

Peran *Big Data Analytics*, *Machine Learning*, dan *Artificial Intelligence* dalam Pendeteksian *Financial Fraud*: A SYSTEMATIC LITERATURE REVIEW

Finecia Shinta Dewi, Totok Dewayanto¹

Departemen Akuntansi, Fakultas Ekonomika dan Bisnis, Universitas Diponegoro

Jl. Prof. Soedharto SH, Tembalang, Semarang 50239, Indonesia

Phone: +62 24 76486851

ABSTRACT

This research aims to explore the critical role of big data analytics, machine learning, and artificial intelligence in detecting financial fraud within financial institutions. The research is based on published research articles.

Utilizing a Systematic Literature Review with the PRISMA protocol, an analysis was conducted on 20 articles published between 2020 and 2024, sourced from the Scopus database. The findings were categorized into three areas: the role of big data analytics in financial fraud detection, the role of machine learning in financial fraud detection, and the role of artificial intelligence in financial fraud detection.

The results indicated that financial fraud detection systems employing big data analytics (BDA) demonstrated a significant average strength (76.67%), particularly in detection effectiveness, accuracy, and data processing speed. The implementation of artificial intelligence (AI) in detection also showed significant strength scores. In contrast to BDA and AI, some machine learning algorithms exhibited substantial weaknesses. Addressing these weaknesses in financial fraud detection at financial institutions, future research on the integration of machine learning algorithms is deemed crucial.

Keywords: artificial intelligence, big data analytics, machine learning, financial fraud, SLR.

PENDAHULUAN

Perjalanan Teknologi dan pengaruhnya terhadap masyarakat telah mengalami transformasi mendalam sejak revolusi industri 1.0 hingga *society 5.0*. Memasuki era *society 5.0*, muncul pengintegrasian teknologi mutakhir seperti AI dan *big data* ke dalam tatanan masyarakat untuk meningkatkan kualitas hidup manusia serta menciptakan harmoni antara kemajuan teknologi dan kesejahteraan sosial (Al-imohammadlou & Khoshsepehr, 2023). Integrasi ini memicu transformasi yang signifikan seperti mengubah fundamental cara lembaga keuangan dan para profesional mengelola dan melaporkan data keuangannya.

Menurut Broby (2021), kemajuan dalam teknologi keuangan telah memberikan dampak yang signifikan terhadap peningkatan efisiensi proses keuangan, termasuk otomatisasi tugas operasional, analisis data yang lebih komprehensif, peningkatan kapasitas penyimpanan data, dan integrasi perangkat lunak yang lebih efisien. Semakin efisien, semakin banyak celah keamanan yang muncul, menjadi pendorong utama bagi peningkatan kecurangan keuangan atau disebut dengan “*financial fraud*”. Dengan kehadiran otomatisasi yang semakin maju, seperti yang disoroti oleh Anna Kooi, pemimpin layanan keuangan nasional di Wipfli LLP, pelaku kejahatan dapat memanfaatkan perangkat lunak dan bot untuk melakukan kecurangan dengan frekuensi yang lebih tinggi dan dalam skala yang lebih luas daripada sebelumnya (Kuehner-Hebert, 2024).

Peningkatan *financial fraud* dapat berdampak buruk bagi sebuah perusahaan dan lembaga keuangan, maka diperlukan pengadopsian pendekatan proaktif untuk mendeteksi kecurangan. Pendekatan ini harus mengimplentasikan teknologi terkini untuk menganalisis data secara mendalam dan mengidentifikasi pola yang mencurigakan. Salah satu solusi yang semakin diandalkan dalam upaya mendeteksi kecurangan keuangan adalah analitik data besar atau dikenal dengan sebutan “*big data analytics*”. Dengan menggunakan algoritma dan teknik analisis yang canggih, analitik data memungkinkan perusahaan untuk menggali informasi berharga dari volume besar data transaksi keuangan dan perilaku pelaku bisnis (Kesuma & Sunarto, 2020).

