

Tinjauan Literatur Sistematis Pendekatan *Machine Learning* untuk Deteksi Kecurangan Laporan Keuangan

Dewa Ferrouzi Diaz Zhah Pahlevi¹, Isti Surjandari^{1*}

¹Department of Industrial Engineering, Faculty of Engineering, Universitas Indonesia, Depok 16424, Indonesia

Abstrak

Kecurangan laporan keuangan merupakan ancaman serius terhadap integritas dan transparansi sistem pelaporan keuangan perusahaan, khususnya bagi perusahaan publik. Untuk mengatasi tantangan ini, pendekatan berbasis machine learning telah banyak dikembangkan sebagai alat bantu dalam mengidentifikasi pola-pola kecurangan secara otomatis dan akurat. Penelitian ini menyajikan tinjauan literatur sistematis terhadap berbagai metode klasifikasi yang digunakan dalam deteksi kecurangan laporan keuangan, berdasarkan 43 publikasi akademik dari tahun 2011 hingga 2025. Metode klasifikasi dikelompokkan ke dalam empat taksonomi utama: metode statistik konvensional, algoritma supervised learning, metode ensemble learning, dan metode deep learning. Studi ini juga membahas berbagai sumber data yang digunakan, faktor-faktor yang memengaruhi hasil klasifikasi, serta kelebihan dan kekurangan masing-masing pendekatan. Hasil tinjauan menunjukkan bahwa meskipun metode statistik konvensional seperti regresi logistik masih sering digunakan karena kemudahan interpretasinya, metode berbasis machine learning seperti Random Forest, XGBoost, dan LSTM memberikan performa yang lebih unggul dalam hal akurasi. Namun demikian, tantangan seperti kebutuhan komputasi yang tinggi dan keterbatasan interpretabilitas tetap menjadi perhatian. Studi ini juga mengidentifikasi arah penelitian masa depan, termasuk penggabungan data terstruktur dan tidak terstruktur, penggunaan teknik text mining pada data naratif, serta pengembangan sistem deteksi adaptif. Dengan semakin berkembangnya teknologi data dan kecerdasan buatan, pendekatan berbasis machine learning memiliki potensi besar untuk meningkatkan efektivitas dan efisiensi dalam mendeteksi kecurangan laporan keuangan.

Kata kunci: Deteksi Kecurangan, Laporan Keuangan, Machine learning, Ensemble learning, Deep learning

Abstract

Financial statement fraud poses a serious threat to the integrity and transparency of corporate financial reporting systems, particularly for publicly listed companies. To address this challenge, machine learning-based approaches have been widely developed as tools to automatically and accurately identify fraudulent patterns. This study presents a systematic literature review of various classification methods used in financial statement fraud detection, based on 43 academic publications from 2011 to 2025. Classification methods are categorized into four main taxonomies: conventional statistical methods, supervised learning algorithms, ensemble learning methods, and deep learning approaches. The study also discusses the different data sources used, factors influencing classification performance, as well as the strengths and limitations of each approach. The findings indicate that while conventional statistical methods such as logistic regression are still commonly used due to their ease of interpretation, machine learning-based methods such as Random Forest, XGBoost, and LSTM demonstrate superior performance in terms of accuracy. Nevertheless, challenges such as high computational requirements and limited interpretability remain concerns. This study also identifies future research directions,

including the integration of structured and unstructured data, the application of text mining techniques to narrative data, and the development of adaptive detection systems. With the continued advancement of data technology and artificial intelligence, machine learning-based approaches hold great potential to enhance the effectiveness and efficiency of financial statement fraud detection.

Keywords: Fraud Detection, Financial Statement, Machine learning, Ensemble learning, Deep learning

PENDAHULUAN

Perkembangan pasar modal memiliki peran krusial dalam mendukung pertumbuhan ekonomi nasional melalui pengalokasian dana dari investor kepada perusahaan terbuka (Sholikah et al., 2022). Dalam beberapa tahun terakhir, pasar modal Indonesia menunjukkan pertumbuhan yang signifikan. Per November 2023, Bursa Efek Indonesia (BEI) mencatat lebih dari 900 perusahaan tercatat, dengan 79 pencatatan saham baru—angka tertinggi sepanjang sejarah pasar modal Indonesia. Partisipasi publik juga meningkat pesat, dengan jumlah investor pasar modal mencapai 12,16 juta, termasuk 5,25 juta investor saham (Indonesia Stock Exchange, 2023). Pertumbuhan ini merupakan hasil dari upaya literasi dan edukasi keuangan yang berkelanjutan, yang bertujuan untuk meningkatkan kepercayaan masyarakat terhadap investasi pada perusahaan yang berada di bawah pengawasan Otoritas Jasa Keuangan (OJK). Sesuai dengan Peraturan Bapepam-LK Nomor X.K.2 Tahun 2011, setiap perusahaan terbuka wajib menyampaikan laporan keuangan secara berkala yang telah diaudit (Otoritas Jasa Keuangan, 2011).

