

IMPROVING FRAUD DETECTION IN FINANCIAL TRANSACTIONS USING SMOTE AND MACHINE LEARNING TECHNIQUES

Zahratul Umamah, Joko Triloka, Sutedi, Sriyanto

Fakultas Ilmu Komputer, Institut Informatika dan Bisnis Darmajaya, Lampung Indonesia
Jalan Z.A. Pagar Alam, No.93 Gedong Meneng, Bandar Lampung, Lampung, Indonesia 35145
zahratul.2421211002@mail.darmajaya.ac.id

ABSTRAK

Perkembangan teknologi informasi mendorong meningkatnya penggunaan sistem transaksi keuangan digital yang memberikan kemudahan dan efisiensi dalam aktivitas ekonomi. Namun, peningkatan tersebut juga diikuti oleh meningkatnya risiko *financial fraud* yang dapat menimbulkan kerugian finansial serta menurunkan kepercayaan terhadap sistem keuangan digital. Permasalahan utama dalam deteksi *fraud* adalah ketidakseimbangan distribusi data antara transaksi normal dan transaksi *fraud* yang menyebabkan model klasifikasi kurang optimal dalam mengenali kelas minoritas. Penelitian ini bertujuan untuk meningkatkan kinerja sistem deteksi *financial fraud* melalui penerapan teknik penyeimbangan data menggunakan *Synthetic Minority Oversampling Technique* (SMOTE) yang dikombinasikan dengan algoritma *machine learning*. Metode penelitian meliputi analisis karakteristik dataset, *preprocessing* data, penerapan SMOTE, pembangunan model menggunakan *Logistic Regression*, *Support Vector Machine*, dan *Random Forest*, serta evaluasi performa menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa penerapan SMOTE mampu meningkatkan kemampuan model dalam mendeteksi transaksi *fraud*, dengan algoritma *Random Forest* memberikan performa terbaik dibandingkan algoritma lainnya.

Kata kunci: *financial fraud*, *machine learning*, SMOTE, *class imbalance*, *Random Forest*, deteksi *fraud*

1. PENDAHULUAN

Perkembangan teknologi informasi telah membawa perubahan signifikan dalam berbagai sektor kehidupan, khususnya pada sistem transaksi keuangan digital. Berbagai layanan keuangan berbasis teknologi seperti pembayaran elektronik, *mobile banking*, dan transaksi daring kini menjadi bagian integral dalam aktivitas ekonomi masyarakat modern. Transformasi digital tersebut memungkinkan proses transaksi dilakukan secara lebih cepat, efisien, serta menjangkau berbagai wilayah tanpa batasan geografis. Selain meningkatkan kemudahan dan kenyamanan bagi pengguna, digitalisasi pada sektor keuangan juga menghasilkan volume data transaksi yang sangat besar dan kompleks. Namun, di balik berbagai manfaat tersebut, peningkatan intensitas transaksi digital juga menghadirkan tantangan baru terkait keamanan sistem keuangan, terutama meningkatnya potensi terjadinya aktivitas penipuan atau *financial fraud* [1], [2].

Fenomena *financial fraud* dalam sistem transaksi digital menjadi permasalahan yang semakin krusial dalam beberapa tahun terakhir. Aktivitas penipuan tidak hanya menimbulkan kerugian ekonomi yang signifikan bagi institusi keuangan dan pengguna layanan, tetapi juga berpotensi menurunkan tingkat kepercayaan masyarakat terhadap sistem keuangan digital secara keseluruhan. Oleh karena itu, diperlukan suatu mekanisme deteksi *fraud* yang mampu bekerja secara efektif, cepat, dan adaptif terhadap pola transaksi yang terus berkembang. Dalam konteks ini, pendekatan berbasis kecerdasan buatan (*Artificial Intelligence*) dan *machine learning* mulai banyak dimanfaatkan sebagai solusi untuk meningkatkan

kemampuan sistem dalam mendeteksi aktivitas transaksi yang mencurigakan [3], [4].

Metode *machine learning* memiliki kemampuan untuk menganalisis pola transaksi dalam jumlah data yang besar serta mengidentifikasi anomali yang berpotensi menunjukkan aktivitas penipuan. Melalui proses pembelajaran dari data historis, algoritma *machine learning* dapat mengklasifikasikan transaksi keuangan ke dalam kategori normal maupun *fraud* secara lebih sistematis dibandingkan pendekatan konvensional berbasis aturan (*rule-based systems*). Berbagai algoritma klasifikasi seperti *Logistic Regression*, *Support Vector Machine*, dan *Random Forest* telah banyak digunakan dalam penelitian sebelumnya untuk membangun model deteksi *fraud* yang lebih adaptif dan akurat. Kemampuan model dalam mempelajari pola kompleks dari data transaksi menjadikan *machine learning* sebagai pendekatan yang menjanjikan dalam meningkatkan efektivitas sistem keamanan pada transaksi keuangan digital [5], [6], [7].

