

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Yang Relevan

Sebagai dasar perbandingan dan penguatan dalam pelaksanaan penelitian ini, bagian ini menyajikan sejumlah studi terdahulu yang memiliki keterkaitan erat dengan topik yang diangkat. Tinjauan terhadap penelitian sebelumnya memberikan landasan yang kuat bagi peneliti untuk mengevaluasi berbagai temuan yang telah diperoleh, sekaligus mengidentifikasi celah atau keterbatasan yang masih ada. Dengan demikian, studi-studi ini tidak hanya menjadi rujukan, tetapi juga berperan penting dalam menyempurnakan pendekatan dan ruang lingkup penelitian yang sedang dilakukan. Berikut disajikan beberapa hasil penelitian relevan yang mendukung pengembangan kajian ini.

Penerapan Data mining Menggunakan Algoritma Apriori untuk Menentukan Pola Golongan Penyandang Masalah Kesejahteraan Sosial (Darmawan et al., 2022). Penelitian ini menerapkan algoritma *Apriori* untuk menemukan pola asosiasi pada data Penyandang Masalah Kesejahteraan Sosial (PMKS). Dengan menggunakan metode *association rule*, penelitian ini berhasil mengidentifikasi *frequent itemset* dari data PMKS periode 2017–2020, menghasilkan 3 aturan dengan nilai *minimum Support* 15% dan *confidence* hingga 100%. Pengujian dilakukan menggunakan RapidMiner dan perhitungan manual, meskipun hasilnya menunjukkan tingkat akurasi relatif rendah (37,5%) dibandingkan hasil aplikasi. Studi ini relevan dengan penelitian saya karena sama-sama memanfaatkan algoritma *Apriori* untuk mengidentifikasi pola hubungan dari data *historis*, namun pada penelitian saya fokusnya pada pola gangguan berulang jaringan PLN di wilayah rawan gangguan.

Penerapan Algoritma Apriori untuk Optimasi Strategi Penjualan Berdasarkan Analisis Pola Pembelian di Torsa Café (Ibezato Zalukhu et al., 2024). Penelitian ini bertujuan menganalisis pola pembelian konsumen di Torsa Café menggunakan algoritma *Apriori* untuk menemukan keterkaitan antar produk yang sering dibeli bersama. Data transaksi bulan Oktober 2024 dianalisis untuk menghitung nilai *Support* dan *confidence*, menemukan kombinasi menu dengan

Support >30% seperti “Sanger Espresso” dan “Avocado Cappuccino Torsa”, serta kombinasi Kopi Tancap dengan beberapa minuman lainnya. Hasilnya digunakan untuk strategi penempatan produk dan pengelolaan stok. Pendekatan ini relevan dengan penelitian saya karena memanfaatkan pola keterkaitan (*association rule*) untuk mendukung keputusan operasional, yang dapat diadaptasi untuk pola kemunculan gangguan pada jaringan listrik.

Penerapan Algoritma Apriori pada Transaksi Penjualan untuk Rekomendasi Menu Makanan dan Minuman (Merliani et al., 2022). Penelitian ini menganalisis data penjualan di Warung Tenda menggunakan algoritma *Apriori* untuk memberikan rekomendasi kombinasi menu. Dari hasil perhitungan, diperoleh kombinasi “Es Teh Manis” dan “Mendoan” dengan *Support* 50% dan *confidence* 76% sebagai rekomendasi utama. Analisis dilakukan menggunakan data transaksi, menghitung *Support* dan *confidence* untuk menentukan menu yang sering dipesan bersama. Studi ini relevan karena menunjukkan penerapan *Apriori* pada data transaksi untuk menemukan pola kemunculan bersamaan, konsep yang dapat diadaptasi untuk mengidentifikasi kombinasi jenis gangguan jaringan yang sering muncul bersamaan.

Analisis Penerapan Data mining dalam Penentuan Tata Letak Barang Menggunakan Algoritma Apriori dan FP-Growth (Harahap et al., 2022). Penelitian ini membandingkan kinerja algoritma *Apriori* dan *FP-Growth* dalam menentukan pola asosiasi untuk penataan tata letak barang di Swalayan Mohare. Dengan *minimum Support* 20% dan *confidence* 70%, *Apriori* menghasilkan 10 aturan, sementara *FP-Growth* menghasilkan 78 aturan dengan akurasi lebih tinggi. Studi ini menekankan peran pola asosiasi dalam optimasi strategi penempatan item. Relevansinya dengan penelitian saya terletak pada penerapan *Apriori* untuk menemukan keterkaitan antar item (atau jenis gangguan), yang kemudian dapat digunakan sebagai dasar strategi penanganan.

Penerapan Data mining Algoritma Apriori dalam Menentukan Stok Bahan Baku pada Restoran Nelayan Menggunakan Metode Association rule (-E & Tamba, 2022). Penelitian ini menerapkan algoritma *Apriori* pada data pengambilan bahan baku untuk menentukan stok optimal di Restoran Nelayan. Analisis menghasilkan 22 aturan asosiasi, misalnya: jika mengambil Ikan Gembung,

kemungkinan besar juga mengambil Ikan Nila (80%) atau Ikan Kerapu (100%). Penerapan *association rule* ini mempermudah pengelolaan stok dan efisiensi operasional. Pendekatan ini sejalan dengan penelitian saya yang bertujuan mengidentifikasi hubungan antar jenis gangguan, sehingga dapat memprediksi kemungkinan gangguan lain yang akan terjadi bersamaan.