Seiring berkembangnya tren dan pola kecurangan, pengoptimalan pendeteksian menjadi hal yang sangat esensial. Gupta (2024) menemukan bahwa pemanfaatan *machine learning*, *data analytics*, dan sistem *artificial intelligence* dapat meningkatkan deteksi transaksi kecurangan dengan presisi yang lebih besar. Dengan memanfaatkan solusi *big data* dan teknik *machine learning*, deteksi kecurangan dapat dilakukan secara real-time dengan tingkat akurasi yang tinggi (Abbassi et al., 2021).

Dengan Demikian, berikut adalah pertanyaan penelitian dalam penelitian ini:

RQ1: Apakah *big data analytics*, *machine learning*, dan *artificial intelligence*

memiliki peran penting dalam pendeteksian *financial fraud* di lembaga keuangan?

RQ2: Bagaimana perkembangan penelitian tentang *big data analytics*, *machine learning*, dan *artificial intelligence*?

RQ3: Apa arah penelitian yang dapat dieksplorasi di masa mendatang dalam usaha pendeteksian *financial fraud*?

Penelitian ini bertujuan untuk melakukan *systematic literature review* (SLR) mengenai pentingnya peran *big data analytics*, *machine learning*, dan *artificial intelligence* dalam pendeteksian *financial fraud* di lembaga keuangan serta mengeksplorasi dan meninjau variasi hasil temuan yang terdapat dalam penelitian empiris terkait peran *big data analytics*, *machine learning*, dan *artificial intelligence* dalam pendeteksian *financial fraud* di lembaga keuangan dan melakukan sintesis pada temuan yang didapat.

TINJAUAN PUSTAKA DAN KERANGKA PEMIKIRAN TEORITIS

Bagian ini menjelaskan teori dan konsep serta kerangka pemikiran yang digunakan dalam penelitian.

Data Science Theory

Data science adalah bidang interdisipliner yang mengeksplorasi proses pengumpulan, analisis, interpretasi, dan pemodelan data untuk mendapatkan wawasan yang berharga.

METODE PENELITIAN

Bagian ini menjelaskan perumusan pertanyaan penelitian, strategi pencarian literatur, kriteria literatur, dan seleksi literatur.

Perumusan Pertanyaan Penelitian (*Research Question*)

Sebagai acuan dalam merumuskan pertanyaan penelitian, diadopsi kerangka kerja *population*, *intervention*, *comparasion* dan *outcome* (PICO) untuk memperoleh kata kunci yang mempermudah proses pencarian.

Tabel 1
Framework PICO

PICO Tool	
Population	Lembaga Keuangan
Intervention	<i>Big data analytics, machine learning, dan artificial intelligence</i>
Comparison	-
Outcome	Pendeteksian <i>Financial Fraud</i>

Sumber: Data Diolah, 2024

Berdasarkan kerangka PICO yang dirumuskan tersebut, kata kunci yang digunakan adalah *big data analytics, machine learning, dan artificial intelligence; financial fraud detection; dan financial institutions*. Kata kunci tersebut merupakan kata kunci dasar yang akan dikembangkan saat melakukan pencarian literatur.

Strategi Pencarian Literatur

Sebelum memulai pencarian, diperlukan pemilihan database yang sesuai untuk meningkatkan kemungkinan menemukan artikel yang sangat relevan. Database yang dipilih adalah Scopus, sehingga data yang dipakai dalam penelitian adalah data sekunder. Hasil-hasil penelitian dalam jurnal yang dipublikasikan secara *on-line* merupakan data tersebut. Untuk mencari artikel yang relevan pada Scopus, perlu mendefinisikan *search string* terlebih dahulu. Berikut merupakan langkah-langkah yang digunakan untuk membuat *search string* pada Scopus:

1. Mengembangkan kata/istilah pencarian dari PICO, terutama dari *population, intervention, dan outcome*
2. Mengembangkan kata/istilah pencarian dengan kata kunci yang relevan
3. Mengembangkan sinonim dan ejaan alternatif dari kata/istilah pencarian
4. Masukkan *search string* ke dalam mesin pencarian Scopus.
5. Mengkombinasikan *search string* menggunakan kata/istilah pencarian yang telah dikembangkan dengan menggunakan operator *boolean* (AND dan OR).