Meskipun demikian, penerapan *machine learning* dalam deteksi *fraud* masih menghadapi tantangan utama yang berkaitan dengan ketidakseimbangan distribusi data (*class imbalance*). Pada dataset transaksi keuangan, jumlah transaksi normal umumnya jauh lebih besar dibandingkan jumlah transaksi yang teridentifikasi sebagai *fraud*. Ketidakseimbangan ini menyebabkan model cenderung lebih banyak mempelajari karakteristik dari kelas mayoritas, sehingga kemampuan dalam mendeteksi transaksi *fraud* sebagai kelas minoritas menjadi kurang optimal. Akibatnya, model dapat menghasilkan tingkat akurasi yang tinggi secara

keseluruhan, namun memiliki kinerja yang rendah dalam mengidentifikasi kasus *fraud* yang sebenarnya. Kondisi ini menunjukkan adanya kesenjangan (*gap*) antara performa model secara umum dan kemampuannya dalam mendeteksi transaksi penipuan secara spesifik [7], [8].

Untuk mengatasi permasalahan tersebut, diperlukan pendekatan yang mampu meningkatkan representasi data pada kelas minoritas agar model klasifikasi dapat mempelajari pola transaksi *fraud* secara lebih optimal. Salah satu teknik yang banyak digunakan adalah *Synthetic Minority Oversampling Technique* (SMOTE), yang bekerja dengan menghasilkan data sintetis pada kelas minoritas berdasarkan karakteristik data yang telah ada. Dengan distribusi dataset yang lebih seimbang, model *machine learning* diharapkan mampu mengidentifikasi pola transaksi *fraud* secara lebih akurat serta meningkatkan performa klasifikasi secara keseluruhan [7], [9].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk meningkatkan kinerja deteksi *financial fraud* pada transaksi keuangan digital melalui penerapan teknik SMOTE yang dikombinasikan dengan beberapa algoritma *machine learning*. Penelitian ini juga melakukan evaluasi komparatif terhadap algoritma *Logistic Regression*, *Support Vector Machine*, dan *Random Forest* untuk menentukan model yang memiliki performa terbaik dalam mendeteksi transaksi *fraud*. Dengan pendekatan tersebut, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem deteksi *fraud* yang lebih efektif, adaptif, dan mampu meningkatkan keamanan pada ekosistem transaksi keuangan digital.

2. TINJAUAN PUSTAKA

2.1. Financial Fraud dalam Sistem Transaksi Digital

Perkembangan teknologi informasi dan digitalisasi sistem keuangan telah memberikan kemudahan dalam melakukan transaksi secara cepat, fleksibel, dan tanpa batas geografis. Namun demikian, peningkatan penggunaan sistem transaksi digital juga diikuti dengan meningkatnya risiko kejahatan finansial, khususnya *financial fraud*. *Fraud* dalam sistem keuangan digital dapat berupa berbagai aktivitas penipuan seperti penyalahgunaan kartu kredit, transaksi tidak sah, manipulasi data transaksi, hingga pencurian identitas digital.

Dalam konteks sistem keuangan modern, *fraud* menjadi permasalahan serius karena dapat menimbulkan kerugian ekonomi yang besar serta menurunkan tingkat kepercayaan masyarakat terhadap layanan keuangan digital. Penelitian menunjukkan bahwa meningkatnya transaksi berbasis internet dan pembayaran digital telah memperluas peluang bagi pelaku kejahatan untuk melakukan aktivitas penipuan yang semakin kompleks dan sulit dideteksi oleh sistem keamanan tradisional [10].

Sistem deteksi *fraud* yang berbasis aturan (*rule-based systems*) yang sebelumnya banyak digunakan

oleh institusi keuangan memiliki keterbatasan dalam menghadapi pola penipuan yang terus berkembang. Sistem tersebut umumnya hanya mampu mendeteksi pola *fraud* yang telah diketahui sebelumnya dan sering menghasilkan tingkat *false positive* yang tinggi. Oleh karena itu, diperlukan pendekatan yang lebih adaptif dan mampu mempelajari pola transaksi secara dinamis.

2.2. Machine Learning dalam Deteksi Financial Fraud

Seiring dengan berkembangnya teknologi analisis data, pendekatan berbasis *machine learning* mulai banyak digunakan dalam sistem deteksi *fraud*. *Machine learning* memungkinkan sistem untuk mempelajari pola transaksi dari data historis dan mengidentifikasi anomali yang berpotensi menunjukkan aktivitas penipuan.

Berbagai algoritma *machine learning* telah digunakan dalam penelitian deteksi *fraud*, di antaranya *Logistic Regression*, *Decision Tree*, *Random Forest*, *Support Vector Machine*, dan *Neural Network*. Algoritma-algoritma tersebut mampu melakukan proses klasifikasi terhadap transaksi keuangan dengan memanfaatkan pola hubungan antar variabel dalam dataset transaksi. Dengan pendekatan ini, sistem dapat mengidentifikasi transaksi yang mencurigakan secara lebih akurat dibandingkan dengan metode konvensional berbasis aturan [11].