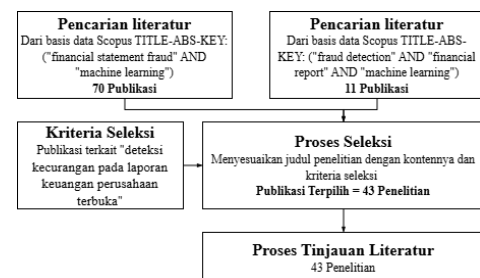
Meskipun terdapat pengawasan regulatif, kecurangan dalam laporan keuangan masih menjadi permasalahan yang sering terjadi. Survei yang dilakukan oleh Association of Certified Fraud Examiners (ACFE) pada tahun 2019 menunjukkan bahwa kecurangan laporan keuangan menyumbang 9,2% dari total kasus fraud di Indonesia, dengan total kerugian mencapai Rp242,26 miliar (ACFE, 2024). Laporan dari KPMG juga mengungkapkan bahwa 70% perusahaan responden pernah mengalami setidaknya satu bentuk fraud (Miller, 2022). Saat ini, pendeteksian fraud umumnya masih bergantung pada proses audit internal dan eksternal, yang secara kumulatif menyumbang sekitar 71,9% dari mekanisme deteksi fraud. Namun, Indonesia menghadapi keterbatasan jumlah profesional akuntansi, dengan rasio satu akuntan untuk setiap 121.000 penduduk—jauh tertinggal dibandingkan Malaysia (1:20.000) dan Singapura (1:5.000) (Herusetya, 2023). Selain itu, proses audit konvensional memerlukan waktu yang lama, dengan sebagian besar kasus kecurangan baru terdeteksi setelah 12 bulan (ACFE, 2024). Kondisi ini menunjukkan perlunya pendekatan deteksi yang lebih efisien dan adaptif.

Dalam beberapa tahun terakhir, metode machine learning (ML) mulai banyak diterapkan sebagai alternatif untuk mendeteksi kecurangan laporan keuangan karena kemampuannya dalam mengidentifikasi pola dan anomali secara otomatis pada data berskala besar. Sejumlah penelitian telah mengembangkan berbagai pendekatan, mulai dari metode statistik, ensemble methods, hingga deep learning (Ali et al., 2022). Namun, literatur yang ada masih tersebar dan bervariasi dalam hal data yang digunakan, teknik yang diterapkan, dan metrik evaluasi yang dipakai, sehingga sulit untuk disintesis secara komprehensif.

Makalah ini bertujuan untuk menyajikan tinjauan pustaka secara sistematis terhadap studi-studi terkini yang menggunakan metode machine learning dalam deteksi kecurangan laporan keuangan perusahaan. Tujuan utama dari studi ini adalah untuk mengklasifikasikan pendekatan yang digunakan, mengevaluasi efektivitasnya, serta mengidentifikasi kesenjangan dan arah penelitian di masa mendatang. Struktur makalah ini disusun sebagai berikut: Bagian 2 menjelaskan metodologi kajian literatur yang digunakan. Bagian 3 membahas klasifikasi metode machine learning berdasarkan studi yang direview. Bagian 4 merangkum temuan utama dan memberikan rekomendasi untuk penelitian selanjutnya.

METODE

Kajian literatur sistematis merupakan suatu metode dan proses penelitian yang bertujuan untuk mengidentifikasi serta menelaah secara kritis studi-studi terdahulu yang relevan, sekaligus mengumpulkan dan menganalisis informasi dari penelitian-penelitian tersebut (Kitchenham et al., 2008). Metodologi kajian literatur yang digunakan dalam studi ini ditunjukkan pada Gambar 1.



Gambar 1. Metodologi Tinjauan Literatur

Kajian literatur dilakukan untuk menjawab pertanyaan penelitian berikut ini:

PP1 Bagaimana perkembangan metode machine learning dalam mendeteksi kecurangan laporan keuangan?

PP2 Faktor-faktor apa saja yang memengaruhi efektivitas deteksi kecurangan berbasis machine learning?

PP3 Apa saja kelebihan dan kekurangan dari masing-masing metode machine learning yang digunakan?

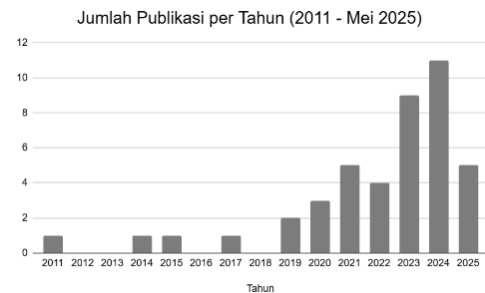
PP4 Masalah dan celah penelitian apa yang dapat menjadi arah pengembangan studi di masa mendatang?