Perbandingan Akurasi Algoritma Apriori dan FP-Growth untuk Mengetahui Pola Penyebab Gangguan Indihome Area Wonosobo (Irgi Sena Fajar Lukerti et al., 2024). Penelitian ini bertujuan mengidentifikasi pola penyebab gangguan Indihome area Wonosobo guna menentukan langkah *preventif* yang tepat. PT Telkom selama ini memiliki 458 data gangguan tahun 2022 namun belum dianalisis secara mendalam. Penelitian ini menggunakan metode KDD (Knowledge Discovery in Database) dengan membandingkan algoritma *Apriori* dan *FP-Growth* melalui RapidMiner (min *Support* 5%, *confidence* 50%). Hasilnya *FP-Growth* lebih akurat (100% *rule valid*) dibanding *Apriori* (17.65%), dengan pola utama gangguan kabel luar putus (offline) dan konfigurasi hilang (online).

Implementasi Algoritma Apriori dalam Menentukan Pola Transaksi Penjualan (Ramadhan & Sari, 2024). Penelitian ini bertujuan menganalisis pola transaksi penjualan The Gade Coffee and Gold cabang Kebayoran Baru untuk pengambilan keputusan strategi bisnis. Kantor cabang selama ini hanya menggunakan data transaksi untuk laporan penjualan tanpa analisis mendalam. Penelitian ini menggunakan metode kuantitatif dengan algoritma *Apriori* melalui *software RapidMiner* untuk menganalisis 3 bulan data transaksi (Januari-Maret 2023). Hasilnya ditemukan 6 aturan asosiasi terbaik termasuk pola pembelian Van Lenning - Iced dengan *confidence* 87.8%.

Korelasi Kemajuan Teknologi terhadap Kinerja Karyawan Menggunakan Metode Apriori, Studi Kasus : PLN Binjai (Alya Fadillah et al., 2023). Penelitian ini bertujuan menganalisis pengaruh teknologi terhadap kinerja karyawan PLN Kota Binjai. Perusahaan selama ini belum memetakan hubungan antara penguasaan teknologi dengan performa kerja. Penelitian ini menggunakan metode kuantitatif dengan algoritma *Apriori* melalui *RapidMiner* (min *Support* 10%, *confidence* 50%) pada 230 data karyawan. Hasilnya ditemukan 44 aturan asosiasi dengan pola terkuat:

penggunaan Ms Office, aplikasi mobile PLN, dan perangkat kantor lengkap berkorelasi dengan kinerja sangat baik (*Support* 30%, *confidence* 96%).

Implementasi Data mining untuk Menganalisis Pola Peminjaman Buku Perpustakaan Dengan Menggunakan Metode Apriori (Yuliyanto & Maedjaja, 2023). Penelitian ini bertujuan memberikan rekomendasi penataan buku di perpustakaan SMK Tunas Harapan Pati berdasarkan pola peminjaman. Perpustakaan selama ini menyusun buku tanpa analisis pola peminjaman yang sistematis. Penelitian ini menggunakan metode deskriptif kuantitatif dengan algoritma *Apriori* melalui *Python* (min *Support* 10%, *confidence* 20%) pada 167 data transaksi. Hasilnya ditemukan pola peminjaman bersamaan buku fiksi dan seni dekorasi dengan *Support* 25.15% serta *confidence* tertinggi 86.96% untuk kombinasi kerajinan tangan dan fiksi.

Implementasi Algoritma Apriori untuk Menentukan Pola Transaksi Penjualan Berbasis Web (Ali Ridla et al., 2023). Penelitian ini bertujuan mengoptimalkan strategi penjualan toko roti berbasis analisis pola transaksi. Pemilik toko selama ini tidak memanfaatkan data *historis* transaksi untuk pengambilan keputusan. Penelitian ini menggunakan metode Waterfall dengan algoritma *Apriori* (min *Support* dan *confidence* 10%) pada dataset publik *Kaggle*. Hasilnya ditemukan 2 kombinasi *itemset* valid termasuk pembelian kopi dan roti (*confidence* 40%) serta medialuna dan roti (*confidence* 50%), yang dapat digunakan untuk strategi penempatan produk.

A Comparative Study on Data Collection Methods: Investigating Optimal Datasets for Data mining Analysis (Jatnika et al., 2024). Penelitian ini membandingkan dua metode pengumpulan data, yaitu *kuesioner* dan *web mining*, untuk menentukan metode yang paling efektif dalam menghasilkan data berkualitas tinggi untuk analisis *data mining* menggunakan algoritma *Support Vector Machine* (SVM) dan *NaiveBayes Classifier* (NBC). Hasil penelitian menunjukkan bahwa data dari *kuesioner* memberikan tingkat akurasi lebih tinggi (di atas 82% untuk SVM dan 81% untuk NBC) dibandingkan dengan *web mining*. Faktor kebersihan data dan minimnya simbol yang mengganggu proses klasifikasi menjadi alasan utama keunggulan *kuesioner*. Penelitian ini menekankan pentingnya pemilihan metode pengumpulan data yang tepat agar hasil analisis *data mining* dapat optimal.

Pendekatan ini relevan dengan penelitian saya karena menunjukkan bahwa kualitas data input sangat mempengaruhi akurasi hasil penerapan algoritma *Apriori* dalam menemukan pola asosiasi gangguan berulang pada jaringan listrik.