Search string yang akhirnya digunakan adalah ((*"big data" OR "big data analy*" OR "artificial intelligence" OR "machine learning" OR "deep learning"*) AND (*fraud OR "fraud detection" OR "financial fraud"*) AND (*"financial institutions" OR "financial industry" OR "bank" OR "insurance" OR corporate OR fintech*)).

Kriteria Literatur

Kriteria literatur dibagi menjadi kriteria inklusi dan eksklusi. Kriteria pada penelitian ini disesuaikan dengan *framework* PICO yang sebelumnya telah dibuat dengan tambahan sebagai berikut

Tabel 2
Kriteria Inklusi dan Eksklusi

Kriteria	Inklusi	Eksklusi
Subjek	Topik penelitian berkaitan dengan pendeteksian <i>financial fraud</i>	Topik penelitian tidak berkaitan dengan pendeteksian <i>financial fraud</i>
Bahasa	Bahasa Inggris	Bahasa selain Inggris
Sumber	Artikel penelitian yang <i>open access</i>	Artikel penelitian yang tidak <i>open access</i>
Jangka Waktu	Tahun 2020-2024	Sebelum tahun 2020-2024
Tema isi Jurnal	Membahas mengenai peran <i>big data analytics</i> , <i>machine learning</i> , dan <i>artificial intelligence</i> dalam pendeteksian <i>financial fraud</i>	Artikel tidak sesuai dengan <i>research question</i> setelah dilakukan analisis secara detail

Sumber: Data Diolah, 2024

Seleksi Literatur

Data dikumpulkan melalui database Scopus dan dianalisis menggunakan kerangka kerja PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) yang terdiri dari 3 tahapan sebagai berikut.

Gambar 3
PRISMA Flow Diagram

;Riyan;

Sumber: Data Diolah, 2024

HASIL PENELITIAN DAN PEMBAHASAN

Bagian ini berisi pembahasan temuan hasil penelitian dan sintesis temuan hasil penelitian.

Temuan Hasil Penelitian

Temuan hasil penelitian dikelompokkan dalam triad klasifikasi, yaitu peran penting *big data analytics* dalam pendeteksian *financial fraud* di lembaga keuangan, serta peran penting *machine learning* dalam pendeteksian *financial fraud* di lembaga keuangan, serta peran penting *artificial intelligence* dalam pendeteksian *financial fraud* di lembaga keuangan.

Tabel 3
Peran Penting *Machine Learning* dalam Pendeteksian *Financial Fraud* di Lembaga Keuangan

No	Judul Artikel	Peneliti	Temuan
----	---------------	----------	--------

1	<i>Auto Insurance Fraud Detection with Multimodal Learning</i>	ıRiyanı	Kerangka kerja <i>auto insurance multi-modal learning</i> (AIML), yang meingintegrasikan pemrosesan bahasa alami teknik visi komputer meningkatkan deteksi kecurangan secara signifikan dengan memanfaatkan data <i>multi-modal</i> untuk prediksi perilaku kecurangan yang lebih akurat
2	<i>Automatic Machine learning Algorithms for Fraud Detection in Digital Payment Systems</i>	ıRiyanı	Penggunaan algoritma pembelajaran mesin otomatis dalam sistem pembayaran digital menunjukkan efektivitas tidak hanya melalui kualitas pengklsifikasian dan pengurangan biaya pengembangan, tetapi juga melalui peningkatan potensi interpretabilitas, yang sangat diperlukan untuk mengatasi tantangan yang berkembang dalam mendeteksi transaksi kecurangan.