Proses pengambilan literatur dilakukan melalui basis data Scopus, yang merupakan salah satu pangkalan data literatur ilmiah terbesar. Pencarian dilakukan dengan menggunakan kombinasi kata kunci pada bidang TITLE-ABS-KEY: ("financial statement fraud" AND "machine learning") dan ("fraud detection" AND "financial report" AND "machine learning"). Dari proses pencarian pertama, diperoleh sejumlah 70 publikasi, sedangkan pencarian kedua menghasilkan 11 publikasi. Setelah dilakukan pemeriksaan manual terhadap hasil penelusuran berdasarkan kata kunci, ditemukan bahwa dalam konteks laporan keuangan, penerapan teknik machine learning untuk deteksi kecurangan dapat dikategorikan ke dalam empat kelompok utama: (1) deteksi kecurangan pada laporan keuangan (financial statement fraud detection), (2) deteksi kecurangan pada transaksi (transaction fraud detection), (3) deteksi anomali dalam audit internal, dan (4) prediksi risiko kredit atau kebangkrutan yang berpotensi mengandung elemen fraud. Studi literatur ini secara khusus difokuskan pada kategori pertama, yaitu deteksi kecurangan pada laporan keuangan perusahaan terbuka. Sebanyak 43 publikasi akhirnya lolos penyaringan dan dikaji lebih lanjut dalam rentang waktu publikasi dari tahun 2011 hingga Mei 2025. Rentang waktu lebih dari satu dekade ini dianggap cukup representatif untuk menangkap perkembangan pendekatan dan tren metode yang digunakan dalam penelitian terkait

HASIL

Terdapat 43 publikasi yang memenuhi kriteria seleksi, terdiri atas 28 artikel jurnal dan 15 prosiding konferensi, yang diterbitkan dalam rentang waktu 2011 hingga Mei 2025. Ringkasan distribusi publikasi per tahun

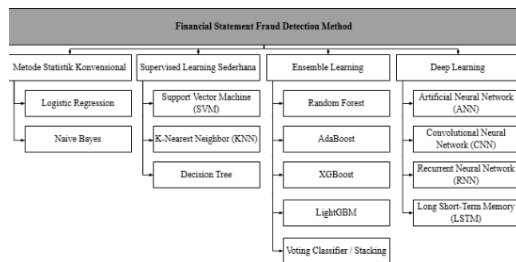
ditampilkan pada Gambar 2. Dalam tiga tahun terakhir, terdapat tren peningkatan jumlah publikasi terkait deteksi kecurangan laporan keuangan menggunakan metode machine learning. Negara yang paling banyak melakukan penelitian di bidang ini adalah Tiongkok dengan 21 publikasi, diikuti oleh Amerika Serikat dengan 10 publikasi, serta India dan Indonesia dengan masing-masing 6 publikasi.



Gambar 2. Distribusi Publikasi Mengenai Deteksi Kecurangan Laporan Keuangan dengan Machine Learning

Berdasarkan penelaahan terhadap pendekatan yang digunakan dalam studi sebelumnya, metode deteksi kecurangan laporan keuangan dapat diklasifikasikan ke dalam empat kategori utama, sebagaimana dirangkum dalam taksonomi pada Gambar 3: (1) metode statistik konvensional, seperti logistic regression dan discriminant analysis; (2) algoritma supervised learning, seperti decision tree, support vector machine, dan naive Bayes; (3) metode ensemble learning, seperti random forest, XGBoost, dan AdaBoost, serta; (4) metode deep learning, yang mencakup penggunaan artificial neural networks (ANN) dan recurrent neural networks (RNN).

Mayoritas studi terdahulu menggunakan pendekatan supervised learning karena kemampuannya dalam melakukan klasifikasi berdasarkan data historis yang telah dilabeli, memungkinkan deteksi pola-pola anomali yang mengindikasikan potensi kecurangan dalam laporan keuangan perusahaan publik. Pendekatan ensemble dan deep learning menunjukkan tren peningkatan pemanfaatan dalam studi-studi terkini, terutama karena keunggulannya dalam menangani data berukuran besar dan kompleksitas hubungan antar variabel.



Gambar 3. Taksonomi Metode Klasifikasi untuk Deteksi Kecurangan Laporan Keuangan

1. Metode Statistik Konvensional

Metode statistik konvensional merupakan pendekatan awal yang digunakan dalam mendeteksi kecurangan laporan keuangan. Pendekatan ini menekankan pada hubungan antara variabel-variabel keuangan untuk mengidentifikasi pola anomali yang mengindikasikan potensi fraud. Salah satu metode yang paling banyak digunakan adalah *logistic regression* (LR). Metode ini memodelkan probabilitas suatu entitas mengalami kecurangan berdasarkan serangkaian variabel prediktor. Studi-studi seperti yang dilakukan oleh (Wan, 2024) dan (Chen et al., 2024) menunjukkan bahwa *logistic regression* cukup efektif dalam mengidentifikasi perusahaan yang terindikasi melakukan manipulasi laporan keuangan dengan akurasi yang cukup tinggi.

Selain *logistic regression*, metode naive bayes juga digunakan untuk membedakan perusahaan yang melakukan fraud dan non-fraud. Naïve Bayes merupakan metode probabilistik yang mengasumsikan independensi antar fitur. Meskipun asumsi ini sering kali tidak terpenuhi dalam data keuangan, beberapa penelitian menunjukkan performa naïve Bayes cukup kompetitif, terutama pada dataset yang besar dan tidak terlalu kompleks (Gupta et al., 2021). Metode ini juga dikenal cepat dan efisien secara komputasi.