Metode Association rule untuk Analisis Citra CT Organ Pasien Kanker Ovarium (Kuswardani et al., 2011). Penelitian ini menerapkan metode *Association rule* dengan algoritma *Apriori* untuk menganalisis citra CT scan organ liver, ginjal, dan uterus pasien kanker ovarium. Prosesnya meliputi ekstraksi fitur ukuran dan intensitas organ, transformasi data numerik menjadi kategori, serta pembentukan tabel transaksi. Uji coba dilakukan pada 17 pasien, menghasilkan ribuan transaksi data. Hasilnya menunjukkan pola-pola asosiasi organ dengan tingkat *Support* dan *confidence* tertentu, misalnya ukuran uterus memiliki *Support* 55% dan hubungan ukuran *liver-uterus* memiliki *Support* 45%. Penerapan *Apriori* di sini membantu mengidentifikasi ciri-ciri organ yang berpotensi menjadi indikator penyakit. Konsep ini relevan dengan penelitian saya karena memperlihatkan bagaimana algoritma *Apriori* mampu mengungkap keterkaitan antar atribut dari data *historis*, yang dapat diadaptasi untuk mengidentifikasi hubungan antar jenis gangguan berulang pada jaringan PLN.

Faktor-Faktor yang Berpengaruh untuk Mengestimasi Waktu Kelulusan Mahasiswa dengan Pendekatan Algoritma C4.5 (Algoritma, 2024). Penelitian ini menggunakan algoritma *decision tree C4.5* untuk memprediksi ketepatan waktu kelulusan mahasiswa berdasarkan beberapa atribut seperti IPK, asal sekolah, keaktifan organisasi, keikutsertaan dalam Unit Kegiatan Mahasiswa (UKM), status bekerja, dan motivasi. Data diuji menggunakan *confusion matrix* dengan hasil akurasi lebih dari 70%. Proses pengolahan data meliputi pengumpulan data mahasiswa, transformasi atribut menjadi kategori, perhitungan *entropy* dan *gain*, hingga pembentukan pohon keputusan dan *rules*. Meskipun algoritma yang digunakan berbeda dengan *Apriori*, penelitian ini relevan karena memperlihatkan pentingnya pemilihan atribut dan pembersihan data sebelum proses *data mining*, yang juga menjadi langkah penting dalam analisis pola asosiasi gangguan berulang jaringan listrik.

Implementation of ECLAT Algorithm to Determine Product Purchasing Pattern at Coffee Shop (Juliandiny et al., 2022). Penelitian ini menerapkan

algoritma ECLAT untuk menemukan pola pembelian produk di sebuah *coffee shop* berdasarkan data transaksi empat bulan terakhir. Dengan memanfaatkan metode *association rule*, penelitian ini mengidentifikasi kombinasi menu yang sering dibeli bersamaan, menggunakan parameter *minimum Support* sebesar 2% dan uji *lift ratio* untuk mengukur kekuatan hubungan antar item. Hasil penelitian menunjukkan bahwa kombinasi minuman lebih dominan dibandingkan kombinasi makanan, sehingga dapat dijadikan rekomendasi paket menu untuk meningkatkan penjualan. Pendekatan ini relevan dengan penelitian saya dalam mengidentifikasi pola asosiasi pada data *historis* gangguan berulang jaringan PLN, karena sama-sama menggunakan teknik *association rule* untuk menemukan hubungan antar kejadian yang sering muncul secara bersamaan.

Klasifikasi Kompetensi Alumni Berdasarkan Masa Tunggu Alumni untuk Mendapatkan Pekerjaan Menggunakan Metode Algoritma C4.5 (Rizqia Cahyaningtyas et al., 2020). Penelitian ini memanfaatkan algoritma C4.5 untuk mengklasifikasikan kompetensi alumni berdasarkan masa tunggu kerja pasca kelulusan. Data diperoleh dari tracer study lima tahun terakhir dan diproses menggunakan *RapidMiner* untuk membentuk pohon keputusan yang menunjukkan atribut kompetensi paling berpengaruh. Hasilnya menunjukkan bahwa kompetensi seperti pengetahuan disiplin ilmu, manajemen waktu, kemampuan analisis, dan keterampilan komputer menjadi faktor dominan. Walaupun fokus penelitian ini pada klasifikasi, konsep pengolahan data *historis* dan penggalian pola yang signifikan dapat diadaptasi pada penelitian saya untuk menentukan faktor penyebab utama gangguan berulang pada jaringan PLN.

Data mining Optimization Uses C4.5 Classification and Particle Swarm Optimization (PSO) in the Location Selection of Student Boardinghouses (Waluyo et al., 2020). Penelitian ini menggabungkan algoritma C4.5 dengan *Particle Swarm Optimization* (PSO) untuk memilih lokasi kos mahasiswa yang optimal berdasarkan lima kriteria (kebersihan air, fasilitas, transportasi, keamanan, dan kondisi). PSO digunakan untuk mengoptimalkan bobot atribut sehingga akurasi model meningkat dari 97,53% menjadi 97,78%. Walaupun konteksnya berbeda, pendekatan optimasi ini dapat menjadi inspirasi pada penelitian saya, misalnya untuk menyempurnakan

parameter algoritma *Apriori* agar hasil pola asosiasi gangguan berulang di jaringan PLN lebih akurat dan relevan untuk pengambilan keputusan.

