

BAB II

PENDAHULUAN

II.1 Konsep Dasar Deteksi Kecurangan

II.1.1 Definisi Kecurangan Tagihan Listrik

Kecurangan tagihan listrik dapat didefinisikan sebagai setiap tindakan yang disengaja atau tidak disengaja yang menyebabkan PLN tidak menerima pembayaran yang sesuai dengan jumlah listrik yang sebenarnya digunakan oleh pelanggan (Budianto & Saragih, 2012). Menurut Peraturan Direksi PLN No. 088-Z/DIR/2016 tentang Penertiban Pemakaian Tenaga Listrik (P2TL), kecurangan listrik termasuk dalam kategori pelanggaran yang dapat dikenai sanksi administrasi maupun pidana.

Secara lebih spesifik, kecurangan tagihan listrik dapat diartikan sebagai setiap aktivitas yang mengakibatkan:

1. Konsumsi listrik tidak tercatat atau tercatat kurang: Pelanggan menggunakan listrik tetapi tidak tercatat pada meteran atau tercatat jumlah yang lebih kecil dari yang sebenarnya digunakan
2. Pembayaran tagihan yang tidak sesuai: Pelanggan membayar tagihan yang lebih rendah dari yang seharusnya dibayarkan berdasarkan konsumsi aktual
3. Penyalahgunaan layanan: Pelanggan menggunakan layanan listrik dengan cara yang tidak sesuai dengan ketentuan atau memanfaatkan kelemahan sistem untuk mendapatkan keuntungan yang tidak sah

Kecurangan tagihan listrik berbeda dengan kesalahan teknis atau administratif yang mungkin terjadi dalam sistem penagihan. Perbedaan utama terletak pada niat dan sengaja; kecurangan dilakukan dengan maksud untuk mendapatkan keuntungan yang tidak sah, sedangkan kesalahan teknis atau administratif terjadi tanpa sengaja dan biasanya dapat diperbaiki dengan proses koreksi yang sesuai.

Menurut Khan et al. (2022), kecurangan tagihan listrik dapat diklasifikasikan menjadi tiga kategori utama berdasarkan sifat dan cara pelaksanaannya:

1. Kecurangan yang bersifat fisik: Melibatkan modifikasi atau manipulasi pada peralatan kelistrikan seperti meteran listrik, kabel, atau perangkat keras lainnya
2. Kecurangan yang bersifat administratif: Melibatkan manipulasi pada data atau proses administratif seperti pembacaan meteran, pencatatan data, atau penetapan tarif
3. Kecurangan yang bersifat teknologi: Melibatkan penggunaan teknologi atau perangkat lunak untuk memanipulasi sistem pembacaan atau penagihan listrik, terutama pada sistem meteran pintar

II.1.2 Jenis-Jenis Kecurangan Dalam Tagihan Listrik

Berdasarkan studi yang dilakukan oleh Budianto & Saragih (2012) serta penelitian lapangan yang dilakukan oleh PLN, jenis-jenis kecurangan tagihan listrik yang umum terjadi antara lain:

1. Manipulasi Meteran Listrik

Merupakan jenis kecurangan yang paling sering ditemui, yang dilakukan dengan cara memodifikasi atau mengganti meteran listrik agar tidak mencatat konsumsi listrik dengan benar. Beberapa bentuk manipulasi meteran listrik antara lain:

a. Penggantian Meteran dengan yang Dimodifikasi

Pelanggan mengganti meteran resmi yang dipasang oleh PLN dengan meteran yang telah dimodifikasi atau dipalsukan. Meteran yang dimodifikasi ini dirancang untuk mencatat konsumsi listrik yang lebih rendah dari yang sebenarnya digunakan atau bahkan tidak mencatat konsumsi sama sekali.



Gambar II. 1 Contoh Meteran Listrik yang di Modifikasi

b. Penggunaan Alat Penghambat

Pelanggan menggunakan alat khusus seperti magnet kuat untuk memperlambat atau menghentikan putaran mekanisme pada meteran listrik analog. Pada meteran digital, dapat digunakan alat untuk mengganggu sinyal atau proses pengukuran sehingga hasil pembacaan menjadi tidak akurat.



Gambar II. 2 Meteran Listrik yang ada Penghambatnya

c. Modifikasi Internal Meteran

Pelanggan membongkar dan memodifikasi komponen internal meteran listrik, seperti mengganti resistor, kapasitor, atau komponen elektronik lainnya agar meteran bekerja tidak sesuai dengan spesifikasi yang ditetapkan.



Gambar II. 3 Internbal Meteran di Modifikasi

d. Pembalikan Arah Putaran Meteran

Pada meteran analog, pelanggan melakukan pembalikan arah putaran meteran dengan cara membalikkan polaritas sambungan kabel, sehingga meteran berputar ke arah yang berlawanan dan mencatat konsumsi yang semakin berkurang.



Gambar II. 4 Kwh Meter Analog

2. Sambungan Ilegal

Sambungan ilegal merupakan jenis kecurangan yang dilakukan dengan cara menyambungkan kabel listrik langsung ke jaringan PLN tanpa melalui meteran resmi atau dengan cara yang tidak sesuai dengan ketentuan.



Gambar II. 5 Sambungan Ilegal pada Meteran listrik

Beberapa bentuk sambungan ilegal antara lain:

a. Sambungan Langsung ke Jaringan Distribusi

Pelanggan membuat sambungan kabel langsung ke saluran distribusi listrik

PLN tanpa melalui meteran atau dengan melewati meteran yang telah dipasang. Hal ini menyebabkan konsumsi listrik tidak tercatat sama sekali.

b. Bypass Meteran

Pelanggan membuat jalur alternatif untuk aliran listrik yang menghindari meteran resmi, sehingga sebagian atau seluruh konsumsi listrik tidak tercatat oleh meteran.

c. Sambungan Antar Pelanggan

Pelanggan yang telah terdaftar dengan benar menyambungkan kabel listrik ke pelanggan lain yang tidak terdaftar atau telah melakukan kecurangan, sehingga beban konsumsi dibagi atau dialihkan.

d. Penggunaan Kabel Sepadu yang Dimodifikasi

Pelanggan memodifikasi kabel sepadu yang menghubungkan jaringan PLN ke rumah pelanggan dengan cara menambahkan konduktor tersembunyi untuk mengalirkan listrik tanpa melalui meteran atau dengan cara memanipulasi isolasi kabel agar dapat mengambil arus listrik tanpa terdeteksi.

3. Manipulasi Pembacaan dan Data Tagihan

Jenis kecurangan ini dilakukan dengan cara memanipulasi data pembacaan meteran atau proses pencatatan tagihan agar nilai yang tercatat lebih rendah dari yang sebenarnya. Bentuk-bentuknya antara lain:

a. Pembacaan Meteran yang Salah

Baik oleh pelanggan maupun petugas pembaca meteran, dilakukan pembacaan nilai meteran yang tidak akurat atau sengaja membaca nilai yang lebih rendah dari yang sebenarnya tercatat.

b. Pemalsuan Bukti Pembacaan

Pelanggan memalsukan formulir atau bukti pembacaan meteran dengan cara mengubah angka pembacaan atau membuat dokumen palsu yang menunjukkan konsumsi listrik yang lebih rendah.

c. Manipulasi Data dalam Sistem Informasi

Melalui akses yang tidak sah ke sistem informasi PLN, pihak yang tidak bertanggung jawab memodifikasi data konsumsi atau tagihan pelanggan

agar nilai yang tercatat menjadi lebih rendah.

d. Penundaan atau Penghapusan Pencatatan Tagihan

Melakukan tindakan untuk menanggukuhkan atau menghapus pencatatan tagihan pada sistem PLN, sehingga pelanggan tidak menerima tagihan atau menerima tagihan dengan nilai yang tidak sesuai.

4. Penipuan Identitas dan Penyalahgunaan Program

Jenis kecurangan ini dilakukan dengan cara memanfaatkan identitas atau program yang ada untuk mendapatkan keuntungan yang tidak sah. Beberapa bentuknya antara lain:

a. Pendaftaran dengan Identitas Palsu

Mendaftarkan layanan listrik menggunakan identitas yang tidak benar atau orang lain tanpa sepengetahuan mereka, sehingga ketika terjadi tagihan yang tidak dibayar, pihak yang bersangkutan tidak dapat dilacak.

b. Penyalahgunaan Program Subsidi

Memanfaatkan program subsidi listrik yang ditujukan untuk masyarakat kurang mampu dengan cara memberikan informasi palsu atau tidak memenuhi syarat tetapi tetap mendapatkan tarif subsidi.

c. Penggantian Kategori Pelanggan

Mengubah kategori pelanggan dari yang memiliki tarif lebih tinggi (misalnya bisnis atau industri) menjadi kategori dengan tarif lebih rendah (misalnya residensial) tanpa melalui proses yang sah.

d. Penipuan dalam Proses Perubahan atau Pemutusan Layanan

Melakukan manipulasi dalam proses perubahan data pelanggan atau pemutusan layanan untuk mendapatkan keuntungan seperti pengembalian uang yang tidak sesuai atau penghapusan tagihan yang belum lunas.

II.1.3 Dampak Kecurangan Bagi PLN Dan Masyarakat

Kecurangan tagihan listrik tidak hanya memberikan dampak negatif bagi PLN tetapi juga bagi masyarakat luas dan pembangunan negara secara keseluruhan. Berikut adalah beberapa dampak utama yang ditimbulkan:

1. Dampak Bagi Perusahaan Listrik Negara (PLN)

a. Kerugian Finansial yang Besar

Kerugian akibat kecurangan listrik mencapai jumlah yang sangat signifikan setiap tahunnya. Menurut laporan PLN tahun 2024, kerugian nasional akibat kecurangan listrik mencapai lebih dari 50 triliun rupiah per tahun. Kerugian ini berasal dari listrik yang digunakan tetapi tidak dibayar oleh pelanggan yang melakukan kecurangan.

b. Gangguan pada Perencanaan Keuangan

Kerugian akibat kecurangan membuat perencanaan keuangan PLN menjadi sulit dilakukan dengan akurat. Pendapatan yang seharusnya diterima tidak dapat diperhitungkan dengan tepat, sehingga mempengaruhi kemampuan PLN untuk melakukan investasi pada infrastruktur baru atau pemeliharaan infrastruktur yang ada.

c. Meningkatnya Beban Operasional

PLN harus mengalokasikan sumber daya yang besar untuk mendeteksi dan menangani kecurangan, termasuk biaya untuk tenaga kerja, peralatan deteksi, serta proses hukum dan penertiban. Beban operasional ini seharusnya dapat dialokasikan untuk meningkatkan kualitas layanan atau mengembangkan infrastruktur.

d. Pengaruh pada Kualitas dan Keandalan Pasokan

Sambungan ilegal dan manipulasi meteran seringkali menyebabkan gangguan pada jaringan distribusi listrik, seperti korsleting, kebakaran, atau kerusakan pada peralatan kelistrikan. Hal ini dapat menyebabkan pemadaman listrik yang tidak terduga dan mengurangi keandalan pasokan listrik bagi seluruh pelanggan.

e. Penurunan Citra dan Kepercayaan Masyarakat

Meskipun kecurangan dilakukan oleh sebagian kecil pelanggan, namun hal ini dapat menurunkan citra PLN sebagai perusahaan yang bertanggung jawab. Selain itu, pelanggan yang jujur dapat merasa tidak adil jika melihat adanya pelanggan lain yang melakukan kecurangan tanpa dikenakan sanksi.

2. Dampak Bagi Masyarakat dan Negara

a. Kenaikan Tarif Listrik

Kerugian yang dialami PLN akibat kecurangan seringkali menjadi alasan untuk kenaikan tarif listrik. Hal ini berarti beban biaya listrik harus ditanggung oleh seluruh pelanggan yang jujur, termasuk masyarakat kurang mampu yang sangat bergantung pada listrik untuk kebutuhan dasar.

b. Ketidakadilan Ekonomi

Pelanggan yang melakukan kecurangan mendapatkan keuntungan yang tidak sah dengan membayar lebih sedikit atau bahkan tidak membayar tagihan listrik sama sekali, sementara pelanggan yang jujur harus membayar sesuai dengan konsumsi mereka. Hal ini menciptakan ketidakadilan ekonomi di antara masyarakat.

c. Gangguan pada Pembangunan Daerah

Pendapatan yang hilang akibat kecurangan dapat mengurangi kemampuan PLN untuk mengembangkan infrastruktur kelistrikan di daerah-daerah yang belum terjangkau atau membutuhkan peningkatan pasokan listrik. Hal ini dapat menghambat pembangunan ekonomi dan sosial di daerah tersebut.

d. Resiko Keamanan yang Tinggi

Sambungan ilegal dan manipulasi peralatan kelistrikan memiliki resiko tinggi menyebabkan kebakaran, sengatan listrik, atau kecelakaan lainnya yang dapat membahayakan keselamatan jiwa dan harta benda masyarakat sekitar.

e. Kerugian Negara Sebagai badan usaha milik negara, kerugian yang dialami PLN juga merupakan kerugian bagi negara. Pendapatan yang seharusnya masuk ke kas negara tidak dapat terkumpul dengan maksimal, sehingga dapat mempengaruhi kemampuan negara untuk membiayai program-program pembangunan dan pelayanan publik lainnya.

II.2 Analitik Data Dan Kecurangan Buatan

II.2.1 Analitik Data Untuk Deteksi Anomali

Analitik data adalah proses mengumpulkan, memproses, dan menganalisis data untuk menemukan pola, tren, dan anomali. (C. L. Pratama et al., 2023) menyatakan bahwa analitik data dapat membantu organisasi dalam pengambilan keputusan yang lebih baik dan lebih cepat. Leonardo pratama juga mencatat bahwa analitik data memainkan peran penting dalam mendeteksi kecurangan dengan mengidentifikasi perilaku yang tidak biasa atau mencurigakan. Analitik data adalah proses mengumpulkan, mengolah, dan menganalisis data untuk mendapatkan wawasan yang bermakna dan berguna dalam pengambilan keputusan. Proses ini melibatkan berbagai teknik statistik, komputasi, dan algoritma untuk mengeksplorasi data mentah dan mengubahnya menjadi informasi yang bermanfaat. Analitik data dapat diterapkan dalam berbagai bidang, termasuk bisnis, kesehatan, ilmu pengetahuan, dan teknik. Analitik data melibatkan pengumpulan, pemrosesan, dan analisis data yang besar untuk mengidentifikasi tren, pola, dan anomali. Dalam konteks ini, data dari meteran pintar, riwayat pembayaran, dan profil pelanggan dianalisis untuk mendeteksi aktivitas yang mencurigakan. Sistem deteksi dini kecurangan tagihan listrik bertujuan untuk mengidentifikasi pola penggunaan energi yang mencurigakan yang dapat mengindikasikan kecurangan, seperti pencurian listrik, manipulasi meteran, atau pelaporan data yang tidak akurat.

Sistem Deteksi adalah mekanisme atau alat yang dirancang untuk mengidentifikasi kejadian, pola, atau anomali tertentu dalam data. Sistem ini memiliki aplikasi luas di berbagai bidang seperti deteksi penipuan, keamanan siber, pemeliharaan prediktif, dan diagnosis medis (Ismawati & Ramadhanti, 2022). Berikut adalah beberapa konsep kunci dan komponen dalam teori sistem deteksi.

1. Komponen Utama Sistem Deteksi

Sensor atau *input* data dengan cara mengumpulkan data dari berbagai sumber seperti sensor fisik, meteran listrik, *log* komputer, atau basis data pelanggan, contohnya dalam sistem deteksi kecurangan tagihan listrik, data bisa diambil dari meteran listrik pintar yang mengirimkan informasi penggunaan energi secara berkala.

2. Preprocessing Data

Melakukan pembersihan data untuk menghilangkan noise, menangani data yang hilang, dan melakukan normalisasi data, contohnya menormalisasi data penggunaan listrik agar berada dalam skala yang konsisten untuk analisis lebih lanjut.

3. Ekstraksi Fitur

Mengidentifikasi dan mengekstraksi fitur relevan dari data mentah yang dapat membantu dalam mendeteksi pola atau anomali. menghitung rata-rata penggunaan listrik per hari, per minggu, dan per bulan.

4. Analitik Data

Menggunakan teknik statistik dan pembelajaran mesin untuk menganalisis data dan mengidentifikasi pola atau anomali, menggunakan algoritma *clustering* untuk mengelompokkan pola penggunaan listrik yang normal dan anomali.

5. Output dan Tindakan

Menghasilkan laporan atau peringatan berdasarkan analisis data dan menginisiasi tindakan yang diperlukan, seperti investigasi lebih lanjut atau pemberitahuan kepada pihak terkait dan mengirimkan peringatan kepada pelanggan atau penyedia listrik jika terdeteksi pola penggunaan yang mencurigakan.

6. Teknik Deteksi Anomali Statistical Methods

Menggunakan teknik statistik untuk mengidentifikasi anomali berdasarkan asumsi distribusi data. Deteksi anomali menggunakan aturan z-score, di mana data yang memiliki z-score di luar batas tertentu dianggap sebagai anomali.

7. *Clustering-Based Methods*

Mengelompokkan data ke dalam cluster dan mengidentifikasi data yang tidak sesuai dengan cluster manapun sebagai anomali. Algoritma K-Means atau DBSCAN yang digunakan untuk mengelompokkan pola penggunaan listrik dan mendeteksi anomali sebagai *outliers*.

8. *Classification-Based Methods*

Menggunakan model klasifikasi untuk membedakan antara data normal dan anomali. Menggunakan *Random Forest* atau *Support Vector Machines* (SVM) untuk mengklasifikasikan pola penggunaan listrik sebagai normal atau mencurigakan.

9. *Proximity-Based Methods*

Mengukur jarak antara titik data dan mengidentifikasi titik yang memiliki jarak besar dari titik lain sebagai anomali. Algoritma k-Nearest Neighbors (k-NN) yang mengidentifikasi anomali berdasarkan jarak ke tetangga terdekatnya.

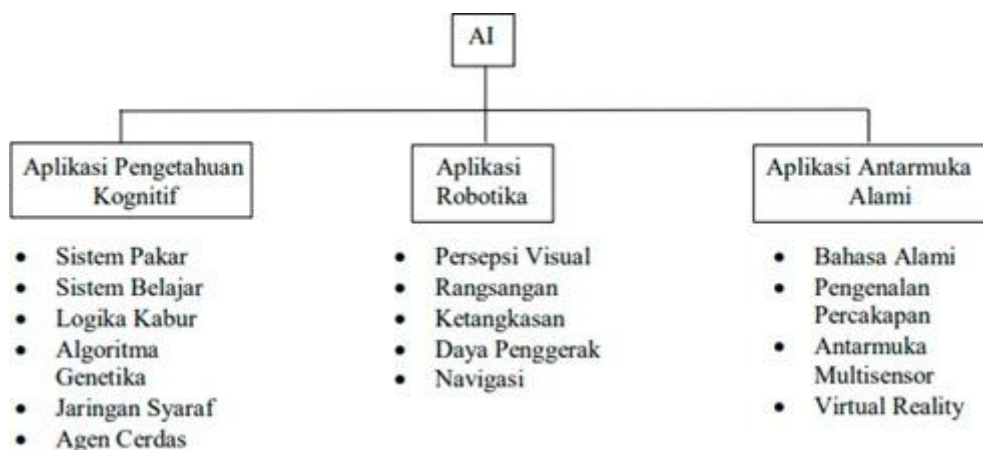
10. *Reconstruction-Based Methods*

Menggunakan model untuk merekonstruksi data dan mengidentifikasi data yang tidak dapat direkonstruksi dengan baik sebagai anomali, *Autoencoders* dalam deep learning yang digunakan untuk mendeteksi anomali berdasarkan perbedaan antara *input* dan

rekonstruksinya.

Menurut Luger dan William, kecerdasan buatan adalah cabang ilmu komputer yang berhubungan dengan otomasi perilaku yang cerdas (1993). Kemudian menurut Haag dan Peter kecerdasan buatan adalah bidang studi yang berhubungan dengan penangkapan, pemodelan, dan penyimpanan kecerdasan manusia ke dalam sebuah sistem teknologi informasi sehingga sistem tersebut dapat digunakan sebagai proses pengambilan keputusan yang dilakukan oleh manusia (1996). Dari dua pengertian tersebut dapat disimpulkan bahwa kecerdasan buatan adalah cabang ilmu komputer yang membahas tentang penangkapan, pemodelan, dan penyimpanan kecerdasan manusia ke dalam sebuah teknologi informasi yang nantinya dapat dimanfaatkan untuk pengambilan keputusan (Octaviani & Dewi, 2020).

Kecerdasan Buatan (*Artificial Intelligence*, AI) adalah bidang ilmu komputer yang berfokus pada penciptaan sistem yang mampu melakukan tugas-tugas yang biasanya memerlukan kecerdasan manusia (Muttaqin, Muhammad Arafa, 2018). AI mencakup berbagai teknik dan pendekatan untuk membuat mesin yang dapat memahami, belajar, beradaptasi, dan melakukan pekerjaan yang memerlukan kecerdasan. Berikut adalah penjelasan lebih mendalam tentang teori dan konsep utama dalam kecerdasan buatan. Menurut O'brien dalam Kadir dan Terra (2003: 331) pengelompokan domain aplikasi kecerdasan buatan dapat dilihat dari gambar II.1



Gambar II. 6 Domain Aplikasi Utama Ai

II.2.2 Konsep Kecerdasan Buatan dan Pembelajaran Mesin

Kecerdasan Buatan (*Artificial Intelligence/AI*) adalah cabang ilmu komputer yang berfokus pada penciptaan sistem yang mampu melakukan tugas-tugas yang biasanya memerlukan kecerdasan manusia (Muttaqin, 2018). Menurut Ismawati & Ramadhanti (2022),

kecerdasan buatan mencakup berbagai teknik dan pendekatan untuk membuat mesin yang dapat memahami, belajar, beradaptasi, dan melakukan pekerjaan yang memerlukan kecurangan.

Pembelajaran Mesin (Machine Learning/ML) adalah cabang dari kecerdasan buatan yang fokus pada pengembangan algoritma yang memungkinkan komputer untuk belajar dari data dan pengalaman tanpa harus diprogram secara eksplisit (Muttaqin, 2018). Dalam konteks deteksi kecurangan, pembelajaran mesin digunakan untuk mengidentifikasi pola-pola anomali dalam data konsumsi listrik yang mengindikasikan adanya kecurangan.

1. Jenis-Jenis Pembelajaran Mesin

Berdasarkan cara model belajar dari data, pembelajaran mesin dapat dibagi menjadi tiga kategori utama:

a. Supervised Learning (Pembelajaran Terawasi)

- Pada pembelajaran terawasi, model dilatih menggunakan data yang telah diberi label (*labeled data*), yang berarti setiap contoh dalam dataset memiliki informasi tentang hasil yang diinginkan. Dalam deteksi kecurangan, label dapat berupa "normal" atau "curang" untuk setiap catatan konsumsi listrik. Jenis tugas yang umum dilakukan dengan pembelajaran terawasi antara lain:
- Klasifikasi: Mengklasifikasikan data ke dalam kategori tertentu (misalnya, membedakan antara tagihan normal dan tagihan curang)
- Regresi: Memprediksi nilai numerik kontinu (misalnya, memprediksi konsumsi listrik pelanggan di masa depan)

Algoritma pembelajaran terawasi yang umum digunakan dalam deteksi kecurangan antara lain *Random Forest*, *Decision Tree*, *Support Vector Machine (SVM)*, *Logistic Regression*, dan *Neural Network*.

b. Unsupervised Learning (Pembelajaran Tidak Terawasi)

Pada pembelajaran tidak terawasi, model dilatih menggunakan data yang tidak diberi label (*unlabeled data*), dan tugas model adalah untuk menemukan pola atau struktur tersembunyi dalam data tersebut. Dalam deteksi kecurangan, pembelajaran tidak terawasi digunakan untuk mengidentifikasi pola konsumsi normal dan kemudian mendeteksi data yang tidak sesuai dengan pola tersebut sebagai anomali. Jenis tugas yang umum dilakukan dengan pembelajaran tidak terawasi antara lain:

- *Clustering*: Mengelompokkan data ke dalam kelompok-kelompok yang mirip

(misalnya, mengelompokkan pelanggan berdasarkan pola konsumsi mereka)

- *Dimensionality Reduction*: Mengurangi jumlah fitur dalam data tanpa kehilangan informasi penting
- *Anomaly Detection*: Mengidentifikasi data yang berbeda secara signifikan dari sebagian besar data lainnya

Algoritma pembelajaran tidak terawasi yang umum digunakan dalam deteksi kecurangan antara lain *K-Means*, *DBSCAN*, *Isolation Forest*, dan *Autoencoder*.

c. *Reinforcement Learning* (Pembelajaran Penguatan)

Pada pembelajaran penguatan, model belajar dengan cara berinteraksi dengan lingkungan dan menerima umpan balik berupa hadiah atau hukuman berdasarkan tindakan yang diambil. Dalam deteksi kecurangan, pembelajaran penguatan dapat digunakan untuk mengoptimalkan keputusan tentang tindakan yang harus diambil terhadap pelanggan yang dicurigai melakukan kecurangan.

2. Deteksi Anomali dalam Pembelajaran Mesin

Deteksi anomali (*anomaly detection*) adalah cabang dari pembelajaran mesin yang fokus pada pengidentifikasian data yang berbeda secara signifikan dari sebagian besar data lainnya (Ismawati & Ramadhanti, 2022). Dalam konteks deteksi kecurangan tagihan listrik, anomali dapat didefinisikan sebagai pola konsumsi listrik yang tidak sesuai dengan pola normal pelanggan atau kelompok pelanggan sejenis.