Dalam beberapa studi, metode statistik ini diperluas untuk mengakomodasi kompleksitas yang lebih tinggi. Misalnya, *stepwise logistic regression* digunakan oleh S. Chen et al. (2014) untuk secara iteratif memilih variabel prediktor paling signifikan, dengan mempertimbangkan kolinearitas dan redundansi antar variabel. Sementara itu, J. Chen et al. (2024) menggabungkan LR dengan teknik pengurangan dimensi seperti *principal component analysis*

(PCA) guna meningkatkan akurasi model dalam dataset berdimensi tinggi.

Meskipun metode statistik konvensional relatif mudah diinterpretasikan dan efisien dari sisi komputasi, pendekatan ini umumnya memiliki performa prediksi yang lebih rendah dibandingkan metode pembelajaran mesin modern, terutama dalam konteks dataset yang besar dan tidak terstruktur. Hal ini mendorong berkembangnya pendekatan berbasis machine learning yang lebih adaptif terhadap keragaman data dan non-linearitas dalam hubungan antar variabel.

2. Metode Algoritma *Supervised Learning*

Supervised learning merupakan pendekatan *machine learning* yang paling banyak diterapkan dalam studi deteksi kecurangan laporan keuangan. Pada pendekatan ini, model dilatih menggunakan dataset berlabel, yaitu laporan keuangan yang telah diklasifikasikan sebagai fraud atau non-fraud. Tujuannya adalah membangun fungsi pemetaan dari fitur-fitur keuangan ke dalam kelas prediksi.

Algoritma yang paling umum digunakan dalam kategori ini adalah *decision tree* (DT), *support vector machine* (SVM), dan *k-nearest neighbor* (KNN). Studi oleh S. Chen et al. (2014) menggunakan DT untuk mengidentifikasi pola-pola pengambilan keputusan yang membedakan laporan keuangan yang curang dan tidak curang. DT memiliki keunggulan dalam hal interpretabilitas, sehingga sangat cocok untuk konteks audit yang membutuhkan transparansi dalam proses pengambilan keputusan.

Sementara itu, SVM menawarkan keunggulan dalam menangani data berdimensi tinggi dan memisahkan kelas menggunakan hyperplane optimal. Penelitian Moepya et al. (2014) menunjukkan bahwa SVM mampu mencapai tingkat akurasi yang tinggi dalam mendeteksi fraud, terutama pada dataset yang tidak seimbang. Namun, kelemahan SVM adalah kurangnya interpretabilitas dan sensitivitas terhadap pemilihan parameter kernel.

KNN digunakan dalam beberapa studi karena kesederhanaan implementasi dan efektivitas dalam kondisi data yang tidak memiliki asumsi distribusi tertentu. Misalnya, Tarjo and Herawati (2017) menggunakan KNN untuk mengklasifikasikan laporan keuangan

berdasarkan kemiripan dengan observasi historis. Namun, performa KNN menurun secara signifikan ketika jumlah fitur meningkat, serta memerlukan optimasi jarak dan nilai K yang tepat.

Kelebihan utama algoritma supervised learning adalah kemampuannya dalam belajar dari data historis dan memberikan akurasi prediksi yang baik. Namun, keterbatasan muncul ketika menghadapi data yang sangat tidak seimbang, noise tinggi, atau fitur-fitur yang saling berkorelasi secara kompleks. Hal ini mendorong pengembangan metode ensemble dan deep learning yang lebih robust terhadap tantangan tersebut.

3. Metode *Ensemble Learning*

Metode *ensemble learning* merupakan pendekatan dalam pembelajaran mesin yang menggabungkan beberapa model pembelajaran (base learners) untuk meningkatkan akurasi prediksi dan mengurangi risiko kesalahan yang berasal dari satu model tunggal. Dalam konteks deteksi kecurangan laporan keuangan, metode ini semakin populer karena mampu menangani data yang tidak seimbang (*imbalanced data*), meningkatkan generalisasi model, dan mengurangi overfitting.

Metode ensemble yang paling sering digunakan dalam literatur adalah *Random Forest* (RF) dan *eXtreme Gradient Boosting* (XGBoost). *Random Forest*, sebagai metode berbasis *bagging*, bekerja dengan membangun banyak pohon keputusan pada subset acak dari data pelatihan dan menggabungkan hasil prediksi melalui mekanisme voting. Studi oleh Zhou et al. (2020) menunjukkan bahwa RF mampu mengidentifikasi kecurangan dengan akurasi yang tinggi dan stabil, serta memiliki ketahanan terhadap noise pada data.

Sementara itu, XGBoost merupakan metode *boosting* yang dikenal karena kecepatan dan efisiensinya dalam pelatihan model. Beberapa studi membandingkan kinerja XGBoost dengan model lain seperti *Logistic Regression* (LR), *Support Vector Machine* (SVM), dan Neural Network, dan menemukan bahwa XGBoost secara konsisten memberikan hasil terbaik dalam hal precision, recall, dan AUC (Area Under Curve) (Li & Zhou, 2024) (Zhu et al., 2024). Penggunaan teknik hyperparameter tuning seperti grid search atau Bayesian optimization juga umum dilakukan untuk meningkatkan performa model ini.