Selain itu, pendekatan *machine learning* juga memiliki kemampuan untuk memproses data dalam jumlah besar serta mendeteksi pola transaksi yang kompleks dan tidak mudah diidentifikasi oleh manusia. Hal ini menjadikan *machine learning* sebagai salah satu pendekatan yang sangat potensial dalam meningkatkan efektivitas sistem keamanan pada transaksi keuangan digital.

2.3. Permasalahan Ketidakseimbangan Data (Class Imbalance)

Meskipun *machine learning* memiliki kemampuan yang kuat dalam mendeteksi *fraud*, penerapannya dalam sistem transaksi keuangan masih menghadapi sejumlah tantangan. Salah satu tantangan utama adalah permasalahan ketidakseimbangan distribusi data atau *class imbalance*. Dalam dataset transaksi keuangan, jumlah transaksi normal biasanya jauh lebih banyak dibandingkan dengan transaksi *fraud*. Kondisi ini menyebabkan distribusi data menjadi tidak seimbang sehingga model *machine learning* cenderung lebih fokus mempelajari karakteristik dari kelas mayoritas.

Akibatnya, model yang dihasilkan sering menunjukkan tingkat akurasi yang tinggi secara keseluruhan, namun memiliki kemampuan yang rendah dalam mendeteksi transaksi *fraud* sebagai kelas minoritas. Kondisi ini dapat menyebabkan banyak kasus penipuan tidak terdeteksi oleh sistem. Penelitian menunjukkan bahwa ketidakseimbangan data merupakan salah satu faktor utama yang

mempengaruhi performa model klasifikasi dalam deteksi fraud pada sistem keuangan digital [12].

2.4. Penanganan Class Imbalance dengan Teknik SMOTE

Untuk mengatasi permasalahan ketidakseimbangan data, berbagai teknik penyeimbangan dataset telah dikembangkan dalam penelitian *machine learning*. Salah satu metode yang paling banyak digunakan adalah *Synthetic Minority Oversampling Technique* (SMOTE). Teknik ini bekerja dengan menghasilkan sampel sintetis pada kelas minoritas berdasarkan kedekatan karakteristik data yang telah ada.

Dengan menambahkan data sintetis pada kelas minoritas, distribusi dataset menjadi lebih seimbang sehingga algoritma klasifikasi dapat mempelajari pola fraud dengan lebih baik. Teknik SMOTE terbukti mampu meningkatkan kemampuan model dalam mendeteksi transaksi fraud serta mengurangi bias terhadap kelas mayoritas [13].

Beberapa penelitian juga mengombinasikan SMOTE dengan algoritma *machine learning* seperti *Random Forest*, *XGBoost*, maupun metode *ensemble learning* untuk meningkatkan performa model klasifikasi pada dataset yang tidak seimbang. Hasil penelitian menunjukkan bahwa penggunaan teknik penyeimbangan data dapat meningkatkan nilai *recall*, *precision*, dan *F1-score* pada proses deteksi fraud [14].

2.5. Relevansi Penelitian

Berdasarkan kajian literatur yang telah dilakukan, berbagai penelitian sebelumnya menunjukkan bahwa penerapan *machine learning* memiliki peran yang signifikan dalam meningkatkan kemampuan sistem deteksi *financial fraud*. Beberapa studi telah memanfaatkan berbagai algoritma seperti *Logistic Regression*, *Support Vector Machine*, dan *Random Forest* untuk mengidentifikasi pola transaksi yang mencurigakan dengan tingkat akurasi yang cukup baik [6], [9], [21]. Selain itu, pendekatan berbasis *ensemble learning* dan metode *deep learning* juga telah dikembangkan untuk meningkatkan performa deteksi *fraud* pada sistem keuangan digital yang semakin kompleks [14], [20]. Di sisi lain, sejumlah penelitian menekankan bahwa permasalahan ketidakseimbangan data (*class imbalance*) menjadi faktor utama yang mempengaruhi kinerja model klasifikasi, sehingga diperlukan teknik penanganan khusus seperti *oversampling*, *undersampling*, maupun metode hibrid seperti SMOTE-ENN untuk meningkatkan kemampuan deteksi pada kelas minoritas [12], [17], [19], [23].

Meskipun berbagai pendekatan tersebut telah menunjukkan hasil yang menjanjikan, masih terdapat keterbatasan dalam mengoptimalkan performa model pada dataset yang memiliki distribusi data tidak seimbang, khususnya dalam meningkatkan kemampuan model dalam mendeteksi transaksi *fraud* secara konsisten. Oleh karena itu, penelitian ini

berfokus pada penerapan teknik penyeimbangan data menggunakan metode *Synthetic Minority Oversampling Technique* (SMOTE) yang dikombinasikan dengan beberapa algoritma *machine learning*, yaitu *Logistic Regression*, *Support Vector Machine*, dan *Random Forest*. Pendekatan ini tidak hanya bertujuan untuk meningkatkan akurasi model, tetapi juga untuk memperbaiki kemampuan deteksi pada kelas minoritas melalui evaluasi komparatif antar model. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem deteksi *fraud* yang lebih adaptif, akurat, dan efektif dalam menghadapi dinamika transaksi keuangan digital yang terus berkembang.