2.1.1 Matriks Penelitian

Berikut merupakan matriks penelitian terdahulu yang berbentuk tabel:

Tabel 2. 1 Matriks Penelitian

Nama Penulis	Tahun Terbit	Judul Penelitian	Metode Penelitian	Hasil Penelitian	Keterkaitan Penelitian
Irwan Adji Darmawan, Muhammad Fakhri Randy, Imam Yunianto, Muhamad Malik Mutoffar, M. Tio Putra Salis	2022	Penerapan <i>Data mining</i> Menggunakan Algoritma <i>Apriori</i> untuk Menentukan Pola Golongan Penyandang Masalah Kesejahteraan Sosial	Algoritma <i>Apriori</i> , <i>Association rule</i> , Pengujian dengan RapidMiner dan perhitungan manual	Menghasilkan 3 aturan asosiasi dengan <i>minimum Support</i> 15% dan <i>confidence</i> hingga 100%; tingkat akurasi manual vs aplikasi hanya 37,5%	Sama-sama menggunakan algoritma <i>Apriori</i> untuk menemukan pola dari data <i>historis</i> , dapat diadaptasi untuk pola gangguan berulang jaringan PLN

Anzas Ibezato Zalukhu, Dewi Sartika, Sri Wahyuni	2024	Penerapan Algoritma <i>Apriori</i> untuk Optimasi Strategi Penjualan Berdasarkan Analisis Pola Pembelian di Torsa Café	Algoritma <i>Apriori</i> , Analisis <i>Support</i> dan <i>Confidence</i>	Menemukan kombinasi menu dengan <i>Support</i> >30% seperti Sanger Espresso dan Avocado Cappuccino Torsa; digunakan untuk strategi promosi dan stok	Relevan karena memanfaatkan pola keterkaitan antar item untuk strategi operasional, dapat diadaptasi untuk identifikasi keterkaitan jenis gangguan
Nanda Nurisya Merliani, Nur Isnaeni Khoerida, Neta Tri Widiawati, Latifah Adi Triana, Pungkas Subarkah	2022	Penerapan Algoritma <i>Apriori</i> pada Transaksi Penjualan untuk Rekomendasi Menu Makanan dan Minuman	Algoritma <i>Apriori</i> , <i>Association rule</i> , Analisis <i>Support</i> dan <i>Confidence</i>	Kombinasi menu “Es Teh Manis” dan “Mendoan” memiliki <i>Support</i> 50% dan <i>confidence</i> 76% sebagai	Relevan dalam penerapan <i>Apriori</i> untuk menemukan kemunculan bersamaan, konsep serupa dapat digunakan untuk identifikasi kombinasi gangguan

				rekomend asi utama	
Afluqfy Harahap, Ade Luthfi Ramadha n Perangin- Angin, Kisen Kumar, Saut Parsaoran Tamba	2022	Analisis Penerapan <i>Data mining</i> dalam Penentuan Tata Letak Barang Menggunak an Algoritma <i>Apriori</i> dan <i>FP-Growth</i>	Perbanding an Algoritma <i>Apriori</i> dan <i>FP-Growth</i> dengan RapidMine r	<i>Apriori</i> menghasil kan 10 aturan, <i>FP- Growth</i> menghasil kan 78 aturan dengan akurasi lebih tinggi	Menunjukkan penggunaan <i>Apriori</i> untuk menemukan pola asosiasi sebagai dasar strategi penempatan, dapat diterapkan pada strategi penanganan gangguan
Saut Parsaoran Tamba, Edbert	2022	Penerapan <i>Data mining</i> Algoritma <i>Apriori</i> dalam Menentukan Stok Bahan Baku pada Restoran Nelayan Menggunak an Metode <i>Association rule</i>	Algoritma <i>Apriori</i> , <i>Association rule</i> , Analisis <i>Support</i> dan <i>Confidence</i>	Menghasil kan 22 aturan asosiasi, misalnya hubungan kuat antara Ikan Gembung dan Ikan Kerapu (100%)	Sama-sama bertujuan mengidentifikasi hubungan antar item untuk prediksi, relevan untuk memprediksi gangguan bersamaan pada jaringan PLN
Irgi Sena Fajar Lukerti,	2024	Perbandinga n Akurasi Algoritma	<i>Apriori</i> vs <i>FP-Growth</i> (min	<i>FP- Growth</i> lebih	Langsung relevan: Menunjukkan

Muhamad Fuat Asnawi, Nulngafan		<i>Apriori</i> dan <i>FP-Growth</i> untuk Pola Gangguan Indihome Area Wonosobo	<i>Support</i> 5%, <i>confidence</i> 50%)	akurat (100% <i>rule</i> valid). Pola utama: kabel luar putus (offline), konfigurasi hilang (online).	penerapan <i>Apriori</i> untuk analisis gangguan jaringan (Indihome), termasuk identifikasi pola berulang seperti kerusakan kabel dan konfigurasi. Dapat diadaptasi untuk analisis gangguan PLN.
Wahyu Satria Ramadhann & Retno Sari	2024	Implementasi Algoritma <i>Apriori</i> untuk Menentukan Pola Transaksi Penjualan	<i>Apriori</i> (RapidMiner, min <i>Support</i> 5%)	Pola pembelian Van Lenning - Iced (<i>confidence</i> 87.8%).	Relevan secara metodologis: Demonstrasi penggunaan <i>Apriori</i> untuk identifikasi pola berulang (transaksi penjualan), mirip dengan pola gangguan berulang PLN.
Alya Fadillah, Akim Manaor Hara Pardede,	2023	Korelasi Kemajuan Teknologi terhadap Kinerja Karyawan	<i>Apriori</i> (RapidMiner, min <i>Support</i> 10%)	44 aturan asosiasi, terbaik: penggunaan Ms Office +	Relevan konteks PLN: Studi kasus di PLN, meskipun fokus pada SDM. Menunjukkan

