.

[7] A. K. Shaikh, M. Al-Shamli, and A. Nazir, "Designing a Relational Model to Identify Relationships between Suspicious Customers in Anti-Money Laundering (AML) Using Social Network Analysis (SNA)," *Journal of Big Data*, vol. 8, no. 1, p. 20, Jan. 2021, doi: 10.1186/s40537-021-00411-3.

[8] T. Zhao, X. Zhang, and S. Wang, "GraphSMOTE: Imbalanced Node Classification on Graphs with Graph Neural Networks," in *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, in WSDM '21. New York, NY, USA: Association for Computing Machinery, Mar. 2021, pp. 833–841. doi: 10.1145/3437963.3441720.

[9] A. A. Taha and S. J. Malebary, "An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine," *IEEE Access*, vol. 8, pp. 25579–25587, Jan. 2020, doi: 10.1109/ACCESS.2020.2971354.

[10] N. R. Al-Molhem, Y. Rahal, and M. Dakkak, "Social network analysis in Telecom data," *Journal of Big Data*, vol. 6, no. 1, p. 99, Nov. 2019, doi: 10.1186/s40537-019-0264-6.

[11] M. K. Anam, T. P. Lestari, M. B. Firdaus, and S. Fadli, "Analisis Kesiapan Masyarakat Pada Penerapan Smart City di Sosial Media Menggunakan SNA | Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, pp. 69–81, Feb. 2021, doi: 10.29207/resti.v5i1.2742.

[12] Eriyanto, *Analisis Jaringan Media Sosial: Dasar-dasar dan Aplikasi Metode Jaringan Sosial untuk Membedah Percakapan di Media Sosial*. Jakarta: Kencana (Prenadamedia Group), 2021.

[13] J. M. P. Sanchez, B. A. Alejandro, M. M. J. Olvido, and I. M. V. Alejandro, "An Analysis of Online Classes Tweets Using Gephi: Inputs for Online Learning," *International Journal of Information and Education Technology*, vol. 11, no. 12, pp. 583–589, 2021, doi: 10.18178/ijiet.2021.11.12.1568.

[14] Y. Ju, G. Sun, Q. Chen, M. Zhang, H. Zhu, and M. U. Rehman, "A Model Combining Convolutional Neural Network and LightGBM Algorithm for Ultra-Short-Term Wind Power Forecasting," *IEEE Access*, vol. 7, pp. 28309–28318, 2019, doi: 10.1109/ACCESS.2019.2901920.

[15] J. Fan, X. Ma, L. Wu, F. Zhang, X. Yu, and W. Zeng, "Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data," *Agricultural Water Management*, vol. 225, p. 105758, Nov. 2019, doi: 10.1016/j.agwat.2019.105758.

[16] I. Alarab, S. Prakoonwit, and M. I. Nacer, "Competence of Graph Convolutional Networks for

ISSN:2302-8149
e-ISSN:2540-9719

Anti-Money Laundering in Bitcoin Blockchain," in *Proceedings of the 2020 5th International Conference on Machine Learning Technologies*, in ICMLT '20. New York, NY, USA: Association for Computing Machinery, Jul. 2020, pp. 23–27. doi: 10.1145/3409073.3409080.

[17] J. Alotibi, B. Almutanni, T. Alsubait, H. Alhakami, and A. Baz, "Money Laundering Detection using Machine Learning and Deep Learning," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 13, no. 10, Art. no. 10, 46/31 2022, doi: 10.14569/IJACSA.2022.0131087.