Menurut Luger & William (1993), metode deteksi anomali dapat dibagi menjadi beberapa kategori utama:

a. Statistical Methods (Metode Statistik)

Menggunakan teknik statistik untuk mengidentifikasi anomali berdasarkan asumsi distribusi data. Contohnya adalah metode z-score, di mana data yang memiliki nilai z-score di luar batas tertentu dianggap sebagai anomali. Metode ini efektif jika data mengikuti distribusi yang diketahui, tetapi kurang efektif untuk data dengan struktur yang kompleks.

b. *Clustering-Based Methods* (Metode Berbasis Klasterisasi)

Mengelompokkan data ke dalam kelompok (*cluster*) dan mengidentifikasi data yang tidak sesuai dengan cluster manapun sebagai anomali. Algoritma seperti *K-Means* atau *DBSCAN* digunakan untuk mengelompokkan pola penggunaan listrik dan mendeteksi anomali sebagai *outlier*.

c. *Classification-Based Methods* (Metode Berbasis Klasifikasi)

Menggunakan model klasifikasi yang telah dilatih untuk membedakan antara data normal dan anomali. Algoritma seperti Random Forest, SVM, atau Neural Network digunakan untuk mengklasifikasikan pola penggunaan listrik sebagai normal atau mencurigakan.

d. *Proximity-Based Methods* (Metode Berbasis Jarak)

Mengukur jarak antara titik data dan mengidentifikasi titik yang memiliki jarak besar dari titik lain sebagai anomali. Algoritma *k-Nearest Neighbors* (k-NN) digunakan untuk mengidentifikasi anomali berdasarkan jarak ke tetangga terdekatnya.

e. *Reconstruction-Based Methods* (Metode Berbasis Rekonstruksi)

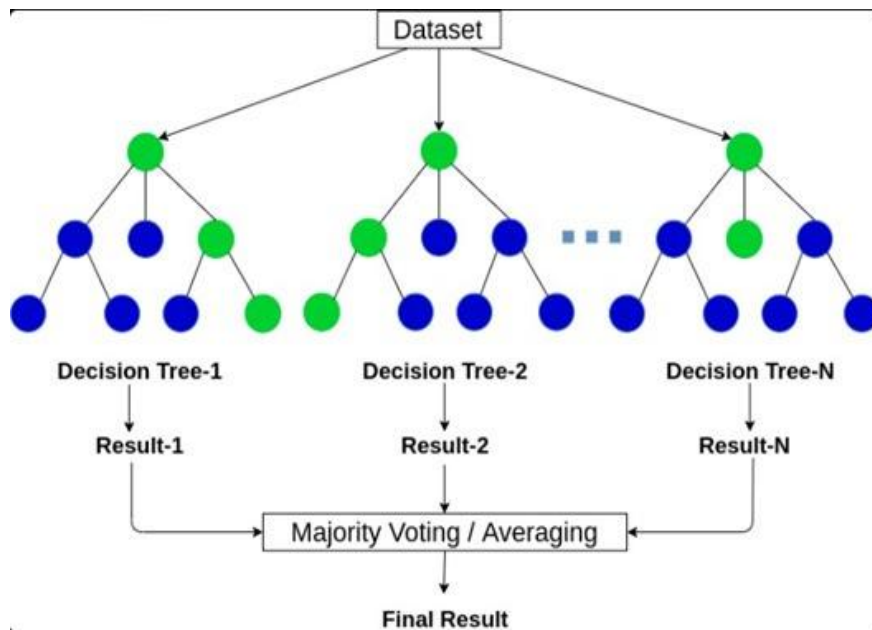
Menggunakan model untuk merekonstruksi data dan mengidentifikasi data yang tidak dapat direkonstruksi dengan baik sebagai anomali. Contohnya adalah Autoencoder dalam deep learning yang digunakan untuk mendeteksi anomali berdasarkan perbedaan antara input dan rekonstruksinya.

Dalam penelitian ini, akan digunakan pendekatan kombinasi antara metode berbasis klasifikasi (menggunakan *Random Forest*) dan metode berbasis rekonstruksi/ekstraksi fitur kompleks (menggunakan CNN) untuk mendeteksi kecurangan tagihan listrik. Pendekatan ini diharapkan dapat menggabungkan keunggulan dari kedua metode dan menghasilkan akurasi deteksi yang lebih tinggi.

II.3 Algoritma *Random Forest*

II.3.1 Konsep Dasar *Random Forest*

Random Forest adalah algoritma *ensemble learning* yang digunakan untuk klasifikasi dan regresi. Algoritma ini pertama kali diperkenalkan oleh Leo Breiman pada tahun 2001 dan merupakan pengembangan dari konsep pohon keputusan (decision tree) tunggal (Anisa & Novita, 2023).



Gambar II. 7 Pohon Keputusan

Konsep dasar dari *Random Forest* adalah menggabungkan banyak pohon keputusan yang dilatih secara independen untuk menghasilkan prediksi yang lebih kuat dan akurat dibandingkan dengan menggunakan satu pohon keputusan saja. Setiap pohon dalam hutan (forest) memproses subset yang berbeda dari data dan fitur secara acak, yang dikenal sebagai metode bagging (*Bootstrap Aggregating*) (Ayub et al., 2022). Dalam konteks deteksi kecurangan tagihan listrik, *Random Forest* bekerja dengan cara berikut:

1. Membangun Banyak Pohon Keputusan: Model membuat banyak pohon keputusan, masing-masing dilatih dengan subset data yang berbeda yang diambil secara acak dari dataset utama dengan metode bootstrap sampling.
2. Memilih Fitur Secara Acak: Pada setiap *node* dalam pohon keputusan, hanya subset acak dari fitur yang dipertimbangkan untuk pemisahan terbaik. Hal ini mengurangi korelasi antar pohon dan meningkatkan keragaman model.
3. Menggabungkan Hasil Prediksi: Untuk tugas klasifikasi, hasil prediksi akhir ditentukan berdasarkan suara mayoritas dari semua pohon dalam hutan. Untuk tugas regresi, hasil prediksi akhir adalah rata-rata dari prediksi semua pohon.

Keunggulan utama dari *Random Forest* dibandingkan dengan pohon keputusan tunggal adalah kemampuannya untuk mengurangi masalah *overfitting*. Pohon keputusan tunggal

cenderung terlalu menyesuaikan diri dengan data pelatihan (*overfit*), sehingga performanya menurun saat diuji dengan data baru. Dengan menggabungkan banyak pohon keputusan yang berbeda, *Random Forest* mampu menghasilkan model yang lebih robust dan memiliki kemampuan generalisasi yang lebih baik.

II.3.2 Komponen Utama Random Forest

Menurut Anisa & Novita (2023), komponen utama dari algoritma *Random Forest* terdiri dari beberapa elemen penting yang bekerja bersama untuk menghasilkan model prediksi yang kuat dan akurat:

1. *Bootstrap Sampling*

- a. Deskripsi: Teknik pengambilan sampel acak dengan penggantian dari dataset asli untuk membuat beberapa subset data yang disebut *bootstrap samples*. Setiap subset ini memiliki ukuran yang sama dengan dataset asli, tetapi karena dilakukan dengan penggantian, beberapa data mungkin muncul lebih dari sekali dan beberapa lainnya mungkin tidak muncul sama sekali.
- b. Jumlah Sampel: Biasanya dibuat sebanyak jumlah pohon keputusan yang akan dibangun (misalnya, 100 atau 200 *bootstrap samples* untuk 100 atau 200 pohon keputusan).

Fungsi:

- Memastikan bahwa setiap pohon keputusan mendapatkan variasi data yang berbeda, sehingga meningkatkan robustness model terhadap *overfitting*
- Mengurangi varians model dengan menggabungkan hasil dari beberapa pohon yang dilatih dengan data yang berbeda
- Memungkinkan evaluasi performa model melalui *out-of-bag* (OOB) error, yaitu error yang dihitung dari data yang tidak termasuk dalam *bootstrap sample* untuk setiap pohon

2. Pohon Keputusan (*Decision Trees*)

- a. Deskripsi: Pohon keputusan adalah model prediksi yang memetakan pengamatan tentang suatu item ke kesimpulan tentang nilai target item tersebut. Setiap node dalam pohon mewakili fitur, setiap cabang mewakili hasil tes pada fitur tersebut, dan setiap daun (leaf node) mewakili label kelas (dalam klasifikasi) atau nilai target (dalam regresi).
- b. Proses Pembuatan Pohon:
 - Memilih fitur terbaik untuk memisahkan data menjadi dua atau lebih kelompok

yang lebih homogen

- Membagi data berdasarkan nilai fitur yang dipilih
- Mengulangi proses ini secara rekursif pada setiap kelompok data yang dihasilkan sampai memenuhi kondisi berhenti

c. Kondisi Berhenti:

- Semua data dalam node memiliki label yang sama
- Tidak ada fitur lagi yang tersedia untuk pemisahan
- Ukuran node telah mencapai batas minimum yang ditentukan

d. Kriteria Pemisahan:

- *Gini Impurity*: Mengukur seberapa sering elemen yang dipilih secara acak akan diklasifikasikan salah jika labelnya diambil secara acak dari distribusi di node tersebut
- *Entropy/Information Gain*: Mengukur perubahan ketidakpastian sebelum dan sesudah pemisahan data berdasarkan suatu fitur

3. Pemilihan Fitur Acak (*Random Feature Selection*)

a. Deskripsi: Pada setiap *node* dalam pohon keputusan, hanya subset acak dari fitur yang dipertimbangkan untuk pemilihan fitur terbaik yang akan digunakan untuk memisahkan data. Jumlah fitur yang dipertimbangkan biasanya adalah akar kuadrat dari jumlah total fitur atau logaritma basis dua dari jumlah total fitur.

b. Contoh: Jika terdapat 16 fitur dalam dataset, maka pada setiap node hanya akan dipertimbangkan 4 fitur secara acak untuk pemisahan terbaik.

c. Fungsi:

- Mengurangi korelasi antar pohon dalam hutan, sehingga hasil prediksi menjadi lebih tidak bergantung pada satu atau beberapa fitur tertentu
- Meningkatkan keragaman antar pohon, yang berkontribusi pada peningkatan performa keseluruhan model
- Mengurangi risiko *overfitting* dengan mencegah pohon-pohon dalam hutan menjadi terlalu mirip satu sama lain

4. Agregasi (*Voting/Averaging*)

a. Deskripsi: Proses menggabungkan hasil prediksi dari semua pohon keputusan dalam hutan untuk menghasilkan prediksi akhir.

b. Untuk Klasifikasi:

- Digunakan metode *voting* mayoritas, di mana setiap pohon memberikan suara

untuk salah satu kelas, dan kelas dengan suara terbanyak menjadi prediksi akhir

- Bisa juga digunakan *voting* berbobot, di mana setiap pohon memiliki bobot berdasarkan akurasinya

c. Untuk Regresi:

- Digunakan metode rata-rata, di mana prediksi akhir adalah rata-rata dari prediksi semua pohon keputusan dalam hutan
- Bisa juga digunakan rata-rata berbobot untuk meningkatkan akurasi

d. Fungsi:

- Menghasilkan prediksi akhir yang lebih stabil dan akurat dengan mengurangi variabilitas yang disebabkan oleh masing-masing pohon individual
- Mengurangi risiko kesalahan yang mungkin dilakukan oleh satu atau beberapa pohon tertentu

5. Validasi Model dengan *Out-of-Bag* (OOB)

- a. Deskripsi: Karena setiap pohon keputusan dilatih dengan bootstrap sample yang berbeda, terdapat sebagian data yang tidak termasuk dalam bootstrap sample untuk setiap pohon (disebut sebagai *out-of-bag* data). Data ini dapat digunakan untuk mengevaluasi performa model tanpa perlu membagi dataset menjadi data latih dan data uji secara terpisah.

b. Cara Kerja:

- Setiap data yang tidak termasuk dalam bootstrap sample untuk suatu pohon digunakan untuk menguji pohon tersebut
- Hasil prediksi dari semua pohon yang tidak menggunakan data tersebut digabungkan untuk menghitung OOB error

c. Fungsi:

- Memberikan estimasi yang akurat tentang performa model tanpa perlu melakukan cross-validation secara terpisah
- Menghemat waktu dan sumber daya karena tidak perlu membagi dataset atau melakukan pelatihan ulang model untuk evaluasi

II3.3 Keunggulan dan Keterbatasan Random Forest

1. Keunggulan *Random Forest*

- a. Tinggi Akurasi: *Random Forest* umumnya menghasilkan akurasi yang tinggi dalam tugas klasifikasi dan regresi, terutama dibandingkan dengan algoritma pembelajaran mesin tunggal lainnya.

- b. Tahan Terhadap *Overfitting* : Karena menggunakan banyak pohon keputusan dan metode *bagging*, *Random Forest* kurang rentan terhadap *overfitting* dibandingkan dengan pohon keputusan tunggal.
 - c. Dapat Menangani Data yang Tidak Seimbang: Dapat dioptimalkan untuk menangani dataset dengan distribusi kelas yang tidak seimbang, seperti dalam kasus deteksi kecurangan di mana jumlah kasus curang jauh lebih sedikit daripada kasus normal.
 - d. Dapat Menangani Berbagai Jenis Fitur: Dapat menangani fitur numerik dan kategorikal tanpa memerlukan pra-pemrosesan yang ekstensif.
 - e. Memberikan Informasi Penting Fitur: Dapat menghitung seberapa penting setiap fitur dalam proses klasifikasi atau regresi, yang berguna untuk pemahaman data dan pemilihan fitur.
 - f. Tahan Terhadap *Noise* dan Data yang Hilang: Kurang sensitif terhadap noise dalam data dan dapat menangani data yang hilang dengan baik.
 - g. Mudah untuk Dikalibrasi: Dapat menghasilkan probabilitas kelas untuk tugas klasifikasi, yang dapat dikalibrasi untuk meningkatkan akurasi prediksi probabilitas.
2. Keterbatasan *Random Forest*
- a. Kompleksitas Komputasi yang Tinggi: Membutuhkan sumber daya komputasi yang lebih besar dan waktu pelatihan yang lebih lama dibandingkan dengan algoritma tunggal seperti pohon keputusan atau regresi logistik, terutama ketika jumlah pohon yang digunakan banyak.
 - b. Sulit untuk Diterjemahkan: Meskipun dapat memberikan informasi tentang pentingnya fitur, *Random Forest* dianggap sebagai model "*black box*" yang sulit untuk dijelaskan secara rinci mengapa suatu keputusan prediksi dibuat.
 - c. Kurang Efektif untuk Data Berukuran Sangat Besar: Meskipun dapat menangani data yang besar, namun untuk dataset dengan ukuran sangat besar (misalnya, jutaan baris), *Random Forest* mungkin kurang efisien dibandingkan dengan algoritma lain seperti *Gradient Boosting* atau *Neural Network*.
 - d. Rentan Terhadap Fitur yang Tidak Relevan: Jika terdapat banyak fitur yang tidak relevan dalam dataset, *Random Forest* mungkin masih dapat menghasilkan prediksi yang baik, tetapi dapat meningkatkan kompleksitas model dan waktu pelatihan.
 - e. Sensitif terhadap Hyperparameter: Performa model sangat bergantung pada pengaturan hyperparameter seperti jumlah pohon, kedalaman maksimum pohon, dan jumlah fitur yang dipertimbangkan pada setiap *node*.

Dalam konteks deteksi kecurangan tagihan listrik, keunggulan *Random Forest* seperti kemampuan menangani data tidak seimbang, memberikan informasi pentingnya fitur, dan tahan terhadap overfitting menjadikannya pilihan yang sangat cocok. Namun, keterbatasannya dalam menangani pola kompleks dan tidak terstruktur dapat diatasi dengan menggabungkannya dengan algoritma lain seperti CNN

II.4 *Convolutional Neural Network (CNN)*

II.4.1 Konsep Dasar CNN

Convolutional Neural Network (CNN) adalah jenis jaringan saraf tiruan yang khusus dirancang untuk mengolah data yang memiliki *grid-like topology*, seperti gambar (dengan struktur *grid piksel*) atau data deret waktu (dengan struktur grid waktu) (A.S.S. Pratama et al., 2022). CNN pertama kali diperkenalkan pada tahun 1980-an, tetapi baru mendapatkan popularitas yang luas dalam beberapa tahun terakhir berkat kemajuan dalam kekuatan komputasi dan ketersediaan data yang besar.

Dalam konteks deteksi kecurangan tagihan listrik, data konsumsi listrik dapat dianggap sebagai data deret waktu atau dapat dikonversi menjadi bentuk visual seperti grafik konsumsi bulanan atau harian. CNN mampu mengenali pola-pola kompleks dalam data ini yang mungkin tidak terlihat oleh algoritma pembelajaran mesin konvensional.

Konsep dasar kerja CNN adalah mengikuti cara kerja sistem visual manusia, di mana sel-sel saraf di otak bereaksi terhadap wilayah tertentu dari bidang pandang (disebut sebagai *receptive fields*). Pada CNN, ini diwujudkan melalui penggunaan lapisan konvolusi yang menerapkan filter atau kernel pada *input* data untuk mendeteksi fitur-fitur lokal seperti tepi, tekstur, atau pola tertentu. Perbedaan utama antara CNN dan jaringan saraf tiruan konvensional (*feedforward neural network*) adalah:

1. *Shared Weights*: Filter yang digunakan dalam lapisan konvolusi memiliki bobot yang sama di seluruh area *input*, sehingga mengurangi jumlah parameter yang perlu dilatih dan memungkinkan model untuk mengenali pola yang sama di berbagai lokasi dalam input.
2. *Spatial Hierarchy of Features*: CNN membangun hierarki fitur dari fitur sederhana seperti tepi dan sudut pada lapisan awal menjadi fitur yang lebih kompleks dan abstrak pada lapisan yang lebih dalam.
3. *Pooling*: Menggunakan lapisan pooling untuk mengurangi dimensi data dan meningkatkan invariansi terhadap pergeseran atau perubahan skala dalam *input*.

II.4.2 Arsitektur dan Komponen CNN

Menurut A.S.S. Pratama et al. (2022), arsitektur CNN yang khas terdiri dari beberapa jenis lapisan yang bekerja bersama untuk memproses *input* data dan menghasilkan *output* prediksi. Berikut adalah komponen utama dari arsitektur CNN:

1. Lapisan *Input* (*Input Layer*)
 - a. Deskripsi: Lapisan pertama dari CNN yang menerima input data dalam bentuk grid. Dalam konteks deteksi kecurangan tagihan listrik, *input* dapat berupa:
 - b. Data deret waktu konsumsi listrik yang dikonversi menjadi matriks (misalnya, konsumsi harian selama beberapa bulan yang diatur dalam bentuk *grid*)
 - c. Grafik visual dari konsumsi listrik yang diubah menjadi bentuk citra digital
 - d. Fitur-fitur numerik yang diatur dalam bentuk matriks untuk memungkinkan operasi konvolusi
 - e. Ukuran *Input*: Bergantung pada jenis dan struktur data yang digunakan. Contohnya, jika menggunakan data konsumsi harian selama 30 hari untuk 12 bulan terakhir, maka ukuran input dapat berupa matriks 12x30.
2. Lapisan Konvolusi (*Convolutional Layer*)
 - a. Deskripsi: Lapisan inti dari CNN yang bertanggung jawab untuk mengekstraksi fitur dari input data. Pada lapisan ini, beberapa filter atau kernel diterapkan pada input data melalui operasi konvolusi.
 - b. Operasi Konvolusi:
 - Setiap filter bergerak (*convolve*) di seluruh input data dengan langkah tertentu yang disebut *stride*
 - Pada setiap posisi, dilakukan operasi perkalian titik demi titik antara filter dan bagian input yang sesuai, kemudian dijumlahkan untuk menghasilkan satu nilai pada peta fitur (*feature map*)
 - Rumus operasi konvolusi adalah:
$$(I * K)(i,j) = \sum_m \sum_n I(i-m, j-n) K(m,n)$$
Dimana I adalah matriks *input*, K adalah filter atau kernel, dan $(I * K)(i,j)$ adalah nilai pada posisi (i,j) pada peta fitur.
 - c. Jenis Filter:
 - Filter Sederhana: Mendeteksi fitur sederhana seperti tepi horizontal, tepi

vertikal, atau tekstur tertentu

- Filter Kompleks: Pada lapisan yang lebih dalam, filter dapat mendeteksi fitur yang lebih kompleks seperti pola konsumsi tertentu atau anomali dalam data
- Jumlah Filter: Biasanya meningkat pada lapisan yang lebih dalam, mulai dari 32 atau 64 filter pada lapisan awal hingga 128 atau 256 filter pada lapisan yang lebih dalam.

3. Lapisan Aktivasi (*Activation Layer*)

a. Deskripsi: Lapisan yang diterapkan setelah lapisan konvolusi untuk memperkenalkan *non-linearitas* ke dalam model. Tanpa fungsi aktivasi, CNN hanya akan menjadi model linier yang memiliki kapasitas pembelajaran yang terbatas.

b. Fungsi Aktivasi yang Umum Digunakan:

- *ReLU (Rectified Linear Unit)*:

$$f(x) = \max(0, x)$$

Mengubah semua nilai negatif menjadi nol dan mempertahankan nilai positif seperti adanya. Keunggulan ReLU adalah menghindari masalah *vanishing gradient* dan mempercepat proses pelatihan.

- *Sigmoid*:

$$f(x) = \frac{1}{1 + e^{-x}}$$

Mengubah nilai input menjadi rentang antara 0 dan 1, cocok untuk tugas klasifikasi biner pada lapisan *output*.

- *Softmax*:

$$f(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

Mengubah nilai *input* menjadi distribusi probabilitas yang jumlahnya sama dengan 1, cocok untuk tugas klasifikasi multi-kelas.

- *Tanh (Hyperbolic Tangent)*

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

Mengubah nilai *input* menjadi rentang antara -1 dan 1.

4. Lapisan *Pooling* (*Pooling Layer*)

a. Deskripsi: Lapisan yang bertujuan untuk mengurangi dimensi peta fitur sambil mempertahankan informasi penting. Hal ini membantu mengurangi jumlah parameter

dan komputasi dalam jaringan, serta meningkatkan invariansi terhadap pergeseran atau perubahan skala dalam input.

b. Jenis Operasi *Pooling*:

- *Max Pooling*: Mengambil nilai maksimum dalam setiap *patch* dari peta fitur.

Rumusnya adalah:

$$P(i,j) = \max\left\{ I(i \times s + m, j \times s + n) \right\}$$

Dimana s adalah ukuran langkah (stride) dan m,n adalah ukuran

filter *pooling*.

- *Average Pooling*: Mengambil nilai rata-rata dalam setiap *patch* dari peta fitur.
 - *Sum Polling* : Mengambil jumlah total nilai dalam setiap *patch* dari peta fitur.
 - *Global Average Pooling*: Mengambil rata-rata dari seluruh peta fitur untuk setiap filter, sering digunakan pada lapisan akhir sebelum lapisan output untuk mengurangi jumlah parameter.
- c. Ukuran *Pooling*: Biasanya menggunakan filter pooling berukuran 2x2 atau 3x3 dengan stride yang sama dengan ukuran filter, sehingga mengurangi dimensi peta fitur menjadi setengah atau sepertiga dari ukuran awal.
5. Lapisan Fully Connected (*Fully Connected Layer*)
6. Lapisan *Dropout* (*Dropout Layer*)
7. Lapisan Normalisasi *Batch* (*Batch Normalization Layer*)
8. H. Lapisan *Output* (*Output Layer*)

II.4.3 Keunggulan dan Keterbatasan CNN

1. Keunggulan CNN
 - a. Kemampuan Mengekstraksi Fitur Otomatis: CNN dapat mengekstraksi fitur penting dari data secara otomatis tanpa perlu definisi manual dari peneliti, yang sangat berguna untuk data dengan struktur kompleks seperti data deret waktu konsumsi listrik.
 - b. Efisien dalam Penggunaan Parameter: Menggunakan konsep *shared weights* dan *pooling* yang mengurangi jumlah parameter yang perlu dilatih, sehingga lebih efisien dibandingkan dengan jaringan saraf tiruan konvensional untuk data *grid-like*.
 - c. Invariansi terhadap Pergeseran dan Skala: Dapat mengenali pola yang sama meskipun terdapat pergeseran atau perubahan skala dalam input, yang berguna dalam mendeteksi pola kecurangan yang mungkin muncul dalam bentuk yang berbeda.
 - d. Kemampuan Menangani Data Besar: Dapat diimplementasikan pada perangkat keras dengan GPU untuk mempercepat proses pelatihan pada dataset yang besar.

- e. Hierarki Fitur yang Kuat: Membangun hierarki fitur dari yang sederhana hingga yang kompleks, yang memungkinkan model untuk memahami pola yang lebih dalam dalam data.
 - f. Dapat Digabungkan dengan Algoritma Lain: Mudah untuk digabungkan dengan algoritma pembelajaran mesin lain seperti *Random Forest* untuk meningkatkan performa deteksi.
2. Keterbatasan CNN
- a. Membutuhkan Data yang Banyak: Memerlukan jumlah data yang cukup besar untuk dilatih dengan baik, terutama jika arsitektur model yang digunakan kompleks.
 - b. Kompleksitas Komputasi yang Tinggi: Membutuhkan sumber daya komputasi yang kuat (terutama GPU) untuk pelatihan model yang besar, yang dapat meningkatkan biaya dan waktu pelatihan.
 - c. Sulit untuk Diterjemahkan: Seperti halnya jaringan saraf tiruan lainnya, CNN dianggap sebagai model "black box" yang sulit untuk dijelaskan mengapa suatu keputusan prediksi dibuat.
 - d. Sensitif terhadap Hyperparameter: Performa model sangat bergantung pada pengaturan hyperparameter seperti ukuran filter, jumlah filter, ukuran pooling, dan learning rate.
 - e. Kurang Efektif untuk Data *Non-Grid*: Kurang efektif untuk data yang tidak memiliki struktur grid-like, seperti data tabular dengan fitur yang tidak memiliki hubungan spasial atau temporal.

Dalam konteks deteksi kecurangan tagihan listrik, keunggulan CNN dalam mengekstraksi fitur kompleks dari data deret waktu atau data visual membuatnya sangat cocok untuk mengidentifikasi pola kecurangan yang mungkin tidak terdeteksi oleh algoritma lain.

Keterbatasannya dalam hal interpretasi dan kebutuhan akan data yang banyak dapat diatasi dengan menggabungkannya dengan *Random Forest*, yang memiliki keunggulan dalam interpretasi dan menangani data tabular.

II.5 Kombinasi *Random Forest* Dan CNN

II.5.1 *Rationale* Kombinasi Dua Algoritma

Penggunaan kombinasi atau *hybrid* dari dua atau lebih algoritma pembelajaran mesin telah menjadi tren dalam penelitian terkini karena mampu menggabungkan keunggulan masing-masing algoritma dan mengatasi keterbatasannya (Faathin et al., 2019). Dalam kasus

deteksi kecurangan tagihan listrik, kombinasi *Random Forest* dan CNN dipertimbangkan karena memiliki keunggulan yang saling melengkapi:

1. Keunggulan yang Saling Melengkapi
 - a. Random Forest Kuat dalam Klasifikasi dan Interpretasi:
 - b. Sangat baik dalam menangani data tabular dengan fitur numerik dan kategorikal
 - c. Memberikan informasi tentang pentingnya setiap fitur, yang berguna untuk memahami faktor-faktor yang menyebabkan kecurangan
 - d. Lebih stabil dan kurang sensitif terhadap noise dalam data
 - e. Mudah untuk diimplementasikan dan dioptimalkan
 - f. CNN Kuat dalam Ekstraksi Fitur Kompleks:
 - Sangat baik dalam menangani data dengan struktur spasial atau temporal seperti data konsumsi listrik
 - Dapat mengekstraksi fitur-fitur tersembunyi yang tidak terlihat oleh algoritma konvensional
 - Mampu mengenali pola kompleks yang mungkin menunjukkan adanya kecurangan
 - Kuat dalam menangani variasi dalam data
2. Alasan Mengapa Kombinasi Ini Efektif untuk Deteksi Kecurangan
 - a. Pola Kecurangan yang Beragam: Kecurangan tagihan listrik dapat muncul dalam berbagai bentuk dan pola, sehingga membutuhkan pendekatan yang dapat menangani berbagai jenis pola tersebut.
 - b. Data yang Bersifat Heterogen: Data konsumsi listrik terdiri dari berbagai jenis fitur, mulai dari fitur numerik yang sederhana hingga pola temporal yang kompleks, sehingga membutuhkan kombinasi algoritma yang dapat menangani kedua jenis data tersebut.
 - c. Tantangan Ketidakseimbangan Data: Dataset kecurangan biasanya memiliki distribusi kelas yang tidak seimbang, dan kombinasi algoritma dapat membantu meningkatkan deteksi pada kelas minoritas (kasus curang).
 - d. Kebutuhan akan Interpretasi dan Akurasi: PLN membutuhkan tidak hanya akurasi deteksi yang tinggi tetapi juga pemahaman tentang mengapa suatu pelanggan dicurigai melakukan kecurangan, yang dapat diberikan oleh Random Forest setelah fitur-fitur kompleks diekstraksi oleh CNN.
 - e. Tinjauan Penelitian Terkait tentang Kombinasi Algoritma

Beberapa penelitian sebelumnya telah menunjukkan keefektifan penggunaan kombinasi Random Forest dan CNN dalam berbagai bidang aplikasi:

1. Penelitian oleh Faathin et al. (2019): Menggunakan kombinasi CNN dan *Random Forest* untuk klasifikasi citra medis, menghasilkan akurasi yang lebih tinggi dibandingkan dengan penggunaan algoritma tunggal.
2. Penelitian oleh Abdullah-Al Nahid et al. (2019): Menggunakan kombinasi CNN dan LSTM untuk deteksi kecurangan listrik, menunjukkan peningkatan akurasi yang signifikan dibandingkan dengan model tunggal.
3. Penelitian oleh Ali Saleh Al-rimy et al. (2021): Menggunakan *framework kombinasi feature engineering* dan *NGBoost* untuk deteksi kecurangan listrik, menghasilkan akurasi 93% dengan presisi 95%.
4. Penelitian oleh Ismini Psychoula et al. (2021): Menggunakan teknik *explainable AI* seperti *LIME* dan *SHAP* untuk menjelaskan hasil dari model kombinasi *Random Forest* dan *Neural Network* dalam deteksi kecurangan, meningkatkan transparansi model.

Dari penelitian-penelitian tersebut dapat disimpulkan bahwa kombinasi algoritma pembelajaran mesin dan *deep learning* seperti *Random Forest* dan CNN memiliki potensi besar untuk meningkatkan akurasi dan keandalan sistem deteksi kecurangan.

II.5.2 Metode Penggabungan Algoritma

Ada beberapa metode yang dapat digunakan untuk menggabungkan *Random Forest* dan CNN dalam satu sistem deteksi. Dalam penelitian ini, akan digunakan metode *feature fusion* atau *ensemble stacking*, di mana CNN digunakan untuk mengekstraksi fitur kompleks dari data, dan kemudian fitur-fitur tersebut digunakan sebagai input untuk Random Forest dalam proses klasifikasi.

Berikut adalah langkah-langkah implementasi kombinasi kedua algoritma:

1. Persiapan Data
 - a. Data *Input*: Data konsumsi listrik pelanggan yang telah melalui tahap pra-pemrosesan, termasuk data numerik dan data deret waktu.
 - b. Pembagian Data: Dataset dibagi menjadi tiga bagian: data latih (60%), data validasi (20%), dan data uji (20%).
 - c. Konversi Data: Data deret waktu konsumsi listrik dikonversi menjadi bentuk matriks atau citra untuk dapat diolah oleh CNN. Misalnya, konsumsi harian selama 30 hari dapat diatur menjadi matriks 5x6 atau dikonversi menjadi grafik visual yang diubah menjadi citra digital dengan resolusi tertentu.
2. Langkah 2: Ekstraksi Fitur dengan CNN

- a. Pelatihan CNN: Model CNN dilatih menggunakan data latih untuk mengekstraksi fitur-fitur penting dari data. Arsitektur CNN yang digunakan dirancang khusus untuk menangani data konsumsi listrik.
- b. Ekstraksi Fitur: Setelah model CNN dilatih dengan baik, lapisan output dari lapisan konvolusi terakhir atau lapisan pooling terakhir digunakan sebagai fitur-fitur yang telah diekstraksi. Fitur-fitur ini mencakup informasi tentang pola-pola kompleks dalam data konsumsi listrik.
- c. Ratakan Fitur: Fitur-fitur yang telah diekstraksi oleh CNN diratakan menjadi vektor satu dimensi untuk dapat digunakan sebagai input untuk Random Forest.

3. Penggabungan Fitur

- a. Fitur dari CNN: Fitur-fitur yang telah diekstraksi oleh CNN yang mencakup informasi tentang pola kompleks dalam data.
- b. Fitur Konvensional: Fitur-fitur konvensional dari data konsumsi listrik seperti rata-rata konsumsi, deviasi standar, rasio tagihan per kWh, dan karakteristik pelanggan lainnya.
- c. Penggabungan: Kedua jenis fitur tersebut digabungkan menjadi satu vektor fitur yang akan digunakan sebagai *input* untuk *Random Forest*. Hal ini memungkinkan *Random Forest* untuk menggunakan informasi dari kedua sumber fitur.

4. Klasifikasi dengan *Random Forest*

- a. Pelatihan Random Forest: Model Random Forest dilatih menggunakan vektor fitur gabungan dari langkah sebelumnya. Hyperparameter Random Forest dioptimalkan menggunakan data validasi untuk mendapatkan performa terbaik.
- b. Prediksi: Setelah model *Random Forest* dilatih, digunakan untuk melakukan prediksi pada data uji untuk menentukan apakah suatu pelanggan melakukan kecurangan atau tidak.
- c. Penggabungan Prediksi: Jika diperlukan, dapat digunakan metode voting atau stacking dengan model lain untuk meningkatkan akurasi prediksi akhir.

5. Evaluasi dan Validasi

- a. Evaluasi Model: Kinerja model kombinasi dievaluasi menggunakan metrik evaluasi seperti akurasi, presisi, recall, F1-score, dan kurva ROC-AUC.
- b. Perbandingan dengan Model Tunggal: Performa model kombinasi

dibandingkan dengan performa model Random Forest tunggal dan CNN tunggal untuk menunjukkan keunggulan dari kombinasi algoritma.

- c. Validasi dengan Data Aktual: Model yang telah dikembangkan divalidasi menggunakan data aktual dari PLN ULP Telda untuk memastikan bahwa model dapat bekerja dengan baik dalam kondisi nyata.

6. Metode Penggabungan Lain yang Dapat Dipertimbangkan

- a. *Ensemble Voting*: Kedua model dilatih secara terpisah, dan hasil prediksi dari masing-masing model digabungkan menggunakan metode voting mayoritas atau *voting* berbobot.
- b. *Ensemble Stacking*: Menggunakan model lain (seperti *Logistic Regression* atau *XGBoost*) sebagai meta-learner yang mengambil hasil prediksi dari *Random Forest* dan CNN sebagai *input* untuk menghasilkan prediksi akhir.
- c. *Feature Selection* dengan *Random Forest*: Menggunakan *Random Forest* untuk memilih fitur-fitur yang paling penting dari data, kemudian menggunakan fitur-fitur tersebut sebagai *input* untuk CNN.
- d. Pembedaan Tingkat Abstraksi: Menggunakan CNN untuk menangani fitur-fitur pada tingkat abstraksi rendah dan *Random Forest* untuk menangani fitur-fitur pada tingkat abstraksi tinggi, kemudian menggabungkan hasilnya.

Dalam penelitian ini, *metode feature fusion* dipilih karena relatif mudah untuk diimplementasikan dan telah terbukti efektif dalam penelitian sebelumnya. Selain itu, metode ini memungkinkan penggunaan keunggulan kedua algoritma secara optimal, dengan CNN menangani ekstraksi fitur kompleks dan *Random Forest* menangani klasifikasi dengan interpretasi yang baik.

II.6 Penelitian Terkait

II.6.1 Penelitian sebelumnya tentang Deteksi Kecurangan Listrik

Tabel II. 1 Penelitian Sebelumnya

Nama peneliti dan Judul Penelitian	Metode Penelitian	Hasil Penelitian

Abdullah-Al Nahid et al. (2019) : - Penggunaan Model CNN-LSTM untuk Deteksi Kecurangan Listrik	Mengembangkan sistem deteksi kecurangan listrik menggunakan kombinasi Convolutional Neural Network (CNN) dan Long Short-Term Memory (LSTM). CNN digunakan untuk ekstraksi fitur otomatis dari data konsumsi listrik, sementara LSTM digunakan untuk menangani sifat deret waktu dari data tersebut.	Model ini mampu menangani masalah ketidakseimbangan kelas dengan menghasilkan data sintetis, yang membantu meningkatkan akurasi deteksi kecurangan baik untuk pengguna normal maupun pengguna yang melakukan pencurian listrik. Akurasi yang dicapai adalah sekitar 89% dengan recall 91%.
Francesco Piccialli et al. (2021) : Energy Theft Detection in Smart Grids: Taxonomy, Comparative Analysis, Challenges, and Future Research Directions	Menggunakan pendekatan taxonomy dan analisis perbandingan terhadap berbagai metode deteksi kecurangan listrik yang ada. Penelitian ini juga membahas tentang penggunaan algoritma LSTM untuk menangani sifat deret waktu dari data konsumsi listrik.	Hasilnya menunjukkan bahwa pendekatan berbasis data dan jaringan memiliki potensi besar, namun masih membutuhkan pengembangan lebih lanjut dalam hal manajemen pengetahuan dan pencegahan. Penelitian ini juga mengidentifikasi tantangan utama dalam deteksi kecurangan seperti ketidakseimbangan data, keragaman pola

		kecurangan, dan kebutuhan akan interpretasi model
Abdullah-Al Nahid (2022) : Energy Theft Detection in Smart Grids: A CNN-AdaBoost Hybrid Approach	Menggunakan kombinasi model Convolutional Neural Network (CNN) dan AdaBoost untuk deteksi kecurangan listrik. Model ini memanfaatkan kekuatan CNN dalam mengenali pola visual dari data konsumsi energi dan menggunakan AdaBoost untuk meningkatkan kinerja klasifikasi.	Model hybrid CNN-AdaBoost menunjukkan peningkatan signifikan dalam akurasi deteksi kecurangan dibandingkan dengan penggunaan CNN atau AdaBoost secara individu. Akurasi yang dicapai adalah sekitar 92% dengan presisi 93%. Penelitian ini juga menangani masalah ketidakseimbangan kelas dengan menggunakan teknik oversampling.
Ali Saleh Al-rimy et al. (2021) : - A Novel Feature-Engineered–NGBoost Machine Learning Framework for Fraud Detection in Electric Power Consumption Data	Menggunakan framework pembelajaran mesin NGBoost yang melibatkan teknik pre-processing data, feature engineering, dan evaluasi model. Teknik imputasi data berbasis Random Forest digunakan untuk mengisi data yang hilang, dan algoritma MWMOTE	Framework ini mencapai akurasi 93%, recall 91%, dan presisi 95%, yang lebih baik dibandingkan model-model pesaing lainnya, menunjukkan efektivitas dalam mendeteksi kecurangan listrik.

	digunakan untuk mengatasi ketidakseimbangan kelas	
Zheng et al. (2022) : Electricity Theft Detection Using Machine Learning Techniques to Secure Smart Grid	Menggabungkan penggunaan Long Short-Term Memory (LSTM) dan Gaussian Mixture Model (GMM) untuk deteksi anomali secara real-time. Metode ini dirancang untuk mendeteksi anomali dalam pola konsumsi listrik yang menunjukkan kemungkinan pencurian listrik.	Hasil menunjukkan bahwa pendekatan kombinasi LSTM-GMM efektif dalam mendeteksi pola anomali dengan tingkat akurasi yang tinggi (sekitar 90%), meningkatkan keandalan deteksi kecurangan dalam sistem jaringan pintar.
Ismini Psychoula et al. (2021) : - Explainable Machine Learning for Fraud Detection	Menggunakan teknik explainable AI seperti LIME (Local Interpretable Model-Agnostic Explanations) dan SHAP (SHapley Additive exPlanations) untuk menjelaskan hasil dari model pembelajaran mesin yang digunakan dalam deteksi kecurangan, termasuk Random Forest dan Neural Networks.	Teknik LIME dan SHAP membantu dalam memberikan penjelasan yang dapat dimengerti tentang fitur mana yang paling berpengaruh terhadap prediksi model. Ini membantu para analis memahami mengapa model mendeteksi transaksi tertentu sebagai kecurangan, meningkatkan transparansi dan

		kepercayaan terhadap sistem deteksi.
Syahputra & Afnan (2020): Pendeteksian Fraud: Peran Big Data dan Audit Forensik	Membahas peran analitik big data dan audit forensik dalam mendeteksi kecurangan. Teknik yang digunakan termasuk analisis data skala besar dan metode audit forensik yang komprehensif.	integrasi big data analytics dan audit forensik dapat memberikan pendekatan yang lebih efektif dalam mendeteksi dan mencegah kecurangan, terutama dalam konteks lembaga keuangan dan pemerintahan. Penelitian ini juga menunjukkan bahwa kombinasi antara teknologi dan keahlian manusia memberikan hasil yang paling optimal.

II.6.2 Perbandingan dengan Penelitian terdahulu

Penelitian ini memiliki beberapa perbedaan dan keunggulan dibandingkan dengan penelitian terdahulu tentang deteksi kecurangan tagihan listrik:

Tabel II. 2 Perbandingan Dengan Penelitian Terdahulu

Aspek Perbandingan	Penelitian Terdahulu	Penelitian Ini
Algoritma yang digunakan	Biasanya menggunakan satu jenis algoritma atau kombinasi algoritma yang berbeda (misalnya	Menggunakan kombinasi Random Forest dan CNN, yang memiliki keunggulan saling melengkapi dalam menangani data tabular dan pola kompleks

	CNN-LSTM, CNN-AdaBoost)	
Fokus Wilayah	Umumnya bersifat umum atau fokus pada wilayah lain	Khusus fokus pada wilayah PLN ULP Telda, sehingga hasil lebih relevan dengan kondisi lokal
Dataset yang digunakan	Biasanya menggunakan dataset publik atau dataset yang tidak spesifik	Menggunakan dataset dari PLN ULP Telda sendiri (disertai validasi dengan dataset publik untuk memastikan generalisasi)
Metode Penggabungan	Menggunakan metode stacking atau voting yang sederhana	Menggunakan metode feature fusion yang menggabungkan fitur dari CNN dengan fitur konvensional untuk meningkatkan performa
Interpretasi Model	Sebagian besar penelitian tidak fokus pada interpretasi model	Menyertakan analisis pentingnya fitur dan teknik explainable AI untuk menjelaskan hasil prediksi model
Evaluasi	Biasanya hanya mengevaluasi berdasarkan metrik klasifikasi	Mengevaluasi tidak hanya berdasarkan metrik klasifikasi tetapi juga berdasarkan implikasi praktis dan biaya-benefit untuk PLN ULP Telda