Indah Ambarita		Menggunakan Metode <i>Apriori</i> , Studi Kasus : PLN Binjai		perangkat kantor → kinerja tinggi.	potensi <i>Apriori</i> untuk analisis data operasional PLN.
Febe Maedjaja, Tri Yuliyanto, Sumeng Sandino Berutu	2023	Implementasi <i>Data mining</i> Untuk Menganalisis Pola Peminjaman Buku Perpustakaan Dengan Menggunakan Metode <i>Apriori</i>	<i>Apriori</i> (Python, min <i>Support</i> 10%)	Pola peminjaman buku fiksi + seni dekorasi (<i>Support</i> 25.15%).	Relevan teknis: Penggunaan <i>Apriori</i> untuk pola asosiasi item (buku) dapat diadaptasi untuk pola gangguan (misal: gangguan transformer + kabel).
Misbahul Marzuqi, Muhammad Ali Ridla, Fajriyanto	2023	Implementasi Algoritma <i>Apriori</i> untuk Menentukan Pola Transaksi Penjualan Berbasis Web	Metode Waterfall, algoritma <i>Apriori</i> dengan <i>minimum Support</i> dan <i>confidence</i> 10%.	Dihasilkan 2 kombinasi <i>itemset</i> yang memenuhi syarat, seperti pembelian roti dan kopi (<i>confidence</i> 40%)	Relevan dasar: Contoh sederhana penerapan <i>Apriori</i> untuk aturan asosiasi, berguna sebagai referensi implementasi dasar untuk penelitian PLN.

				serta medialuna dan roti (<i>confidence</i> 50%).	
Hendra Jatnika, Ari Waluyo, Abdul Azis	2024	A Comparative Study on Data Collection Methods: Investigating Optimal Datasets for <i>Data mining</i> Analysis	Perbandingan metode pengumpulan data (kuesioner & web mining) dengan algoritma SVM dan NBC	Data kuesioner menghasilkan akurasi lebih tinggi (>82% SVM; >81% NBC) dibanding web mining karena lebih bersih dari simbol yang mengganggu	Menunjukkan pentingnya kualitas data input untuk meningkatkan akurasi algoritma, relevan untuk memastikan data <i>historis</i> gangguan listrik yang digunakan <i>Apriori</i> berkualitas baik
Dwina Kuswardani, M. Rahmat Widyanto, Indang	2011	Metode <i>Association rule</i> untuk Analisis Citra CT Organ	Algoritma <i>Apriori</i> pada <i>Association rule</i> , ekstraksi	Menghasilkan pola asosiasi organ dengan nilai	Menunjukkan penerapan <i>Apriori</i> dalam menemukan pola asosiasi dari data kompleks, dapat

Trihandini		Pasien Kanker Ovarium	fitur ukuran & intensitas citra CT scan, transformasi data numerik menjadi kategori	<i>Support & confidence</i> tertentu, misalnya ukuran uterus <i>Support</i> 55%, hubungan liver-uterus <i>Support</i> 45%	diadaptasi untuk menganalisis hubungan antar jenis gangguan berulang pada jaringan PLN
Dewi Arianti Wulandari, Herman Bedi, Luqman, Ahmad Arfan Mashuda	2024	Faktor-Faktor yang Berpengaruh untuk Mengestimasi Waktu Kelulusan Mahasiswa dengan Pendekatan Algoritma C4.5	Algoritma C4.5 (decision tree), perhitungan entropy dan gain, confusion matrix	Akurasi prediksi kelulusan >70%, ditemukan faktor IPK, motivasi, dan keaktifan organisasi sebagai penentu utama	Menunjukkan pentingnya pemilihan atribut dan pembersihan data sebelum <i>data mining</i> , relevan untuk tahap <i>preprocessing</i> sebelum menjalankan <i>Apriori</i> pada data gangguan listrik
Sarah Juliandiny, Pritasari Palupinin	2022	Implementation of ECLAT Algorithm to	<i>Data mining – Association rule</i> menggunakan	Mengidentifikasi pola pembelian produk	Relevan karena menggunakan metode <i>association rule</i> untuk

gsih, Herman Bedi Agtriadi, Budi Prayitno, Eka Putra		Determine Product Purchasing Pattern at Coffee Shop	an algoritma ECLAT, analisis <i>Support</i> , <i>confidence</i> , dan <i>lift</i> ratio	secara bersamaan di coffee shop. Kombinas i minuman lebih dominan dibanding kan makanan, direkomen dasikan untuk paket menu.	menemukan hubungan antar item yang sering muncul bersama, sama seperti penerapan <i>Apriori</i> untuk pola asosiasi gangguan jaringan PLN.
Rizqia Cahyanin gtyas, Luqman, R. Yolana Inas H.	2020	Klasifikasi Kompetensi Alumni Berdasarkan Masa Tunggu Alumni untuk Mendapatka n Pekerjaan Menggunak an Metode Algoritma C4.5	<i>Data mining</i> – Algoritma C4.5, pengolahan tracer study alumni, pengujian akurasi dengan Confusion Matrix	Menemuk an kompetens i utama yang mempeng aruhi masa tunggu kerja, seperti pengetahu an disiplin ilmu, manajeme n waktu, kemampu	Memberikan gambaran pemanfaatan data <i>historis</i> untuk mengidentifikasi faktor dominan. Pendekatan ini relevan dalam menentukan faktor penyebab utama gangguan berulang PLN.