Selain XGBoost ada beberapa metode boosting yang kerap digunakan sebagai metode deteksi kecurangan laporan keuangan. Seperti penelitian Mishra (2021) yang memanfaatkan dan menunjukkan bahwa AdaBoost efektif dalam mendeteksi anomali pada laporan keuangan. AdaBoost bekerja dengan menambahkan model secara bertahap, di mana model baru berfokus pada kesalahan prediksi dari model sebelumnya. Setiap instance data yang salah diklasifikasikan akan diberi bobot lebih tinggi untuk iterasi berikutnya. Dalam konteks laporan keuangan, AdaBoost dapat memperkuat kemampuan klasifikasi terhadap sampel data fraud yang berupa minoritas. Namun, metode ini cenderung sensitif terhadap noise dan outlier.

LightGBM adalah algoritma boosting berbasis pohon yang dirancang untuk efisiensi tinggi dan kecepatan pelatihan. Berbeda dari metode gradient boosting tradisional, LightGBM menggunakan pendekatan leaf-wise growth yang memungkinkan model belajar lebih dalam dan akurat, namun berisiko overfitting jika tidak dikendalikan. Penelitian oleh (Zhao et al., 2024) menunjukkan bahwa LightGBM memiliki performa terbaik dalam deteksi kecurangan pada data keuangan China, dengan waktu komputasi yang jauh lebih rendah dibanding XGBoost.

Selain RF dan XGBoost, beberapa penelitian juga mengeksplorasi ensemble stacking, seperti kombinasi dari SVM dan XGBoost atau LR dan RF, untuk memanfaatkan keunggulan masing-masing algoritma dasar. Metode stacking, yaitu pendekatan yang menggabungkan prediksi dari beberapa model dengan meta-learner di atasnya, juga mulai digunakan untuk meningkatkan akurasi deteksi kecurangan yang kompleks dan tersembunyi (Achakzai & Peng, 2023).

Keunggulan utama metode ensemble adalah ketahanannya terhadap outlier dan kemampuannya menangani data berskala besar dan berdimensi tinggi. Namun, tantangan yang sering dihadapi meliputi kebutuhan komputasi yang tinggi dan kesulitan dalam interpretasi hasil model, khususnya pada pendekatan boosting yang kompleks.

4. Metode Deep Learning

Metode deep learning merupakan pendekatan pembelajaran mesin lanjutan yang menggunakan jaringan saraf tiruan (artificial

neural networks) dengan banyak lapisan tersembunyi untuk menangkap pola dan relasi kompleks dalam data. Dalam konteks deteksi kecurangan laporan keuangan, deep learning memberikan potensi signifikan karena kemampuannya dalam mengekstraksi fitur non-linier dari data keuangan yang bersifat kompleks dan tidak terstruktur.

Model deep learning yang umum digunakan dalam literatur antara lain adalah Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), dan Long Short-Term Memory (LSTM). ANN digunakan sebagai pendekatan dasar untuk klasifikasi biner (fraud atau tidak fraud), dan telah menunjukkan performa yang kompetitif dibandingkan metode machine learning konvensional seperti Logistic Regression dan SVM (Wu & Chen, 2024). CNN, meskipun lebih populer untuk pengolahan citra, mulai diaplikasikan untuk mendeteksi anomali dalam struktur laporan keuangan melalui representasi matriks atau gambar heatmap dari data numerik. LSTM, sebagai bagian dari Recurrent Neural Network (RNN), digunakan untuk menangkap pola temporal dari data keuangan historis dan memprediksi kemungkinan kecurangan berdasarkan urutan waktu data.

Salah satu tantangan utama dalam penerapan deep learning untuk deteksi kecurangan adalah jumlah data yang terbatas dan tidak seimbang (fraud vs. non-fraud). Untuk mengatasi hal ini, beberapa penelitian menggunakan teknik augmentasi data seperti SMOTE (Synthetic Minority Oversampling Technique) atau menerapkan pendekatan unsupervised learning untuk mengidentifikasi outlier dalam data keuangan (Z. Zhao & Bai, 2022). Selain itu, interpretabilitas menjadi isu

penting dalam model deep learning. Model ini sering dikritik sebagai black box karena sulit menjelaskan alasan di balik prediksi tertentu. Untuk mengatasi hal ini, beberapa studi menggunakan teknik interpretabilitas seperti LIME (Local Interpretable Model-Agnostic Explanations) atau SHAP (SHapley Additive exPlanations) untuk menjelaskan kontribusi fitur terhadap keputusan model.

Meskipun memiliki kompleksitas tinggi dan memerlukan sumber daya komputasi besar, deep learning menunjukkan tren yang terus meningkat dalam riset deteksi kecurangan karena potensinya dalam menangani volume data besar, kompleksitas relasi fitur, serta peningkatan performa klasifikasi dalam berbagai studi empiris (Psychoula et al., 2021) (Jan, 2021).