3. METODE PENELITIAN

Penelitian ini bertujuan untuk meningkatkan kinerja sistem deteksi *financial fraud* pada transaksi keuangan digital melalui penerapan teknik penyeimbangan data menggunakan *Synthetic Minority Oversampling Technique* (SMOTE) yang dikombinasikan dengan beberapa algoritma *machine learning*. Secara umum, tahapan penelitian meliputi proses pengumpulan dataset, *preprocessing* data, penanganan ketidakseimbangan data, pembangunan model klasifikasi, serta evaluasi performa model. Seluruh tahapan penelitian dirancang secara sistematis agar model yang dihasilkan mampu mengidentifikasi transaksi yang berpotensi *fraud* secara lebih akurat dan adaptif terhadap karakteristik data transaksi [15], [16].

Alur penelitian yang digunakan dalam studi ini ditunjukkan pada Gambar 1. Diagram tersebut menggambarkan tahapan utama penelitian mulai dari proses pengolahan dataset hingga evaluasi performa model klasifikasi.



Gambar 1. Framework penelitian deteksi financial fraud

Berdasarkan Gambar 1, penelitian dimulai dengan proses pengumpulan dataset transaksi keuangan yang kemudian diproses melalui tahap *preprocessing*. Setelah dataset dipersiapkan, dilakukan proses penyeimbangan data menggunakan metode SMOTE untuk mengatasi permasalahan ketidakseimbangan distribusi data. Dataset yang telah diproses selanjutnya digunakan dalam pembangunan model klasifikasi menggunakan beberapa algoritma *machine learning*. Tahap terakhir adalah evaluasi performa model menggunakan berbagai metrik pengukuran yang umum digunakan dalam penelitian deteksi fraud.

3.1. Dataset Penelitian

Dataset yang digunakan dalam penelitian ini merupakan data transaksi keuangan yang memuat berbagai atribut yang merepresentasikan aktivitas transaksi pengguna. Dataset tersebut mencakup sejumlah variabel yang menggambarkan karakteristik transaksi, seperti informasi *merchant*, kategori transaksi, jumlah transaksi, serta atribut yang berkaitan dengan perilaku pengguna dalam melakukan aktivitas finansial [11].

Untuk memberikan gambaran yang lebih jelas mengenai atribut yang digunakan dalam penelitian ini, deskripsi variabel dataset disajikan pada Tabel 1. Tabel tersebut menunjukkan beberapa variabel utama yang digunakan sebagai fitur dalam proses pembangunan model klasifikasi.

Tabel 1. Deskripsi variabel dataset transaksi keuangan

No	Variabel	Deskripsi
1	merchant	nama merchant tempat transaksi dilakukan
2	category	kategori transaksi yang dilakukan pengguna
3	amt	jumlah nilai transaksi
4	gender	jenis kelamin pengguna
5	job	pekerjaan pengguna
6	city_pop	populasi kota tempat pengguna berada
7	unix_time	waktu transaksi dalam format timestamp
8	is_fraud	label transaksi (0 = normal, 1 = fraud)

Sebagaimana ditunjukkan pada Tabel 1, variabel *is_fraud* digunakan sebagai label target dalam proses klasifikasi untuk membedakan transaksi normal dan transaksi fraud. Secara umum, dataset transaksi keuangan memiliki distribusi data yang tidak seimbang karena jumlah transaksi normal jauh lebih besar dibandingkan dengan jumlah transaksi fraud. Kondisi ini menyebabkan terjadinya permasalahan *class imbalance* yang dapat memengaruhi performa model klasifikasi dalam mendeteksi aktivitas fraud.

3.2. Preprocessing Data

Tahap preprocessing data dilakukan untuk mempersiapkan dataset sebelum digunakan dalam proses pelatihan model machine learning. Proses ini bertujuan untuk meningkatkan kualitas data serta memastikan bahwa dataset memiliki format yang sesuai untuk proses analisis. [17].

Beberapa langkah yang dilakukan pada tahap ini meliputi pembersihan data (*data cleaning*), penghapusan atribut yang tidak relevan, serta transformasi data yang diperlukan. Atribut yang bersifat kategorikal dikonversi ke dalam bentuk numerik menggunakan teknik encoding sehingga dapat diproses oleh algoritma machine learning. Setelah proses transformasi data selesai dilakukan, dataset kemudian dipisahkan menjadi dua bagian, yaitu data pelatihan (*training data*) dan data pengujian (*testing data*) untuk keperluan evaluasi model.

3.3. Penanganan Ketidakseimbangan Data dengan SMOTE

Salah satu tantangan utama dalam dataset transaksi keuangan adalah ketidakseimbangan distribusi data antara kelas transaksi normal dan transaksi fraud. Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan metode Synthetic Minority Oversampling Technique (SMOTE) sebagai teknik penyeimbangan data [18].