				an analisis, dan keterampil an komputer.	
Ari Waluyo, Hendra Jatnika, Marlina Rahmi Shinta Permatas ari, T. Tuslaela, Indah Purnamas ari, Agus Perdana Windarto	2020	<i>Data mining</i> Optimizatio n uses C4.5 Classificatio n and Particle Swarm Optimizatio n (PSO) in the Location Selection of Student Boardingho uses	<i>Data mining</i> – Algoritma C4.5 dikombinas ikan dengan optimasi Particle Swarm Optimizati on (PSO)	Akurasi model meningkat dari 97,53% menjadi 97,78% setelah optimasi PSO. Faktor paling berpengar uh adalah kebersihan air dan keamanan.	Memberikan ide optimasi parameter untuk meningkatkan akurasi model. Dapat diterapkan pada optimasi parameter <i>Apriori</i> untuk menghasilkan pola asosiasi gangguan yang lebih tepat.

2.2 Landasan Teori

2.2.1 *Data mining*

Data mining merupakan proses penggalian informasi dan penemuan pola tersembunyi dari dataset berukuran besar melalui penerapan berbagai metode analisis. Proses ini melibatkan beberapa tahapan utama, yaitu pemilihan data (data selection), pembersihan data (data *cleaning*), transformasi data (data transformation), penambangan pola (pattern mining), dan evaluasi hasil (pattern *evaluation*) (Peralatan et al., 2020).

Metode *data mining* dapat digunakan dalam berbagai konteks, termasuk asosiasi, klasifikasi, klustering, dan prediksi, bergantung pada tujuan analisis. Dalam konteks penelitian ini, *data mining* menyediakan kerangka kerja yang sistematis untuk menganalisis data gangguan berulang pada jaringan PLN. Data *historis* gangguan yang bersifat kompleks dapat diproses mulai dari tahap ekstraksi, pembersihan, transformasi atribut (seperti kondisi cuaca, waktu, jenis peralatan), hingga penambangan pola menggunakan algoritma *Apriori*. Hasil analisis berupa aturan asosiasi dengan nilai *support*, *confidence*, dan *lift* yang tinggi dapat digunakan sebagai dasar penyusunan strategi mitigasi gangguan. Dengan proses yang terstruktur, *data mining* tidak hanya membantu menemukan pola, tetapi juga memastikan pola tersebut relevan secara teknis dan dapat diterapkan dalam perencanaan operasional.

2.2.2 *Unsupervised learning*

Unsupervised learning adalah metode pembelajaran mesin yang menganalisis data tanpa label target, dengan tujuan menemukan pola, struktur, atau hubungan tersembunyi. Pendekatan ini mencakup teknik seperti *clustering*, *dimensionality reduction*, dan *association rule mining* yang terakhir kerap menggunakan algoritma *Apriori* untuk mengidentifikasi keterkaitan antar atribut berdasarkan frekuensi kemunculannya (Fauzi et al., 2025).

Keunggulan metode ini terletak pada kemampuannya mengolah data besar dan kompleks tanpa persiapan label, sehingga cocok untuk eksplorasi awal. Namun, kualitas hasil sangat bergantung pada pra-pemrosesan data dan interpretasi berbasis pengetahuan domain. Dalam konteks penelitian ini, *unsupervised learning* relevan karena data gangguan PLN bersifat heterogen dan sering tanpa label penyebab yang jelas. Dengan *association rule mining*, kombinasi faktor pemicu gangguan misalnya kondisi cuaca tertentu yang sering disertai jenis gangguan spesifik dapat terungkap dan dimanfaatkan untuk strategi pemeliharaan *preventif* di wilayah rawan gangguan.

2.2.3 *Association rule*

Association rule atau aturan asosiasi merupakan salah satu metode dalam *data mining* yang digunakan untuk mengidentifikasi keterkaitan signifikan antara item atau atribut dalam sebuah dataset. Hubungan ini biasanya dituliskan dalam bentuk aturan $A \rightarrow B$, di mana A adalah *antecedent* (kondisi awal) dan B adalah *consequent*

(hasil atau kejadian yang mengikuti). *Metrik* utama yang digunakan untuk menilai kualitas aturan meliputi *support*, yang mengukur proporsi transaksi yang mengandung kombinasi item tertentu, dan *confidence*, yang menunjukkan probabilitas *consequent* muncul jika *antecedent* terjadi. Selain itu, *lift* digunakan untuk menilai kekuatan asosiasi, dengan nilai lebih dari satu menandakan adanya hubungan positif yang signifikan (Fauzi et al., 2025).