Setiap metode klasifikasi untuk mendeteksi kecurangan pada laporan keuangan perusahaan memiliki kelebihan dan kekurangan, yang dijelaskan melalui Tabel 1. Pemilihan metode klasifikasi dalam mendeteksi kecurangan laporan keuangan bergantung pada tujuan analisis, karakteristik data, serta kebutuhan interpretabilitas. Berdasarkan hasil dapat dilihat bahwa tidak ada satu metode yang secara mutlak paling unggul; pemilihan model harus mempertimbangkan kompleksitas data, kebutuhan akurasi, transparansi model, serta ketersediaan sumber daya komputasi. Penggunaan metode yang lebih kompleks seperti *ensemble* dan *deep learning* sangat potensial dalam meningkatkan akurasi deteksi, namun tantangan interpretasi dan kebutuhan data tetap harus diperhatikan dalam konteks implementasi praktis dan kepatuhan terhadap regulasi

Tabel 1. Kelebihan dan Kekurangan Metode Klasifikasi pada Deteksi Kecurangan Laporan Keuangan

Referensi	Metode Utama	Kelebihan	Kekurangan
Wan (2024), J. Chen et al. (2024) , S. Chen et al. (2014)	Logistic Regression	Mudah diimplementasikan, hasil mudah diinterpretasikan, cocok untuk baseline model.	Rentan terhadap multikolinearitas dan hanya mampu menangkap hubungan linier.
(Gupta et al., 2021)	Naive Bayes	Cepat, efisien pada data besar; bekerja baik untuk data teks	Asumsi independensi antar fitur sering tidak terpenuhi
S. Chen et al. (2014), Tang et al. (2018)	Decision Tree	Dapat menangani data numerik dan kategorik; interpretasi mudah.	Rentan terhadap overfitting dan bias pada data yang tidak seimbang.
S. Chen et al. (2014), Moepya et al. (2014)	Support Vector Machine	Efektif untuk data berdimensi tinggi dan dapat digunakan pada dataset kecil.	Sensitif terhadap pemilihan kernel; sulit diinterpretasikan.
Tarjo and Herawati (2017)	K-Nearest Neighbor	Algoritma sederhana, tidak membutuhkan asumsi distribusi data	Sensitif terhadap noise; tidak efektif pada dataset besar.
Li and Zhou (2024)	Random Forest	Stabil terhadap outlier dan overfitting; memiliki kemampuan estimasi feature importance.	Interpretasi kompleks; waktu pelatihan cukup lama untuk dataset besar.
Achakzai and Peng (2023), Mishra (2021), X. Chen and Zhai (2023)	AdaBoost	Performa tinggi pada dataset yang seimbang; meningkatkan model lemah	Sensitif terhadap outlier dan data noise
Li and Zhou (2024), Zhu et al. (2024), Achakzai and Peng (2023), X. Chen and Zhai (2023)	XGBoost	Cepat, akurat, mampu menangani missing value dan regulasi overfitting	Kompleksitas parameter tuning; konsumsi memori tinggi
R. Zhao et al. (2024), Achakzai and Peng (2023), X. Chen and Zhai (2023), Huang (2020)	LightGBM	Sangat cepat; efisien pada dataset besar; mendukung categorical feature langsung	Sulit diinterpretasi; kurang optimal pada dataset kecil
Zhu et al. (2024), Achakzai and Peng (2023)	Stacking	Menggabungkan keunggulan berbagai model; hasil prediksi lebih stabil	Waktu pelatihan lama; kompleksitas meningkat
Jan (2021), Rahayu and Widuri (2025)	Artificial Neural Network	Mampu menangkap pola nonlinier kompleks; cocok untuk prediksi jangka panjang.	Black box model; membutuhkan data dalam jumlah besar dan komputasi tinggi.
Wu and Chen (2024), Z. Zhao and Bai (2022)	Convolutional Neural Network	Cocok untuk data berbentuk matriks atau gambar laporan (misal hasil OCR); mampu ekstraksi fitur spasial.	Tidak optimal untuk data tabular atau sekuensial; kompleksitas tinggi dan interpretasi rendah.
Oguntimilehin et al. (2024)	Recurrent Neural Network	Baik untuk data sekuensial dan temporal, seperti tren waktu laporan keuangan.	Rawan vanishing gradient; pelatihan lambat; sulit menangani dependensi jangka panjang.
Z. Zhao and Bai (2022), Bhattacharya et al. (2024)	Long Short-Term Memory	Mengatasi kelemahan RNN; mampu mengingat dependensi jangka panjang; cocok untuk tren historis.	Butuh banyak data dan daya komputasi; model kompleks dan interpretasi rendah.

SIMPULAN DAN SARAN

Kecurangan laporan keuangan merupakan salah satu isu krusial dalam dunia bisnis dan keuangan yang berdampak besar terhadap kepercayaan investor dan stabilitas pasar. Untuk mendeteksi kecurangan secara lebih efektif, pendekatan berbasis machine learning telah banyak dikembangkan dan digunakan dalam penelitian akademik. Studi ini menyajikan tinjauan literatur sistematis terkait metode klasifikasi yang digunakan untuk deteksi kecurangan laporan keuangan, serta mengidentifikasi arah potensial penelitian di masa depan.

Dalam tinjauan literatur ini, sebanyak 43 publikasi yang relevan telah dianalisis, mencakup publikasi dari tahun 2011 hingga Mei 2025. Metode klasifikasi dalam penelitian-penelitian tersebut dapat dikategorikan ke dalam empat taksonomi utama: metode statistik konvensional, algoritma supervised learning, metode ensemble learning, dan metode deep learning. Selain itu, kajian ini juga membahas berbagai sumber data yang digunakan, faktor-faktor yang memengaruhi hasil klasifikasi, serta kelebihan dan kekurangan masing-masing pendekatan.

Metode statistik konvensional seperti regresi logistik masih banyak digunakan karena interpretabilitasnya, namun mulai tergeser oleh metode yang lebih kompleks seperti random forest, XGBoost, dan neural network yang mampu memberikan akurasi lebih tinggi. Meskipun begitu, peningkatan akurasi sering kali dibarengi dengan meningkatnya kompleksitas dan keterbatasan interpretabilitas. Oleh karena itu, isu interpretabilitas dan transparansi model menjadi perhatian penting, terutama dalam konteks audit dan pengambilan keputusan regulasi.

Sebagai arahan penelitian masa depan, dibutuhkan eksplorasi lebih lanjut terhadap pendekatan yang menggabungkan data terstruktur dan tidak terstruktur, serta pengembangan model adaptif yang mampu menyesuaikan diri dengan dinamika perubahan pola kecurangan. Selain itu, integrasi model klasifikasi ke dalam sistem pendukung keputusan yang interaktif dan ramah pengguna dapat meningkatkan kebermanfaatan model dalam praktik.

Dengan semakin berkembangnya teknik analisis data dan ketersediaan data keuangan, penelitian mengenai deteksi kecurangan laporan keuangan berbasis machine learning memiliki potensi besar untuk terus berkembang dan berkontribusi dalam menciptakan sistem pelaporan keuangan yang lebih transparan dan akuntabel..

UCAPAN TERIMA KASIH

Penulis menyampaikan apresiasi dan rasa terima kasih kepada Fakultas Teknik Universitas Indonesia atas pendanaan penelitian ini melalui Hibah Riset Profesor FTUI Skema Seed Funding dengan Nomor: NKB-3452/UN2.F4.D/PPM.00.00/2024.

DAFTAR PUSTAKA

- ACFE Report to the Nations | 2024 Global Fraud Study.* (n.d.). <https://legacy.acfe.com/report-to-the-nations/2024/>
- Achakzai, M. a. K., & Peng, J. (2023). Detecting financial statement fraud using dynamic ensemble machine learning. *International Review of Financial Analysis*, 89, 102827. <https://doi.org/10.1016/j.irfa.2023.102827>
- Ali, A., Razak, S. A., Othman, S. H., Eisa, T. a. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022). Financial Fraud Detection Based on Machine Learning: A Systematic Literature review. *Applied Sciences*, 12(19), 9637. <https://doi.org/10.3390/app12199637>
- Bhattacharya, R., Kumar, R., Rajeswari, U., Prakash, J. A., Barodia, N., & Sawadkar, S. (2024). An analysis on Financial Statement Fraud Detection for Listed Companies using DCNN-LSTM-AE-AM Model. *2024 Asian Conference on Intelligent Technologies (ACOIT)*, 1–6. <https://doi.org/10.1109/acoit62457.2024.10941438>
- Chen, J., Qian, H., & Yao, W. (2024). Efficient credit card fraud detection based on binary logistic regression. *Applied and Computational Engineering*, 115(1),