SMOTE bekerja dengan menghasilkan data sintesis pada kelas minoritas berdasarkan karakteristik data yang telah ada. Proses ini dilakukan dengan membuat sampel baru yang berada di antara titik-titik data yang memiliki kedekatan karakteristik dalam ruang fitur. Dengan pendekatan ini, jumlah sampel pada kelas minoritas, dapat ditingkatkan sehingga distribusi dataset menjadi lebih seimbang [19].

Penerapan teknik SMOTE dilakukan pada data pelatihan sebelum proses pembangunan model klasifikasi. Dengan demikian, model *machine learning* dapat mempelajari pola transaksi fraud secara lebih representatif dan mengurangi kecenderungan bias terhadap kelas mayoritas.

3.4. Pembangunan Model Machine Learning

Setelah dataset berhasil diseimbangkan menggunakan teknik SMOTE, tahap selanjutnya adalah pembangunan model klasifikasi menggunakan beberapa algoritma *machine learning*. Dalam penelitian ini digunakan tiga algoritma yang umum digunakan dalam penelitian deteksi fraud, yaitu *Logistic Regression*, *Support Vector Machine*, dan *Random Forest* [20].

Setiap algoritma digunakan untuk membangun model klasifikasi yang bertujuan membedakan transaksi normal dan transaksi fraud berdasarkan pola yang terdapat dalam dataset. Model dilatih menggunakan data pelatihan yang telah diproses sebelumnya dan kemudian diuji menggunakan data pengujian untuk mengetahui kemampuan model dalam

melakukan prediksi terhadap data baru yang belum pernah dipelajari sebelumnya.

Penggunaan beberapa algoritma klasifikasi dalam penelitian ini juga bertujuan untuk melakukan analisis perbandingan performa antar model sehingga dapat diperoleh model yang memiliki kinerja terbaik dalam mendeteksi aktivitas fraud pada transaksi keuangan [21].

3.5. Evaluasi Model

Tahap akhir dalam penelitian ini adalah evaluasi performa model klasifikasi yang telah dibangun. Evaluasi dilakukan menggunakan beberapa metrik pengukuran yang umum digunakan dalam penelitian deteksi fraud, yaitu accuracy, precision, recall, dan F1-score. Selain itu, analisis juga dilakukan menggunakan confusion matrix untuk memberikan gambaran yang lebih rinci mengenai kemampuan model dalam mengklasifikasikan transaksi normal dan transaksi fraud [22].

Penggunaan metrik *precision* dan *recall* menjadi sangat penting dalam penelitian ini karena berkaitan langsung dengan kemampuan sistem dalam mendeteksi transaksi fraud sebagai kelas minoritas. Dengan melakukan evaluasi menggunakan berbagai metrik tersebut, penelitian ini dapat memberikan gambaran yang lebih komprehensif mengenai performa model klasifikasi yang digunakan dalam sistem deteksi fraud pada transaksi keuangan digital [23].

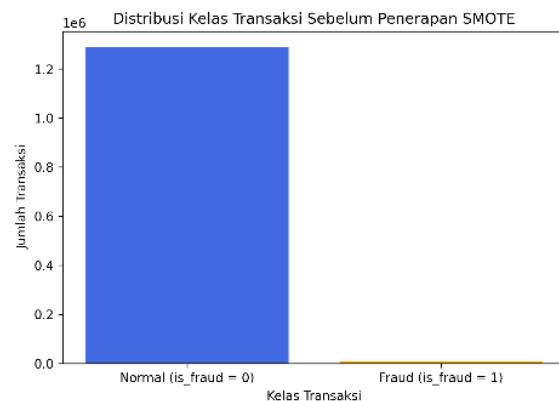
4. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk meningkatkan kinerja sistem deteksi *financial fraud* pada transaksi keuangan digital melalui penerapan teknik penyeimbangan data menggunakan *Synthetic Minority Oversampling Technique* (SMOTE) yang dikombinasikan dengan beberapa algoritma *machine learning*. Proses analisis dilakukan secara bertahap, dimulai dari analisis karakteristik dataset, penerapan teknik penyeimbangan data, hingga evaluasi performa model klasifikasi yang dibangun. Pendekatan ini diharapkan mampu menghasilkan model yang lebih adaptif dalam mendeteksi transaksi yang berpotensi fraud pada dataset yang memiliki distribusi data tidak seimbang [18].

Tahap awal penelitian difokuskan pada analisis karakteristik dataset transaksi keuangan yang digunakan. Dataset tersebut terdiri dari sejumlah atribut yang merepresentasikan aktivitas transaksi pengguna, seperti informasi *merchant*, kategori transaksi, jumlah transaksi, serta atribut lain yang berkaitan dengan perilaku pengguna dalam melakukan aktivitas finansial [11]. Karakteristik atribut yang digunakan dalam penelitian ini telah dijelaskan sebelumnya pada Tabel 1. Deskripsi atribut dataset transaksi keuangan pada bagian metode penelitian.

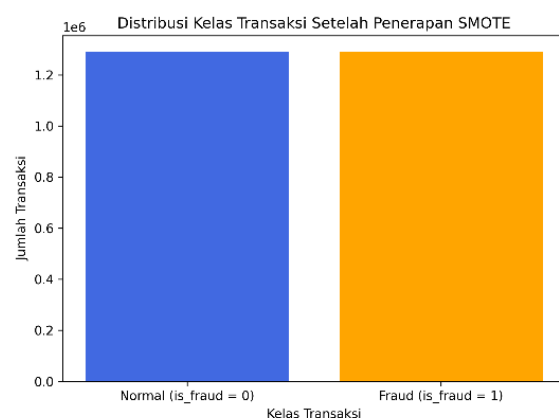
Hasil analisis awal terhadap dataset menunjukkan bahwa distribusi data antara transaksi normal dan transaksi fraud tidak seimbang. Jumlah

transaksi normal mendominasi dataset secara signifikan dibandingkan jumlah transaksi fraud. Kondisi ini merupakan karakteristik umum dalam sistem deteksi fraud, di mana aktivitas penipuan biasanya hanya terjadi dalam proporsi yang relatif kecil dibandingkan dengan total transaksi yang terjadi. Distribusi kelas transaksi sebelum penerapan teknik penyeimbangan data ditunjukkan pada Gambar 2 [13], [14].



Gambar 2. Distribusi kelas transaksi sebelum penerapan SMOTE.

Ketidakseimbangan distribusi data tersebut berpotensi mempengaruhi performa model klasifikasi yang dibangun. Dalam kondisi dataset yang tidak seimbang, algoritma *machine learning* cenderung lebih dominan mempelajari pola dari kelas mayoritas sehingga kemampuan model dalam mendeteksi transaksi fraud sebagai kelas minoritas menjadi kurang optimal. Akibatnya, model dapat menghasilkan nilai akurasi yang tinggi secara keseluruhan, namun kurang efektif dalam mengidentifikasi transaksi yang benar-benar termasuk dalam kategori fraud [12], [14].



Gambar 3. Distribusi kelas transaksi setelah penerapan SMOTE.

Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan teknik penyeimbangan data menggunakan *Synthetic Minority Oversampling Technique* (SMOTE). Teknik SMOTE bekerja dengan menghasilkan data sintesis pada kelas minoritas

berdasarkan karakteristik data yang telah ada, sehingga distribusi dataset menjadi lebih seimbang. Dengan distribusi data yang lebih proporsional, model *machine learning* memiliki peluang yang lebih baik untuk mempelajari pola transaksi fraud secara lebih representatif. Perubahan distribusi dataset setelah penerapan teknik penyeimbangan data ditunjukkan pada Gambar 3 [13].

Setelah proses penyeimbangan data dilakukan, tahap selanjutnya adalah pembangunan model klasifikasi menggunakan beberapa algoritma *machine learning*, yaitu *Logistic Regression*, *Support Vector Machine*, dan *Random Forest*. Ketiga algoritma tersebut digunakan untuk membangun model deteksi fraud yang kemudian dievaluasi menggunakan beberapa metrik pengukuran yang umum digunakan dalam penelitian klasifikasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil evaluasi performa model klasifikasi dalam penelitian ini ditunjukkan pada Tabel 2 [6].

Tabel 2. Perbandingan performa model machine learning dalam deteksi financial fraud.

Algoritma	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.93	0.88	0.84	0.86
Support Vector Machine	0.95	0.91	0.88	0.89
Random Forest	0.97	0.94	0.92	0.93

Berdasarkan hasil evaluasi yang ditampilkan pada Tabel 2, terlihat bahwa setiap algoritma memiliki kemampuan yang berbeda dalam mendeteksi transaksi fraud. Perbedaan performa tersebut dipengaruhi oleh karakteristik masing-masing algoritma dalam mempelajari hubungan antar fitur dalam dataset transaksi keuangan. Algoritma *Logistic Regression* menunjukkan performa yang cukup baik dalam melakukan klasifikasi, namun memiliki keterbatasan dalam menangkap pola hubungan yang lebih kompleks dalam dataset [6].

Sementara itu, algoritma *Support Vector Machine* menunjukkan peningkatan performa dibandingkan *Logistic Regression* karena kemampuannya dalam membangun batas keputusan yang lebih optimal dalam ruang fitur. Namun demikian, hasil penelitian menunjukkan bahwa algoritma *Random Forest* memberikan performa terbaik dibandingkan algoritma lainnya dalam mendeteksi transaksi fraud. Hal ini disebabkan oleh kemampuan algoritma *Random Forest* dalam menggabungkan beberapa pohon keputusan sehingga mampu menangkap pola hubungan yang lebih kompleks dalam dataset transaksi keuangan [9].