Association rule atau aturan asosiasi adalah metode dalam *data mining* yang digunakan untuk menemukan keterkaitan atau hubungan signifikan antara item dalam Proses pembentukan *association rule* pada umumnya melalui dua tahapan inti berikut:

1. *Frequent Itemset Generation*

Tahap ini bertujuan untuk menemukan kumpulan item (*itemset*) yang sering muncul bersama dalam dataset. Proses dimulai dengan menghitung nilai *support* untuk setiap kombinasi item. Jika nilai *support* suatu *itemset* melebihi ambang batas *minimum support* yang telah ditentukan, *itemset* tersebut dianggap *frequent* dan akan digunakan untuk membentuk kombinasi *itemset* yang lebih besar pada iterasi berikutnya. Sebaliknya, *itemset* yang tidak memenuhi *minimum support* akan dieliminasi sejak awal. Tahap ini sangat penting karena berfungsi sebagai penyaring awal untuk memastikan hanya pola yang cukup sering muncul yang dipertahankan untuk proses berikutnya.

2. *Rule Generation*

Setelah *frequent itemset* diperoleh, tahap selanjutnya adalah membentuk aturan asosiasi dari *itemset* tersebut. Proses ini dilakukan dengan memisahkan *frequent itemset* menjadi bagian *antecedent* dan *consequent*, kemudian menghitung nilai *confidence* untuk setiap aturan yang dihasilkan. Aturan hanya dipertahankan jika nilai *confidence* memenuhi ambang *minimum confidence* yang telah ditetapkan. Pada tahap ini, analisis lanjutan seperti perhitungan *lift*, *conviction*, atau *leverage* dapat dilakukan untuk memvalidasi bahwa hubungan antar item bukanlah hasil kebetulan. Tahap ini memastikan aturan yang dipilih memiliki relevansi praktis dan dapat digunakan sebagai dasar pengambilan keputusan (Arifin et al., 2025).

2.2.4 Algoritma *Apriori*

Algoritma *Apriori* adalah metode klasik yang digunakan untuk menemukan *frequent itemset* dan menghasilkan aturan asosiasi. Algoritma ini bekerja secara iteratif dengan membentuk kandidat *itemset* dari *frequent itemset* sebelumnya, kemudian menghitung nilai *support* dan menyaring hanya kombinasi item yang memenuhi ambang batas *minimum support*. Prinsip utama yang digunakan adalah sifat anti-monotonic atau *downward closure*, yaitu jika suatu himpunan item tidak memenuhi *minimum support*, maka semua superset-nya juga tidak akan memenuhi syarat tersebut (Arifin et al., 2025).

Secara matematis, nilai *support* dan *confidence* pada algoritma *Apriori* dirumuskan sebagai berikut:

$$Support(A \rightarrow B) = \frac{\text{Jumlah transaksi yang mengandung } (A \cup B)}{\text{Total jumlah transaksi}} \quad (2.1)$$

$$Confidence(A \rightarrow B) = \frac{\text{Jumlah transaksi yang mengandung } (A \cup B)}{\text{Jumlah transaksi yang mengandung } A} \quad (2.2)$$

Selain dua parameter tersebut, *lift* juga digunakan sebagai indikator kekuatan asosiasi, dengan rumus:

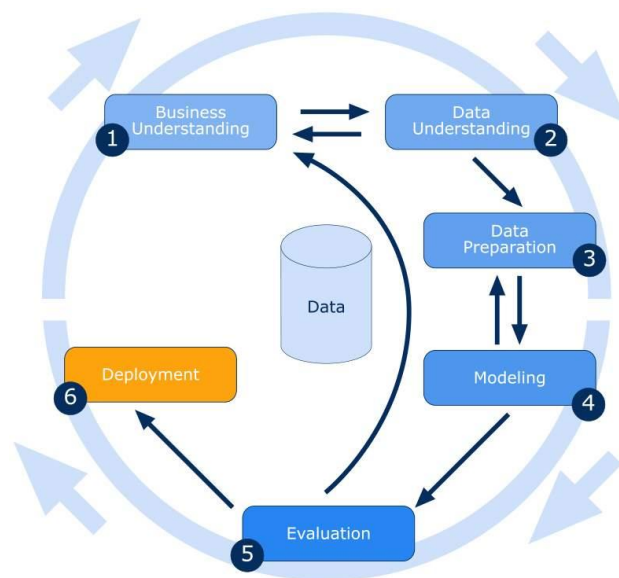
$$Lift(A \rightarrow B) = \frac{Confidence(A \rightarrow B)}{Support(B)} \quad (2.3)$$

$lift > 1$ menunjukkan bahwa kemunculan *antecedent* meningkatkan kemungkinan munculnya *consequent*.

Dalam penelitian-penelitian terdahulu, *Apriori* terbukti efektif dalam menghasilkan pola asosiasi yang jelas meskipun memerlukan waktu komputasi lebih lama dibandingkan algoritma lain seperti *FP-Growth* (Peralatan et al., 2020). Relevansi algoritma ini dengan penelitian skripsi terletak pada kemampuannya dalam menghasilkan pola yang dapat dipertanggungjawabkan secara transparan. Dengan menentukan parameter seperti *minimum support* dan *minimum confidence* yang tepat, peneliti dapat menemukan hubungan signifikan antara faktor-faktor penyebab gangguan pada jaringan PLN, sehingga dapat dijadikan acuan dalam penyusunan rencana pemeliharaan yang lebih efektif.

2.2.5 Cross Industry Standard Process for Data mining (CRISP-DM)

CRISP-DM (*Cross Industry Standard Process for Data mining*) merupakan kerangka kerja yang diakui secara luas dalam proses *data mining* di berbagai sektor industri. Kerangka ini memberikan panduan terstruktur untuk menganalisis data dan memecahkan masalah kompleks di bidang bisnis maupun penelitian. Dengan fokus pada pemahaman masalah, pengolahan data, dan penerapan solusi berbasis pengetahuan, CRISP-DM memfasilitasi penemuan pola, tren, dan informasi penting dalam data besar. Fleksibilitasnya memungkinkan penerapan di berbagai domain, serta menjadi acuan utama dalam strategi analisis menyeluruh yang mencakup tahapan dari pemahaman konteks hingga pemanfaatan hasil analisis untuk pengambilan keputusan berbasis data (*Abdulhakim et al., 2021*).



Gambar 2. 1 Visualisasi CRISP-DM

1. *Business Understanding*

Tahap awal dalam kerangka CRISP-DM adalah *Business Understanding*, yang berfungsi untuk menjabarkan tujuan proyek serta strategi pemanfaatan data guna mendukung pencapaian target tersebut. Fokus utamanya adalah memahami konteks permasalahan, kebutuhan pemangku kepentingan, dan relevansi data terhadap model bisnis yang dianalisis, sehingga proses *data mining* menghasilkan temuan yang tidak

hanya tepat secara teknis, tetapi juga sejalan dengan sasaran bisnis atau penelitian.

Kegiatan pada tahap ini mencakup diskusi dengan pihak terkait, telaah dokumen, analisis laporan *historis*, hingga pembatasan ruang lingkup permasalahan agar analisis lebih terfokus. Pendekatan seperti analisis kebutuhan, *brainstorming*, atau *workshop* digunakan untuk menggali informasi secara mendalam dan menyamakan persepsi tim. Hasil yang dicapai adalah pemahaman jelas mengenai latar belakang, tujuan, dan batasan proyek. Penetapan tujuan sejak awal menjadi landasan seluruh proses berikutnya, mulai dari pengumpulan data hingga penerapan hasil, menjadikan *Business Understanding* sebagai fondasi penting yang menjaga konsistensi setiap tahap (Singh et al., 2023).

2. Data Understanding

Tahap Data Understanding pada kerangka CRISP-DM merupakan langkah kedua yang bertujuan memperoleh pemahaman mendalam mengenai karakteristik, potensi, dan keterbatasan data. Fokus utamanya adalah menilai kualitas data dari aspek kelengkapan, distribusi nilai, konsistensi format, serta kepatuhan terhadap prinsip data governance. Kualitas data pada tahap ini sangat berpengaruh terhadap akurasi dan keandalan model. Kegiatan yang dilakukan meliputi eksplorasi awal, deskripsi data, dan verifikasi integritas untuk mendeteksi data hilang, nilai ekstrem, atau inkonsistensi. Analisis statistik digunakan untuk melihat distribusi, hubungan variabel, dan potensi bias. Selain itu, data scientists meninjau penggunaan data saat ini dan membandingkannya dengan kebutuhan ideal untuk mendukung analisis serta algoritma yang digunakan. Output tahap ini berupa dokumentasi kondisi data, permasalahan yang ditemukan, dan rekomendasi perbaikan, yang menjadi dasar penting untuk tahap Data Preparation berikutnya (Singh et al., 2023).

3. Data Preparation

Tahap *Data Preparation* bertujuan mengubah data mentah menjadi dataset akhir yang siap dipakai untuk pemodelan. Proses ini meliputi seleksi tabel, catatan, dan atribut sesuai kriteria inklusi-eksklusi, pembersihan data

untuk mengatasi nilai kosong, duplikasi, atau format tidak konsisten, serta transformasi seperti normalisasi, standarisasi, dan pengkodean variabel. Pada tahap ini juga dilakukan *feature engineering* untuk menghasilkan atribut baru yang relevan dengan model. Langkah-langkahnya bersifat berulang dan disesuaikan dengan jenis algoritma yang digunakan, sehingga dihasilkan dataset bersih, terstruktur, dan sesuai kebutuhan analisis (Brzowska et al., 2023).

4. *Modelling*

Tahap Modelling merupakan bagian penting dalam proses *data mining*, di mana teknik atau algoritma yang sesuai dipilih, diterapkan, dan disesuaikan parameternya guna memperoleh hasil optimal. Pemilihan metode harus mempertimbangkan permasalahan yang ingin dipecahkan, tujuan bisnis, serta karakteristik data. Berbagai algoritma seperti classification, clustering, atau *association rule mining* dapat digunakan untuk masalah serupa, namun masing-masing memiliki persyaratan spesifik terhadap format, tipe, dan struktur data. Beberapa metode membutuhkan data numerik terstandarisasi, sementara lainnya dapat memproses data kategorikal. Karena itu, tahap data preparation kerap memerlukan penyesuaian seperti encoding, normalisasi, atau penanganan outlier (Brzowska et al., 2023).

Pemodelan juga melibatkan hyperparameter tuning untuk mengoptimalkan kinerja model. Setelah model dibangun, evaluasi dilakukan dengan *metrik* seperti accuracy, precision, recall, F1-score, atau *Support* dan *confidence* pada *association rule*. Penting bagi peneliti menjelaskan alasan pemilihan metode berdasarkan relevansi data, keterbatasan, dan efisiensi komputasi. