BAB 2
LANDASAN TEORI

2.1 Penelitian Terdahulu

Pengembangan model *Hybrid* semakin banyak digunakan dalam sistem *fraud detection* dalam beberapa tahun terakhir, khususnya untuk mendeteksi anomali pada transaksi keuangan secara lebih adaptif. Pendekatan ini dikembangkan untuk mengatasi keterbatasan metode tradisional berbasis aturan (*rule-based*) yang cenderung kurang fleksibel dalam menghadapi pola *fraud* yang terus berkembang. Dengan memanfaatkan teknik *machine learning* dan *unsupervised learning*, model dapat mengidentifikasi pola transaksi yang tidak biasa secara lebih akurat serta mampu diterapkan pada berbagai kondisi penggunaan.

Tabel 2.1 merangkum sepuluh penelitian terdahulu yang menjadi dasar konseptual dalam penelitian ini. Penelitian-penelitian tersebut membahas berbagai pendekatan dalam deteksi anomali dan *fraud*, termasuk penggunaan metode seperti *Isolation Forest*, *Autoencoder*, serta teknik *Explainable Artificial Intelligence (XAI)* seperti *lime* dan *shap*. Hasil dari penelitian-penelitian tersebut menunjukkan bahwa pendekatan berbasis *machine learning* dan *hybrid model* mampu meningkatkan performa deteksi *fraud* dengan tingkat akurasi yang dalam beberapa kasus dapat mencapai lebih dari 90%.

Tabel 2.1. Penelitian Terkait

No	Nama Jurnal	Judul Artikel	Penulis	Latar Belakang, Metode, dan Hasil
1	*Information Sciences (2021)*	*Combining Unsupervised and Supervised Learning for Fraud Detection in Financial Transactions[16]*	Carcillo, F.; Dal Pozzolo, A.; Le Borgne, Y.; Caelen, O.; Bontempi, G.	Penelitian ini membahas deteksi fraud pada transaksi finansial digital dengan pendekatan *hybrid machine learning*. Metode yang digunakan menggabungkan *unsupervised anomaly detection* dan *supervised learning*. Hasil penelitian menunjukkan peningkatan performa deteksi fraud dengan nilai *AUC* hingga 0.97.

Tabel 2.1. Penelitian Terkait (Lanjutan)

No	Nama Jurnal	Judul Artikel	Penulis	Latar Belakang, Metode, dan Hasil
2	*ACM Computing Surveys (2021)*	*Deep Learning for Anomaly Detection: A Review[17]*	Pang, G.; Shen, C.; Cao, L.; Hengel, A.	Penelitian ini mengkaji berbagai metode *deep learning* untuk *anomaly detection* pada data kompleks. Metode yang dibahas termasuk *Autoencoder* dan *Deep Neural Network*. Hasil penelitian menunjukkan metode ini mampu meningkatkan akurasi deteksi anomali hingga lebih dari 90%.
3	*IEEE Transactions on Knowledge and Data Engineering (2021)*	*Extended Isolation Forest[18]*	Hariri, S.; Kind, M.; Brunner, R.	Penelitian ini mengembangkan metode *Isolation Forest* untuk meningkatkan kemampuan deteksi anomali pada dataset besar. Hasil penelitian menunjukkan metode ini mampu mendeteksi anomali secara lebih efisien pada data dengan dimensi tinggi.
4	*Neurocomputing (2020)*	*Deep Autoencoder for Unsupervised Anomaly Detection in Financial Data[19]*	Zantedeschi, V.; Ene, F.; Baroni, M.	Penelitian ini mengembangkan model *Deep Autoencoder* untuk mendeteksi anomali pada data transaksi finansial. Hasil penelitian menunjukkan model mampu mengidentifikasi pola anomali dengan tingkat akurasi lebih dari 94%.
5	*Expert Systems with Applications (2021)*	*Machine Learning Approaches for Financial Anomaly Detection[20]*	Bahnsen, A.; Aouada, D.; Ottersten, B.	Penelitian ini membahas penerapan berbagai algoritma *machine learning* untuk mendeteksi transaksi finansial yang tidak normal. Hasil penelitian menunjukkan bahwa metode berbasis pembelajaran mesin mampu meningkatkan akurasi deteksi fraud secara signifikan dibanding metode tradisional.

Tabel 2.1. Penelitian Terkait (Lanjutan)

No	Nama Jurnal	Judul Artikel	Penulis	Latar Belakang, Metode, dan Hasil
6	*ACM SIGKDD (2016)*	*Why Should I Trust You? Explaining the Predictions of Any Classifier[21]*	Ribeiro, M.; Singh, S.; Guestrin, C.	Penelitian ini mengembangkan metode *Explainable Artificial Intelligence* yang disebut *LIME*. Metode ini memungkinkan interpretasi lokal terhadap prediksi model *machine learning* sehingga meningkatkan transparansi model.
7	*NeurIPS (2017)*	*A Unified Approach to Interpreting Model Predictions[15]*	Lundberg, S.; Lee, S.	Penelitian ini memperkenalkan metode *SHAP* untuk menjelaskan kontribusi setiap fitur terhadap hasil prediksi model berbasis teori *Shapley Value*. Metode ini memberikan interpretasi yang konsisten terhadap model kompleks.
8	*Information Fusion (2020)*	*Explainable Artificial Intelligence (XAI): Concepts and Challenges[12]*	Arrieta, A.; Diaz-Rodriguez, N.	Penelitian ini membahas konsep dan tantangan dalam penerapan *Explainable Artificial Intelligence*. Hasil penelitian menunjukkan pentingnya interpretasi model dalam sistem yang bersifat kritis seperti sistem keuangan dan deteksi fraud.
9	*IEEE Access (2022)*	*Anomaly Detection in Digital Payment Systems Using Machine Learning[22]*	Zhang, Y.; Wang, X.; Li, H.	Penelitian ini mengembangkan sistem deteksi anomali pada transaksi pembayaran digital menggunakan pendekatan *machine learning*. Hasil penelitian menunjukkan metode ini mampu meningkatkan akurasi deteksi anomali hingga 93%.

Tabel 2.1. Penelitian Terkait (Lanjutan)

No	Nama Jurnal	Judul Artikel	Penulis	Latar Belakang, Metode, dan Hasil
10	*Journal of Financial Crime (2022)*	*Hybrid Machine Learning Approach for Financial Transaction Fraud Detection[23]*	Ahmad, M.; Mahmood, A.; Hu, J.	Penelitian ini mengembangkan sistem deteksi fraud berbasis *hybrid machine learning*. Metode yang digunakan adalah kombinasi *Isolation Forest* dan *Autoencoder*. Hasil penelitian menunjukkan pendekatan hybrid mampu meningkatkan akurasi deteksi fraud hingga 96%.

Berdasarkan sepuluh penelitian terdahulu yang telah dirangkum pada Tabel 2.1, dapat disimpulkan bahwa penggunaan metode *machine learning* dan *deep learning* memiliki peran penting dalam meningkatkan kemampuan sistem dalam mendeteksi anomali dan *fraud* pada data transaksi keuangan[19]. Berbagai pendekatan seperti *Isolation Forest*, *Autoencoder*,serta metode klasifikasi berbasis *Random Forest* dan *Support Vector Machine* telah terbukti mampu mengidentifikasi pola transaksi yang tidak biasa dengan tingkat akurasi yang relatif tinggi.

Selain itu, beberapa penelitian juga menunjukkan bahwa pendekatan berbasis *unsupervised learning* memiliki keunggulan dalam mendeteksi anomali pada dataset yang tidak memiliki label yang lengkap atau memiliki distribusi data yang tidak seimbang. Metode seperti *Isolation Forest* mampu mendeteksi anomali secara efisien pada dataset berskala besar, sedangkan *Autoencoder* dapat mempelajari pola normal dari data melalui proses rekonstruksi untuk mengidentifikasi transaksi yang menyimpang dari pola tersebut[19].

Di sisi lain, perkembangan *Explainable Artificial Intelligence (XAI)* seperti *lime* dan *shap* juga memberikan kontribusi penting dalam meningkatkan transparansi model *machine learning*. Dengan adanya metode interpretasi ini, keputusan yang dihasilkan oleh model dapat dijelaskan secara lebih jelas sehingga meningkatkan tingkat kepercayaan terhadap sistem deteksi *fraud*.

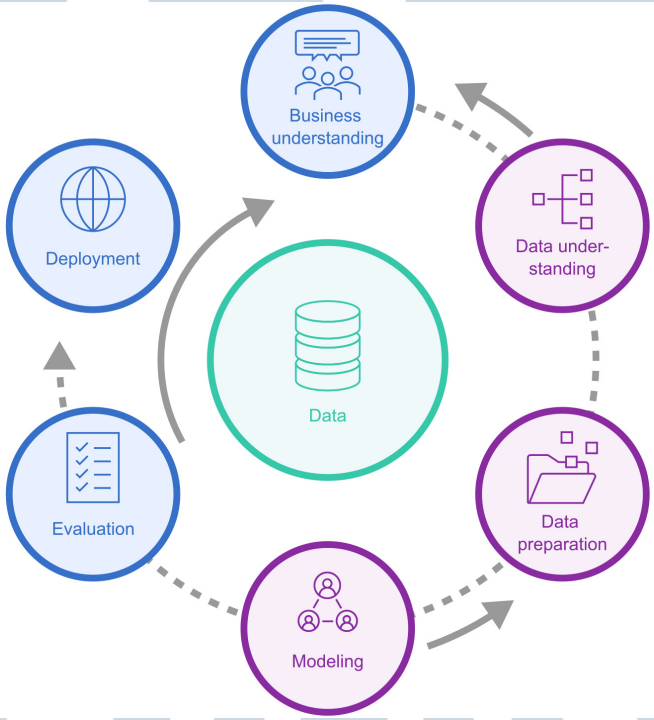
Secara keseluruhan, penelitian-penelitian terdahulu menunjukkan bahwa kombinasi beberapa metode *machine learning* dalam pendekatan *hybrid* berpotensi menghasilkan sistem deteksi anomali yang lebih akurat dan adaptif dibandingkan dengan metode konvensional berbasis aturan (*rule-based*). Oleh karena itu, penelitian ini mengusulkan penggunaan pendekatan *Hybrid Machine Learning* yang menggabungkan metode *Isolation Forest* dan *Autoencoder* serta didukung oleh teknik *Explainable Artificial Intelligence* untuk meningkatkan kemampuan sistem dalam mendeteksi transaksi anomali secara lebih efektif[12].

2.2 *Cross Industry Standard Process for Data Mining* (CRISP-DM)

Penelitian ini menggunakan *Cross Industry Standard Process for Data Mining* (CRISP-DM) yang merupakan sebuah kerangka kerja yang banyak digunakan dalam proses pengembangan sistem *data mining* dan *machine learning*. Metode ini pertama kali diperkenalkan pada tahun 1996

melalui kolaborasi antara beberapa perusahaan teknologi dan institusi penelitian seperti Daimler-Benz, SPSS, dan NCR. CRISP-DM dirancang untuk menyediakan pendekatan yang sistematis dan terstruktur dalam proses analisis data sehingga dapat diterapkan pada berbagai bidang industri[24].

CRISP-DM menyediakan alur kerja yang jelas dalam proses pengembangan model analisis data mulai dari tahap pemahaman permasalahan hingga evaluasi hasil model yang dihasilkan. Pendekatan ini bersifat *iterative*, yang berarti setiap tahapan dapat diulang kembali apabila ditemukan kebutuhan untuk memperbaiki proses sebelumnya. Kerangka kerja CRISP-DM terdiri dari enam tahapan utama yang dapat dilihat pada Gambar 2.1 yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Keenam tahapan tersebut saling terhubung dan membentuk suatu siklus proses pengembangan sistem berbasis data[25].



Gambar 2.1. Visualisasi CRISP-DM

Sumber: [26]

1. ***Business Understanding***

   Tahapan pertama dalam CRISP-DM adalah memahami permasalahan bisnis atau tujuan penelitian yang ingin dicapai. Pada tahap ini dilakukan identifikasi permasalahan, penentuan tujuan analisis, serta penentuan kriteria keberhasilan dari model yang akan dibangun.

2. ***Data Understanding***

   Tahapan ini bertujuan untuk memahami karakteristik data yang akan digunakan dalam penelitian. Proses yang dilakukan meliputi pengumpulan data, eksplorasi awal terhadap *dataset*, serta analisis struktur dan distribusi data yang tersedia.

3. ***Data Preparation***

Tahap *data preparation* merupakan proses pengolahan data sebelum digunakan dalam proses pemodelan. Pada tahap ini dilakukan beberapa proses seperti pembersihan data (*data cleaning*), transformasi data, normalisasi data, serta pemilihan fitur yang relevan untuk digunakan dalam proses pelatihan model.

4. ***Modeling***

   Pada tahap *modeling* dilakukan proses pembangunan model menggunakan algoritma *machine learning* atau metode analisis data yang sesuai dengan tujuan penelitian. Model yang dibangun kemudian dilatih menggunakan *training data* agar mampu mempelajari pola yang terdapat dalam *dataset*.

5. ***Evaluation***

   Tahap *evaluation* bertujuan untuk mengevaluasi performa model yang telah dibangun. Evaluasi dilakukan dengan menggunakan berbagai metrik pengukuran untuk memastikan bahwa model yang dihasilkan mampu menyelesaikan permasalahan yang telah didefinisikan pada tahap awal.

6. ***Deployment***

   Tahap *deployment* merupakan tahap akhir dalam CRISP-DM yang bertujuan untuk menerapkan model yang telah dibangun ke dalam sistem nyata atau lingkungan operasional sehingga dapat digunakan dalam proses pengambilan keputusan.

Dengan menggunakan kerangka kerja *CRISP-DM*, proses pengembangan model *machine learning* dapat dilakukan secara sistematis dan terstruktur sehingga memudahkan peneliti dalam mengelola proses analisis data dari tahap awal hingga tahap evaluasi model.

2.3 Pembayaran Berbasis QR

Pembayaran berbasis QR merupakan salah satu metode pembayaran digital yang menggunakan kode dua dimensi yang dikenal sebagai *Quick Response Code* sebagai media untuk menyimpan informasi transaksi. Kode QR dapat menyimpan berbagai jenis informasi seperti identitas *merchant*, nomor akun tujuan pembayaran, maupun detail transaksi yang diperlukan untuk memproses pembayaran[27].

Dalam proses pembayaran berbasis QR, pengguna cukup memindai kode QR menggunakan kamera pada perangkat *smartphone*. Setelah kode dipindai, aplikasi pembayaran digital akan menampilkan informasi transaksi yang kemudian dapat dikonfirmasi oleh pengguna untuk menyelesaikan proses pembayaran. Metode ini dinilai lebih praktis dibandingkan metode pembayaran konvensional karena tidak memerlukan kartu fisik maupun perangkat tambahan[28].

Di Indonesia, implementasi pembayaran berbasis QR telah distandarisasi melalui sistem *Quick Response Code Indonesian Standard (QRIS)* yang dikembangkan oleh *Bank Indonesia*. Standar ini memungkinkan berbagai aplikasi pembayaran digital menggunakan satu kode QR yang sama sehingga meningkatkan interoperabilitas antara berbagai penyedia layanan pembayaran.

2.4 Fraud Detection System

Fraud Detection System merupakan sistem yang dirancang untuk mengidentifikasi aktivitas transaksi yang berpotensi mengarah pada tindakan *fraud*. Sistem ini bekerja dengan cara menganalisis data transaksi untuk menemukan pola atau karakteristik yang menunjukkan adanya aktivitas yang tidak normal[29].

Dalam implementasinya, *Fraud Detection System* biasanya terdiri dari beberapa komponen utama, yaitu proses pengumpulan data transaksi, analisis pola transaksi, serta proses pengambilan keputusan berdasarkan hasil analisis tersebut. Sistem ini dapat menggunakan berbagai pendekatan analisis seperti metode statistik, aturan berbasis logika (*rule-based system*), maupun metode *machine learning*. Pendekatan *machine learning* banyak digunakan dalam sistem deteksi *fraud* karena mampu mempelajari pola transaksi dari data historis dan mengidentifikasi aktivitas transaksi yang memiliki karakteristik berbeda dari pola tersebut[30].

2.5 Machine Learning

Machine learning merupakan cabang dari *artificial intelligence* yang berfokus pada pengembangan algoritma yang mampu mempelajari pola dari data dan membuat prediksi berdasarkan pola tersebut. Dalam pendekatan ini, sistem komputer tidak diprogram secara eksplisit untuk menyelesaikan suatu tugas, melainkan belajar dari data yang tersedia melalui proses pelatihan (*training*).

Metode *machine learning* dapat digunakan untuk berbagai jenis permasalahan seperti klasifikasi (*classification*), regresi (*regression*), pengelompokan data (*clustering*), serta deteksi anomali (*anomaly detection*)[31]. Dengan memanfaatkan *dataset* yang cukup besar, model *machine learning* dapat menemukan pola tersembunyi dalam data yang sulit diidentifikasi menggunakan metode analisis konvensional.Dalam penelitian ini, metode *machine learning* digunakan untuk menganalisis pola transaksi pengguna dalam sistem pembayaran digital dan mengidentifikasi transaksi yang memiliki karakteristik berbeda dari pola transaksi normal[32].

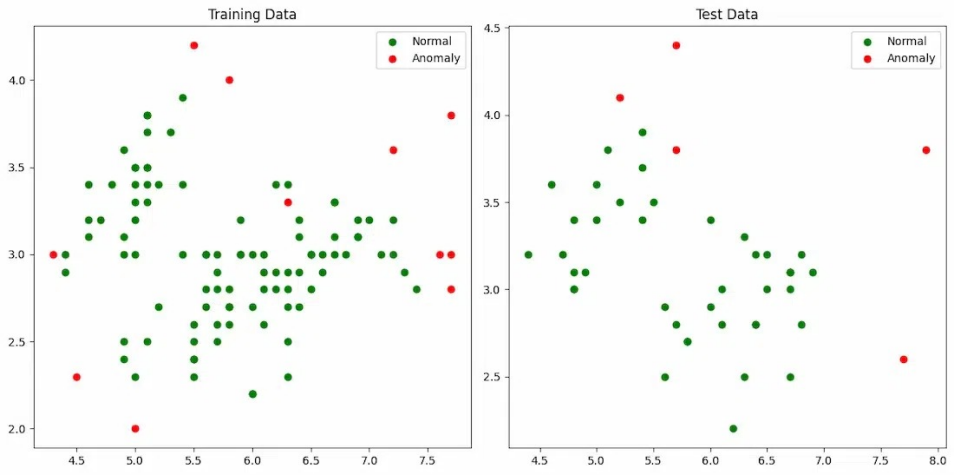
2.6 Anomaly Detection

Anomaly detection adalah teknik analisis data yang digunakan untuk mengidentifikasi data yang memiliki karakteristik berbeda dari sebagian besar data lainnya dalam suatu *dataset*. Data yang memiliki karakteristik tersebut biasanya disebut sebagai anomali atau *outlier*.

Dalam konteks sistem transaksi digital, *anomaly detection* digunakan untuk mendeteksi transaksi yang memiliki pola tidak normal dibandingkan dengan transaksi sebelumnya. Sebagai contoh, transaksi dengan nominal yang sangat besar dibandingkan transaksi sebelumnya atau transaksi yang terjadi dalam frekuensi yang tidak wajar dapat dikategorikan sebagai anomali. Pendekatan *anomaly detection* sering digunakan dalam berbagai bidang seperti keamanan jaringan, deteksi *fraud*, serta *monitoring* sistem. Dalam penelitian ini, teknik *anomaly detection* digunakan untuk mengidentifikasi transaksi yang memiliki pola berbeda dari perilaku transaksi normal pengguna.

2.7 Isolation Forest

Isolation Forest merupakan metode *unsupervised anomaly detection* yang diperkenalkan oleh Liu et al. untuk mendeteksi data anomali dengan cara mengisolasi observasi dalam struktur pohon acak (*random tree*). Berbeda dengan metode deteksi anomali lainnya yang mengukur jarak atau kepadatan data, Isolation Forest bekerja dengan memisahkan data secara rekursif menggunakan pemilihan atribut dan nilai split secara acak[10]. seperti yang ada dari Gambar 2.2 ,Konsep utama dari metode ini adalah bahwa data anomali cenderung lebih mudah diisolasi dibandingkan data normal. Hal ini menyebabkan panjang jalur (*path length*) pada pohon isolasi untuk data anomali biasanya lebih pendek dibandingkan data normal.



Gambar 2.2. Ilustrasi Isolation Forest

Sumber: [33]

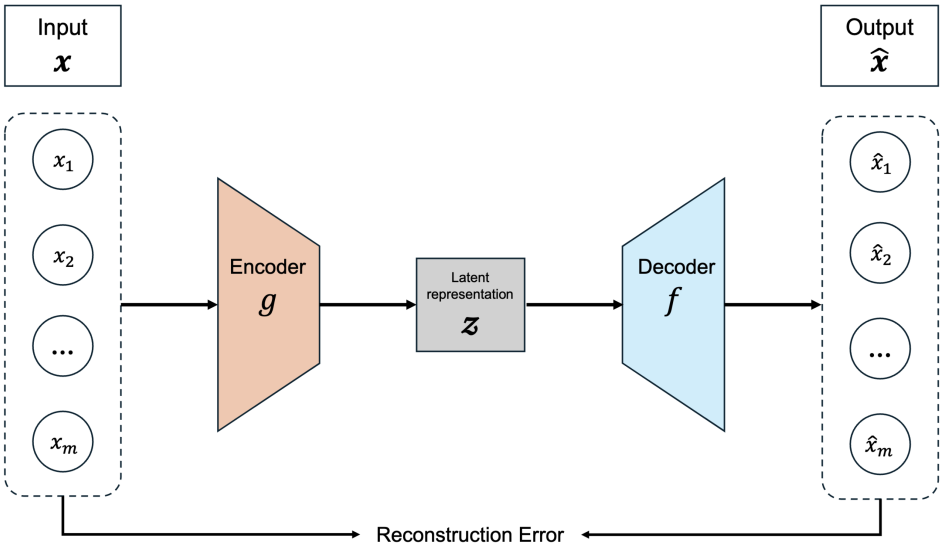
Nilai skor anomali berada pada rentang 0 hingga 1. Jika nilai $s(x,n)$ mendekati 1, maka data tersebut memiliki kemungkinan tinggi sebagai anomali karena lebih cepat terisolasi dalam struktur pohon [34]. Kondisi tersebut menunjukkan bahwa karakteristik data cukup berbeda dibanding mayoritas data lainnya.

Sebaliknya, jika nilai skor mendekati 0.5 atau lebih kecil, maka data tersebut cenderung merupakan data normal. Hal ini menandakan bahwa sampel memerlukan jalur pemisahan yang lebih panjang sehingga masih berada dalam pola distribusi umum dataset. Dengan demikian, semakin tinggi skor anomali, semakin besar potensi data diklasifikasikan sebagai penyimpangan [34].

2.8 Autoencoder

Autoencoder merupakan model *neural network* yang digunakan untuk mempelajari representasi data melalui proses rekonstruksi data. Model ini terdiri dari dua komponen utama yaitu *encoder* dan *decoder*. Bagian *encoder* berfungsi untuk mengubah data input menjadi representasi yang lebih ringkas, sedangkan bagian *decoder* digunakan untuk merekonstruksi kembali data tersebut menjadi bentuk yang mendekati data input[19].

Dalam proses pelatihan, model *autoencoder* dilatih menggunakan data normal sehingga model mampu mempelajari pola data tersebut. Ketika model menerima data yang memiliki karakteristik berbeda dari pola yang telah dipelajari, proses rekonstruksi data akan menghasilkan kesalahan yang lebih besar. Kesalahan rekonstruksi ini dapat digunakan sebagai indikator untuk mendeteksi data yang memiliki karakteristik anomali.



Gambar 2.3. Ilustrasi AutoEncoder
[35]

Kesalahan rekonstruksi (*reconstruction error*) merupakan selisih antara data asli dengan hasil keluaran yang direkonstruksi oleh model *Autoencoder*. Nilai ini digunakan untuk mengukur seberapa baik model mampu mempelajari pola normal pada data. Semakin kecil nilai kesalahan rekonstruksi, semakin baik model dalam merepresentasikan karakteristik data tersebut.

Perhitungan kesalahan rekonstruksi pada penelitian ini menggunakan metode *Mean Squared Error* (*MSE*)[19], yaitu rata-rata kuadrat selisih antara nilai asli dan nilai hasil rekonstruksi. Metode ini dipilih karena sensitif terhadap selisih yang besar, sehingga efektif dalam mendeteksi data yang menyimpang. Rumus *MSE* ditunjukkan sebagai berikut:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2 \tag{2.1}$$

Berdasarkan rumus 2.1, *Mean Squared Error* (MSE) digunakan untuk mengukur rata-rata kesalahan kuadrat antara nilai aktual dan nilai hasil rekonstruksi model. Semakin kecil nilai MSE, maka semakin baik model dalam merekonstruksi data input.

Keterangan:

- n adalah jumlah total data atau fitur yang digunakan dalam perhitungan error.
- x_i adalah nilai aktual pada data ke-i.
- $\hat{x}_i$ adalah nilai hasil prediksi atau rekonstruksi model pada data ke-i.

- $(x_i - \hat{x}_i)^2$ adalah selisih kuadrat antara nilai aktual dan hasil prediksi pada data ke-i.

Sumber: [19]
Keterangan:

- x_i : nilai data asli pada fitur ke-i
- $\hat{x}_i$: nilai hasil rekonstruksi oleh model *Autoencoder* pada fitur ke-i
- n : jumlah seluruh fitur pada data

Apabila nilai *reconstruction error* yang dihasilkan tinggi, maka data tersebut memiliki pola yang berbeda dari data normal sehingga berpotensi dikategorikan sebagai anomali. Sebaliknya, jika nilai kesalahan rekonstruksi rendah, maka data dianggap masih sesuai dengan pola normal yang telah dipelajari model. Oleh karena itu, nilai *MSE* digunakan sebagai dasar dalam menentukan tingkat keanehan suatu transaksi.

2.9 *Threshold Anomaly Detection*

Threshold merupakan nilai batas yang digunakan untuk memisahkan data normal dan anomali berdasarkan skor tertentu yang dihasilkan model. Dalam konteks *anomaly detection*, model umumnya menghasilkan nilai skor atau error yang menunjukkan tingkat penyimpangan suatu data terhadap pola normal. Data dengan nilai skor yang melebihi *threshold* akan diklasifikasikan sebagai anomali, sedangkan data di bawah nilai tersebut diklasifikasikan sebagai normal[14].

Penentuan nilai *threshold* merupakan tahap penting karena secara langsung memengaruhi performa model klasifikasi. Nilai *threshold* yang terlalu rendah dapat meningkatkan jumlah *False Positive*, yaitu data normal yang salah terdeteksi sebagai anomali. Sebaliknya, nilai *threshold* yang terlalu tinggi dapat meningkatkan *False Negative*, yaitu data anomali yang gagal terdeteksi.

Salah satu pendekatan yang umum digunakan dalam menentukan *threshold* adalah metode berbasis persentil (*percentile-based thresholding*), yaitu menetapkan nilai batas berdasarkan distribusi skor anomali atau error pada dataset. Pendekatan ini banyak digunakan pada data dengan distribusi tidak seimbang (*imbalanced dataset*) karena mampu mengidentifikasi nilai ekstrem yang berpotensi sebagai anomali.Secara matematis, penentuan *threshold* berbasis persentil dapat dirumuskan sebagai berikut[36].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2 \tag{2.2}$$

Berdasarkan rumus 2.2, *Threshold* digunakan sebagai nilai batas untuk menentukan apakah suatu data diklasifikasikan sebagai normal atau anomali berdasarkan distribusi nilai *anomaly score*. Data dengan nilai *score* yang melebihi nilai *threshold* akan dikategorikan sebagai anomali atau fraud.

Keterangan:

- $Threshold$ adalah nilai batas klasifikasi untuk membedakan data normal dan anomali.
- P_n adalah fungsi persentil ke-n yang digunakan untuk menentukan nilai batas berdasarkan distribusi data.
- *score* adalah nilai *anomaly score* yang dihasilkan model untuk setiap data.

- n adalah persentase tertentu yang digunakan sebagai cutoff, misalnya persentil ke-95 atau ke-99.

Sumber: [14]
dengan:

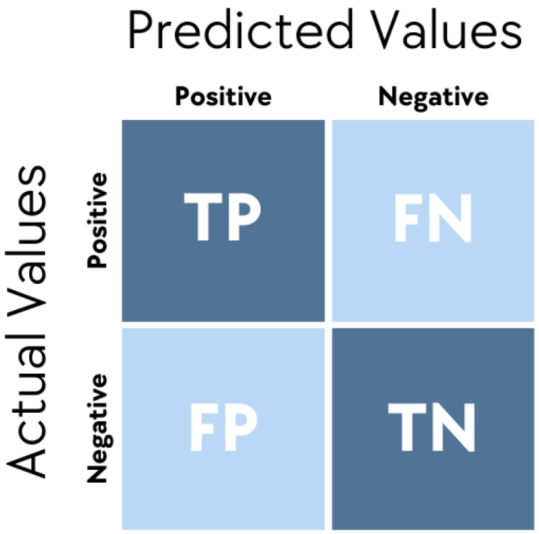
- $Threshold$ = nilai batas klasifikasi,
- P_n = persentil ke-n,
- $score$ = skor anomali atau error yang dihasilkan model.

2.10 *Confussion Matrix*

Confusion Matrix merupakan metode evaluasi yang digunakan untuk mengukur kinerja model klasifikasi dengan membandingkan hasil prediksi model terhadap label aktual. Matriks ini menyajikan jumlah prediksi yang benar maupun salah ke dalam empat kategori utama, yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*.

True Positive (TP) menunjukkan jumlah data positif yang berhasil diprediksi positif oleh model, sedangkan *True Negative (TN)* menunjukkan jumlah data negatif yang berhasil diprediksi negatif. Sebaliknya, *False Positive (FP)* merupakan kondisi ketika data negatif diprediksi sebagai positif, dan *False Negative (FN)* terjadi ketika data positif diprediksi sebagai negatif.

Penggunaan *Confusion Matrix* penting untuk mengevaluasi performa model, khususnya pada klasifikasi tidak seimbang (*imbalanced dataset*). Berdasarkan nilai *True Positive*, *True Negative*, *False Positive*, dan *False Negative*, metrik evaluasi model dapat dihitung sebagaimana ditunjukkan pada Gambar 2.4 berikut.



Gambar 2.4. Confusion Matrix

Sumber: [37]

Berikut merupakan formulasi metrik evaluasi yang digunakan untuk mengukur performa model klasifikasi berdasarkan hasil perhitungan pada *Confusion Matrix*. Adapun metrik evaluasi yang digunakan meliputi *Accuracy*, *Precision*, *Recall*, dan *F1-Score* sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2.3}$$

Rumus (2.3) digunakan untuk menghitung tingkat akurasi keseluruhan model dalam mengklasifikasikan data secara benar, baik data normal maupun fraud. Nilai accuracy yang tinggi menunjukkan bahwa model memiliki performa yang baik dalam melakukan klasifikasi secara umum.

$$Precision = \frac{TP}{TP+FP} \tag{2.4}$$

Precision pada rumus (2.4) digunakan untuk mengukur tingkat ketepatan model dalam mengidentifikasi data fraud dari seluruh prediksi yang diklasifikasikan sebagai fraud. Semakin tinggi nilai precision, semakin kecil kemungkinan model menghasilkan kesalahan klasifikasi positif palsu (*false positive*).

$$Recall = \frac{TP}{TP+FN} \tag{2.5}$$

Recall pada rumus (2.5) digunakan untuk mengukur kemampuan model dalam mendeteksi

seluruh data fraud yang sebenarnya terdapat dalam dataset. Nilai recall yang tinggi menunjukkan bahwa model mampu meminimalkan kesalahan negatif palsu (*false negative*).

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{2.6}$$

Nilai F1-Score pada rumus (2.6) digunakan untuk mengukur keseimbangan antara precision dan recall, terutama pada dataset yang memiliki distribusi kelas tidak seimbang. Metrik ini penting digunakan ketika diperlukan trade-off antara kemampuan mendeteksi fraud dan meminimalkan kesalahan prediksi.

2.11 *Explainable Artificial Intelligence*

Explainable Artificial Intelligence (XAI) merupakan pendekatan dalam pengembangan sistem *machine learning* yang bertujuan untuk meningkatkan transparansi dan interpretabilitas model. Dalam banyak kasus, model *machine learning* yang kompleks sering kali sulit dipahami oleh pengguna karena proses pengambilan keputusan dalam model tidak dapat dijelaskan secara langsung[38].

Pendekatan *XAI* bertujuan untuk memberikan penjelasan mengenai bagaimana model menghasilkan suatu prediksi. Dengan demikian, pengguna dapat memahami faktor-faktor yang mempengaruhi keputusan model serta meningkatkan kepercayaan terhadap sistem yang digunakan.

Metode yang umum digunakan dalam pendekatan *XAI* antara lain *SHAP (SHapley Additive exPlanations)* dan *LIME (Local Interpretable Model-agnostic Explanations)*, yang digunakan untuk menganalisis kontribusi setiap fitur terhadap hasil prediksi model.

2.11.1 SHAP (SHapley Additive exPlanations)

SHAP (*SHapley Additive exPlanations*) merupakan metode *Explainable Artificial Intelligence (XAI)* yang digunakan untuk menjelaskan kontribusi setiap fitur terhadap hasil prediksi model *machine learning*. Metode ini berasal dari konsep *Shapley Value* dalam *cooperative game theory*, yang menghitung kontribusi setiap pemain terhadap hasil akhir suatu permainan. Dalam konteks *machine learning*, setiap fitur dianggap sebagai pemain yang berkontribusi terhadap nilai prediksi model[21].

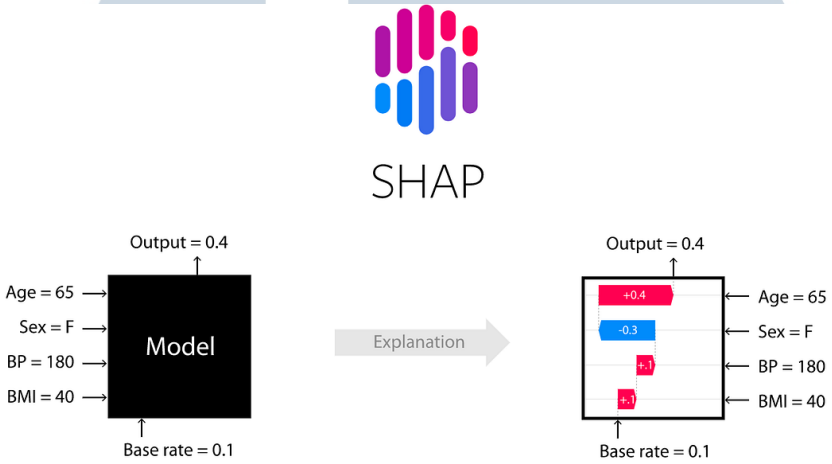
A Keunggulan SHAP

Beberapa keunggulan dari metode SHAP antara lain:

- Memberikan interpretasi model yang konsisten secara matematis
- Dapat digunakan pada berbagai jenis model (*model-agnostic*)
- Menyediakan penjelasan baik secara global maupun lokal terhadap hasil prediksi

B Relevansi SHAP dalam Penelitian

Dalam penelitian deteksi anomali transaksi, SHAP digunakan untuk menjelaskan fitur apa saja yang menyebabkan suatu transaksi diklasifikasikan sebagai anomali atau transaksi normal oleh model. Hal ini penting untuk meningkatkan transparansi sistem deteksi *fraud*, khususnya pada sektor keuangan yang membutuhkan tingkat interpretabilitas dan akuntabilitas yang tinggi seperti yang ada dari Gambar 2.4.



Gambar 2.5. Visualisasi SHAP

Sumber: [39]

Gambar 2.4 menunjukkan visualisasi metode *shap* dalam menjelaskan hasil prediksi model. Setiap fitur diberikan nilai kontribusi terhadap output prediksi, baik bersifat positif maupun negatif. Pada ilustrasi tersebut, fitur yang mendorong kenaikan nilai prediksi ditunjukkan pada satu sisi, sedangkan fitur yang menurunkan prediksi berada pada sisi lainnya. Melalui visualisasi ini, pengguna dapat memahami alasan model menghasilkan suatu keputusan secara lebih transparan. Dengan demikian, *shap* membantu mengidentifikasi fitur mana yang paling berpengaruh terhadap hasil prediksi pada suatu data tertentu.

2.12 LIME (Local Interpretable Model-Agnostic Explanations)

LIME (*Local Interpretable Model-Agnostic Explanations*) merupakan metode *Explainable Artificial Intelligence (XAI)* yang digunakan untuk menjelaskan prediksi model *machine learning* secara lokal. Metode ini diperkenalkan oleh Ribeiro et al. untuk membantu memahami bagaimana sebuah model membuat keputusan terhadap suatu data tertentu.*lime* bersifat *model-agnostic*, yang berarti metode ini dapat digunakan untuk menjelaskan berbagai jenis model *machine learning* tanpa bergantung pada struktur internal model tersebut.

Tujuan utama dari *lime* adalah memberikan interpretasi yang sederhana dan mudah dipahami mengenai alasan suatu prediksi dihasilkan oleh model. Dengan demikian, pengguna

dapat mengetahui fitur apa saja yang paling berpengaruh terhadap keputusan model pada suatu data tertentu[21].

A Konsep Dasar *lime*

Metode *lime* bekerja dengan membangun model interpretasi sederhana di sekitar data yang ingin dijelaskan. Proses ini dilakukan dengan menghasilkan beberapa variasi data di sekitar sampel yang dianalisis, kemudian mengamati bagaimana perubahan pada fitur tersebut mempengaruhi hasil prediksi model.

Secara umum proses *lime* dapat dijelaskan melalui beberapa tahapan sebagai berikut:

1. Membuat beberapa sampel data baru di sekitar data yang akan dijelaskan.
2. Menghitung prediksi model terhadap sampel-sampel tersebut.
3. Memberikan bobot pada setiap sampel berdasarkan kedekatannya dengan data asli.
4. Membangun model sederhana, seperti *linear regression*, untuk menjelaskan hubungan antara fitur dan prediksi model.

Melalui pendekatan ini, *lime* dapat memberikan interpretasi lokal terhadap keputusan model dengan menunjukkan fitur-fitur yang paling berpengaruh terhadap prediksi.

B Keunggulan LIME

Beberapa keunggulan dari metode *lime* antara lain:

- Dapat digunakan pada berbagai jenis model *machine learning* (*model-agnostic*)
- Memberikan interpretasi yang mudah dipahami oleh pengguna
- Mampu menjelaskan prediksi model secara lokal pada data tertentu

C Relevansi *lime* dalam Penelitian

Dalam penelitian deteksi anomali transaksi, *lime* digunakan untuk menjelaskan alasan suatu transaksi diklasifikasikan sebagai transaksi normal atau anomali oleh model. Dengan menggunakan metode ini, fitur-fitur yang memiliki pengaruh signifikan terhadap keputusan model dapat diidentifikasi, sehingga membantu meningkatkan transparansi serta kepercayaan terhadap sistem deteksi fraud yang dikembangkan.