- 103–108. <https://doi.org/10.54254/2755-2721/2025.18483>
- Chen, S., Goo, Y. J., & Shen, Z. (2014). A hybrid approach of stepwise regression, logistic regression, support vector machine, and Decision Tree for forecasting fraudulent financial statements. *The Scientific World JOURNAL*, 2014, 1–9. <https://doi.org/10.1155/2014/968712>
- Chen, X., & Zhai, C. (2023). Bagging or boosting? Empirical evidence from financial statement fraud detection. *Accounting and Finance*, 63(5), 5093–5142. <https://doi.org/10.1111/acfi.13159>
- Gupta, A., Lohani, M. C., & Manchanda, M. (2021). Financial fraud detection using naive bayes algorithm in highly imbalance data set. *Journal of Discrete Mathematical Sciences and Cryptography*, 24(5), 1559–1572. <https://doi.org/10.1080/09720529.2021.1969733>
- Huang, K. (2020). An optimized LightGBM model for fraud detection. *Journal of Physics Conference Series*, 1651(1), 012111. <https://doi.org/10.1088/1742-6596/1651/1/012111>
- Indonesia Stock Exchange. (2023, December). *Melalui Berbagai Pencapaian Tahun 2023, Pasar Modal Indonesia Tunjukkan Optimisme Hadapi Tahun 2024* [Press release]. Retrieved May 13, 2025, from <https://www.idx.co.id/en/news/press-release/2080>
- Jan, C. (2021). Detection of Financial Statement Fraud Using Deep Learning for Sustainable Development of Capital Markets under Information Asymmetry. *Sustainability*, 13(17), 9879. <https://doi.org/10.3390/su13179879>
- Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2008). Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology*, 51(1), 7–15. <https://doi.org/10.1016/j.infsof.2008.09.009>
- Li, T., & Zhou, J. (2024). Evaluating the effectiveness of machine learning algorithms for financial fraud detection. In *Frontiers in artificial intelligence and applications*. <https://doi.org/10.3233/faia241178>
- Miller, M. (2022, January). *Insights*. KPMG. <https://kpmg.com/xx/en/home/insights/2022/01/kpmg-fraud-outlook-survey.html>
- Mishra, A. (2021). Fraud Detection: A study of AdaBoost Classifier and K-Means Clustering. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3789879>
- Moepya, S. O., Nelwamondo, F. V., & Van Der Walt, C. (2014). A support vector machine approach to detect Financial statement Fraud in South Africa: A first look. In *Lecture notes in computer science* (pp. 42–51). https://doi.org/10.1007/978-3-319-05458-2_5
- Oguntimilehin, A., Ngaikedi, F., Atachin, A. J., Toyin, O., Ogundipe, A. T., Mebawondu, J. O., Babalola, G. O., Oluwatoki, T. G., & Sanya, O. A. (2024). Financial Fraud Detection Model using Recurrent Neural Network. *2024 IEEE 5th International Conference on Electro-Computing Technologies for Humanity (NIGERCON)*, 1–5. <https://doi.org/10.1109/nigercon62786.2024.10927232>
- Otoritas Jasa Keuangan. (2011, July 5). *X. K2. Penyampaian Laporan Keuangan Berkala Emiten atau Perusahaan Publik Berdasarkan Keputusan Ketua Badan Pengawas Pasar Modal dan Lembaga Keuangan*. <https://ojk.go.id/id/regulasi/Pages/BAPEPAM-XK2-tentang-Penyampaian-Laporan-Kuangan-Berkala-Emiten-atau-Perusahaan-Publik.aspx>
- Psychoula, I., Gutmann, A., Mainali, P., Lee, S. H., Dunphy, P., & Petitcolas, F. a. P. (2021, May 13). *Explainable machine learning for fraud detection*. arXiv.org. <https://arxiv.org/abs/2105.06314>
- Rahayu, D. V. R. W., & Widuri, R. (2025). Artificial neural network methodology in financial statements fraud: An empirical

- study in the property and real estate sector. *Risk Governance and Control Financial Markets & Institutions*, 15(1, special issue), 237–248. <https://doi.org/10.22495/rgcv15i1sip9>
- Sholikah, N. F. P., Putri, W., & Djangi, N. R. M. (2022). Peranan pasar modal dalam perekonomian negara Indonesia. *ARBITRASE Journal of Economics and Accounting*, 3(2), 341–345. <https://doi.org/10.47065/arbitrase.v3i2.496>
- Tang, X., Liu, G., Yang, J., & Wei, W. (2018). Knowledge-based Financial Statement Fraud Detection System: based on an ontology and a decision tree. *KNOWLEDGE ORGANIZATION*, 45(3), 205–219. <https://doi.org/10.5771/0943-7444-2018-3-205>
- Tarjo, T., & Herawati, N. (2017). The comparison of two data mining method to detect financial fraud in Indonesia. *GATR Accounting and Finance Review*, 2(1), 01–08. [https://doi.org/10.35609/afr.2017.2.1\(1](https://doi.org/10.35609/afr.2017.2.1(1)
- Tristiawati, P. (2023, November 4). Guru Besar Akuntansi: Indonesia kekurangan Akuntan publik. *liputan6.com*. <https://www.liputan6.com/regional/read/5443669/guru-besar-akuntansi-indonesia-kekurangan-akuntan-publik#:~:text=%E2%80%9CBerdasarkan%20data%20awal%20tahun%202023,dikukuhkan%20sebagai%20Guru%20Besar%20dalam>
- Wan, Y. (2024). Financial Fraud Detection Study - based on Logit Model. *Financial Economics Insights.*, 1(1), 40–49. <https://doi.org/10.70088/r74m3975>
- Wu, P., & Chen, Y. (2024). Enhanced detection of accounting fraud using a CNN-LSTM-Attention model optimized by Sparrow search. *PeerJ Computer Science*, 10, e2532. <https://doi.org/10.7717/peerj-cs.2532>
- Zhao, R., Zhang, L., & Li, Z. (2024). Identification of Financial Fraud in Listed Companies Based on Bayesian-LightGBM Model. *Proceedings of the 2024 International Conference on Artificial Intelligence and Computer Science*, 339–344. <https://doi.org/10.1145/3714334.3714391>
- Zhao, Z., & Bai, T. (2022). Financial fraud detection and prediction in listed companies using SMOTE and machine learning algorithms. *Entropy*, 24(8), 1157. <https://doi.org/10.3390/e24081157>
- Zhu, S., Wu, H., Ngai, E. W. T., Ren, J., He, D., Ma, T., & Li, Y. (2024). A financial fraud prediction framework based on stacking ensemble learning. *Systems*, 12(12), 588. <https://doi.org/10.3390/systems12120588>