Selain itu, pendekatan *ensemble learning* yang digunakan dalam *Random Forest* juga membantu meningkatkan stabilitas model serta mengurangi risiko *overfitting* pada proses pelatihan model. Kemampuan

ini menjadikan *Random Forest* lebih efektif dalam mengidentifikasi transaksi fraud yang memiliki pola yang tidak selalu linier dan seringkali tersembunyi dalam dataset transaksi [5].

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa penerapan teknik penyeimbangan data menggunakan SMOTE mampu meningkatkan kemampuan model *machine learning* dalam mendeteksi transaksi fraud pada dataset yang memiliki distribusi data tidak seimbang. Kombinasi teknik SMOTE dengan algoritma *Random Forest* menghasilkan performa klasifikasi yang lebih optimal dibandingkan dengan algoritma lainnya dalam penelitian ini. Temuan ini menunjukkan bahwa pendekatan berbasis *machine learning* yang didukung dengan teknik penyeimbangan data dapat menjadi solusi yang efektif dalam meningkatkan keamanan sistem transaksi keuangan digital.

5. KESIMPULAN DAN SARAN

Kesimpulan dari penelitian ini menunjukkan bahwa penerapan teknik penyeimbangan data menggunakan *Synthetic Minority Oversampling Technique* (SMOTE) yang dikombinasikan dengan algoritma *machine learning* mampu meningkatkan kinerja sistem deteksi *financial fraud* pada transaksi keuangan digital. Proses penyeimbangan data terbukti efektif dalam mengatasi permasalahan *class imbalance* yang umum terjadi pada dataset transaksi keuangan, sehingga model klasifikasi dapat mempelajari pola transaksi *fraud* secara lebih representatif. Berdasarkan hasil evaluasi performa model, algoritma *Random Forest* menunjukkan kinerja terbaik dibandingkan *Logistic Regression* dan *Support Vector Machine* dalam mendeteksi transaksi *fraud*, karena kemampuannya dalam menangkap pola hubungan yang lebih kompleks pada dataset transaksi keuangan. Temuan ini menunjukkan bahwa kombinasi antara teknik penyeimbangan data dan pendekatan *machine learning* dapat menjadi solusi yang efektif untuk meningkatkan akurasi dan keandalan sistem deteksi *fraud* pada lingkungan transaksi digital yang semakin berkembang. Untuk penelitian selanjutnya, disarankan agar dilakukan eksplorasi terhadap penggunaan algoritma *machine learning* lainnya seperti *gradient boosting* atau *deep learning*, serta pengujian pada dataset dengan skala yang lebih besar dan variasi fitur yang lebih beragam guna memperoleh model deteksi *fraud* yang lebih robust dan adaptif terhadap perkembangan pola kejahatan finansial di masa mendatang.

DAFTAR PUSTAKA

- [1] K. P. Ramesh, R. Amudha, K. Prasob, and K. S. Kanna, "Fintech innovations in E-payments: Privacy and security in cybercrime threats," *Multidiscip. Sci. J.*, vol. 5, p. 2023ss0320, Aug. 2023, doi: 10.31893/multiscience.2023ss0320.
- [2] O. Angela, Adesola Soyelee, and Emmanuel Ogunwobi, "Enhancing fraud detection and

- prevention in fintech: Big data and machine learning approaches,” *World J. Adv. Res. Rev.*, vol. 24, no. 2, pp. 2301–2319, Nov. 2024, doi: 10.30574/wjarr.2024.24.2.3617.
- [3] R. Y. Yadav and R. Maurya, “The Role of Ai and Machine Learning in Financial Fraud Detection,” *IJFMR*, vol. 7, no. 3, p. 46023, Jun. 2025, doi: 10.36948/ijfmr.2025.v07i03.46023.
- [4] Iftekhar Rasul, S M Iftekhar Shaboj, Mainuddin Adel Rafi, Md Kauser Miah, Md Redwanul Islam, and Abir Ahmed, “Detecting Financial Fraud in Real-Time Transactions Using Graph Neural Networks and Anomaly Detection,” *JEFAS*, vol. 6, no. 1, pp. 131–142, Feb. 2024, doi: 10.32996/jefas.2024.6.1.13.
- [5] A. Rajkumar, “Machine Learning Techniques for Real-Time Fraud Detection in Financial Transactions,” *lbr*, pp. 97–103, May 2025, doi: 10.56595/lbr.v4i1.53.
- [6] P. Manorom, U. Detthamrong, and W. Chansanam, “Comparative Assessment of Fraudulent Financial Transactions using the Machine Learning Algorithms Decision Tree, Logistic Regression, Naïve Bayes, K-Nearest Neighbor, and Random Forest,” *Eng. Technol. Appl. Sci. Res.*, vol. 14, no. 4, pp. 15676–15680, Aug. 2024, doi: 10.48084/etasr.7774.
- [7] Akshita Sunerah, “Detection of Credit Card Fraud Utilizing Transaction History Using Machine Learning,” *jisem*, vol. 10, no. 10s, pp. 660–669, Feb. 2025, doi: 10.52783/jisem.v10i10s.1457.
- [8] N. Susmitha and Smt. A.Kalpna, “Fraud Detection in Banking Data Using Machine Learning Techniques,” *JES*, vol. 16, no. 04, pp. 120–125, 2025, doi: 10.36893/JES.2025.V16I04.020.
- [9] T. K. Gawali *et al.*, “Machine Learning-Based Credit Card Fraud Detection: A Comparative Study Using Logistic Regression, SVM, and KNN,” *Int. J. Environ. Sci.*, vol. 11, no. 8s, pp. 375–380, Jun. 2025, doi: 10.64252/1ph35z52.
- [10] Srinivasa Kalyan Vangibhurathachhi, “Machine Learning for Fraud Detection in Financial Transactions,” *IJSAT*, vol. 16, no. 1, p. 3727, Mar. 2025, doi: 10.71097/IJSAT.v16.i1.3727.
- [11] Precious Ngulube and Ms Fanny Chatola, “Machine Learning for Fraud Detection In Financial Transactions,” *IJARST*, pp. 417–424, Jan. 2025, doi: 10.48175/IJARST-22945.
- [12] M. Isangediok and K. Gajamannage, “Fraud Detection Using Optimized Machine Learning Tools Under Imbalance Classes,” in *2022 IEEE International Conference on Big Data (Big Data)*, Osaka, Japan: IEEE, Dec. 2022, pp. 4275–4284. doi: 10.1109/BigData55660.2022.10020723.
- [13] Ajay Kumar, “Voting Classifier as a Balanced Framework for Fraud Detection in Imbalanced Credit Card Transactions,” *jisem*, vol. 9, no. 4s, pp. 2393–2410, Dec. 2024, doi: 10.52783/jisem.v9i4s.12735.
- [14] T. Albalawi and S. Dardouri, “Enhancing credit card fraud detection using traditional and deep learning models with class imbalance mitigation,” *Front. Artif. Intell.*, vol. 8, p. 1643292, Oct. 2025, doi: 10.3389/frai.2025.1643292.
- [15] N. Susmitha and Smt. A.Kalpna, “Fraud Detection in Banking Data Using Machine Learning Techniques,” *JES*, vol. 16, no. 04, pp. 120–125, 2025, doi: 10.36893/JES.2025.V16I04.020.
- [16] S. T. Bhavitavya Isukapati-, S. J. -, and S. G. S. -, “Machine Learning Approaches for Effective Credit Card Fraud Detection: Addressing Imbalance and Enhancing Accuracy,” *IJFMR*, vol. 7, no. 1, p. 35045, Jan. 2025, doi: 10.36948/ijfmr.2025.v07i01.35045.
- [17] E. M. Al-dahasi, R. K. Alsheikh, F. A. Khan, and G. Jeon, “Optimizing fraud detection in financial transactions with machine learning and imbalance mitigation,” *Expert Systems*, vol. 42, no. 2, p. e13682, Feb. 2025, doi: 10.1111/exsy.13682.
- [18] M. Tayebi and S. El Kafhali, “Generative Modeling for Imbalanced Credit Card Fraud Transaction Detection,” *JCP*, vol. 5, no. 1, p. 9, Mar. 2025, doi: 10.3390/jcp5010009.
- [19] I. Syamsuddin, S. Omsa, A. Rustam, and D. Hasan, “Detecting fraud activities in financial transactions using SMOTENN model,” *jour*, vol. 25, no. 3, pp. 466–474, Jul. 2025, doi: 10.17586/2226-1494-2025-25-3-466-474.
- [20] C.-W. Lee, M.-W. Fu, C.-C. Wang, and Muh. I. Azis, “Evaluating Machine Learning Algorithms for Financial Fraud Detection: Insights from Indonesia,” *Mathematics*, vol. 13, no. 4, p. 600, Feb. 2025, doi: 10.3390/math13040600.
- [21] F. Adamu-Fika *et al.*, “A Comparative Evaluation of Machine Learning Models for an Enhanced Fraud Detection in Financial Systems,” *AIMS Research Journal Publication*, vol. 16, no. 2, pp. 13–22, Jun. 2025, doi: 10.22624/AIMS/CISDI/V16N2P2.
- [22] S. S. Kaddi and M. M. Patil, “Ensemble learning based health care claim fraud detection in an imbalance data environment,” *IJECS*, vol. 32, no. 3, p. 1686, Dec. 2023, doi: 10.11591/ijeecs.v32.i3.pp1686-1694.
- [23] Ms. Hemangini Patel, Mrugesh Patel, and Ms. Krinal Savani, “An Optimized XGBoost Framework for Real-Time Credit Card Fraud Detection: Addressing Class Imbalance with Hybrid SMOTE-ENN Resampling,” *Int J Sci Res Sci & Technol*, vol. 12, no. 3, pp. 1129–1136, Jun. 2025, doi: 10.32628/IJSRST25123122.