3	<i>Chinese Corporate Fraud Risk Assessment with Machine Learning</i>	ıRiyanı	Memanfaatkan berbagai algoritma <i>machine learning</i> untuk meningkatkan akurasi model dan <i>area under the curve</i> (AUC) dan <i>receiver operating characteristic</i> (ROC) adalah kunci dalam deteksi kecurangan, dengan sumber data <i>real-time</i> dan analisis sentimen dari berita atau media sosial yang memperkuat kerangka kerja melawan taktik kecurangan yang terus berkembang.
4	<i>Detecting Fraud Transaction using Ripper Algorithm Combines with Ensemble Learning Mode</i>	ıRiyanı	Penggabungan algoritma <i>Ripper</i> dengan model <i>ensemble learning</i> terutama menggunakan <i>gradient boosting</i> , menghasilkan akurasi deteksi lebih dari 99%, menunjukkan efektivitas pembelajaran mesin dalam meningkatkan kemampuan deteksi kecurangan.
5	<i>Detecting Insurance Fraud Using Supervised and Unsupervised Machine Learning</i>	ıRiyanı	Studi ini mengevaluasi metode <i>unsupervised</i> dan <i>supervised machine learning</i> untuk deteksi <i>insurance fraud</i> menggunakan data klaim eksklusif, menemukan bahwa <i>isolation forest</i> dalam <i>unsupervised learning</i> sangat efektif, sementara metode <i>supervised learning</i> juga berkinerja baik meskipun dengan jumlah kasus kecurangan berlabel terbatas.

6	<i>Federated Learning Model for Credit Card Fraud Detection with Data Balancing Techniques</i>	ıRiyanı	<i>Federated learning model meningkatkan deteksi kecurangan kartu kredit dengan memungkinkan kolaborasi antar bank tanpa berbagi data sensitif, manajemen ketidakseimbangan kelas melalui resampling hibrida meningkatkan kinerja model klasifikasi, dan pengklasifikasi random forest menonjol sebagai metode terbaik berdasarkan parameter kinerja dalam mendeteksi transaksi fraud.</i>
---	--	---------	--

Sumber: Data Diolah, 2024

Berikut merupakan kesimpulan mengapa *machine learning* berperan penting dalam deteksi *financial fraud* menurut hasil penelitian yang telah dijabarkan di atas.

1. Data *multi-modal* meningkatkan akurasi deteksi kecurangan, penggunaan data teks dan visual bersama-sama memperkaya informasi yang dimiliki model, sehingga dapat memprediksi perilaku curang dengan lebih baik.
2. Algoritma *automated machine learning* dapat menyintesis model deteksi kecurangan dengan lebih efisien dan mengurangi biaya pengembangan, sambil menjaga kualitas hasil deteksi.
3. Pemanfaatan beragam algoritma *machine learning* dan sumber data *real-time* dapat memperkuat model deteksi kecurangan terhadap taktik kecurangan yang terus berkembang.
4. Model *machine learning* seperti *random forest* yang dioptimalkan dengan teknik *oversampling* dapat unggul dalam mendeteksi kecurangan akuntansi, menunjukkan pentingnya pemilihan dan optimasi model.
5. Metode *federated learning* memungkinkan kolaborasi antar bank tanpa mengorbankan privasi data, meningkatkan deteksi kecurangan kartu kredit dengan mengumpulkan lebih banyak data untuk pelatihan model.

Sintesis Temuan Hasil Penelitian

Bagian ini akan menyajikan sintesis dari 20 artikel temuan, yang secara khusus dirancang untuk menjawab pertanyaan penelitian atau *research question* yang telah dirumuskan sebelumnya.

Kekuatan dan Kelamahan *Big Data Analytics*, *Machine Learning*, dan *Artificial Intelligence* dalam Pendeteksian *Financial Fraud* di Lembaga Keuangan

Untuk mengetahui apakah *big data analytics*, *machine learning*, dan *artificial intelligence* berperan penting dalam pendeteksian *financial fraud* di lembaga keuangan, maka dibutuhkan penjabaran tentang kekuatan dan kelemahan masing-masing variabel tersebut dalam pendeteksian *financial fraud* di lembaga keuangan.

Tabel 6

Persentase Penilaian *Big Data Analytics* dalam Pendeteksian *Financial Fraud* di Lembaga Keuangan

Kategori	Kekuatan (%)	Kelemahan (%)
Efektivitas Deteksi	90	10
Akurasi	85	15
Kecepatan	80	20
Kemudahan Penggunaan	70	30
Biaya	60	40
Adaptabilitas	75	25

Sumber: Data Diolah, 2024

Perkembangan Penelitian BDA, ML, dan AI dalam Pendeteksian *Financial Fraud*

Penelitian tentang analisis *big data* dalam deteksi kecurangan telah berkembang pesat sejak awal abad ke-21, didorong oleh krisis keuangan tahun 2008

dan regulasi yang memaksa sektor perbankan untuk lebih transparan serta memanfaatkan data pelanggan secara luas. Deteksi kecurangan diidentifikasi sebagai aplikasi penting *big data analytics* (BDA) dalam perbankan, membantu meningkatkan pengambilan keputusan dan manajemen risiko. Penelitian *machine learning* dalam deteksi kecurangan telah berkembang dari model tunggal ke model ansambel untuk meningkatkan akurasi dan efisiensi, dengan pengenalan algoritma seperti regresi logistik, SVM, dan *XGBoost*. Peran AI juga telah berkembang untuk memerangi ancaman *cyber* yang kompleks seperti pencurian identitas dan *malware*, serta meningkatkan langkah-langkah keamanan secara keseluruhan. Prospek masa depan AI dalam deteksi kecurangan sangat menjanjikan, dengan kemajuan berkelanjutan yang diharapkan dapat memperkuat langkah-langkah keamanan siber secara efektif.

Ekplorasi Arah Penelitian Masa Depan

Dari pembahasan sintesis temuan hasil penelitian, dapat terlihat bahwa *big data analytics* dan *artificial intelligence* sudah cukup efektif dalam mendeteksi *financial fraud* menurut skor penilaian. Sedangkan, beberapa algoritma pada *machine learning* memiliki skor kelemahan yang lumayan besar, contohnya adalah *federated learning* dan jaringan *neural grafik*, khususnya model HAN. Untuk itu, ada beberapa integrasi antar algoritma yang dapat dilakukan untuk mengatasi kelemahan yang ada pada kedua algoritma tersebut.

Kesimpulan dan Rekomendasi

Bagian ini berisi kesimpulan dan rekomendasi untuk penelitian selanjutnya.

Kesimpulan

Kesimpulan dari penelitian ini adalah penerapan *big data analytics* (BDA), *machine learning* (ML), dan *artificial intelligence* (AI) sangat berperan penting dalam pendeteksian *financial fraud*, dilihat dari rata-rata kekuatan masing-masing teknologi tersebut yang menunjukkan kekuatan yang signifikan. Sistem deteksi *financial fraud* dengan menggunakan BDA dan AI masing-masing menunjukkan rata-rata kekuatan yang sama, yaitu sebesar 76.67%. Di sisi lain, dalam konteks pendeteksian *financial fraud* menggunakan *machine learning*, terdapat beberapa algoritma yang memiliki skor kelemahan yang tinggi. Algoritma tersebut adalah *GNN* dan *federated learning*. Untuk mengatasi kelemahan tersebut, dibutuhkan pengintegrasian dengan algoritma yang memiliki skor kekuatan yang tinggi. Oleh

karena itu, diperlukan penelitian tentang integrasi algoritma *machine learning* di masa mendatang.

Rekomendasi

Rekomendasi yang ditujukan pada peneliti yang akan menggarap penelitian sejenis adalah perlunya indikator yang dapat menjadi acuan untuk penilaian variabel, karena penilaian indikator dalam penelitian ini masih dilakukan secara subjektif. Topik yang dapat digunakan untuk penelitian lebih lanjut adalah mengenai integrasi algoritma *machine learning* yang kurang efektif dengan algoritma yang memiliki efektivitas tinggi. Tujuannya adalah untuk memperkuat efektivitas, akurasi, dan berbagai aspek lain dari deteksi *financial fraud* di lembaga keuangan. Salah satu contoh penelitian yang bisa dieksplorasi di masa mendatang adalah penggabungan atau integrasi algoritma *graph neural networks* (GNN) dengan *random forest*, serta kombinasi *federated learning* dengan *gradient boosting*. Untuk membuktikan hasil dari integrasi tersebut, maka perlu dilakukan penelitian secara empiris.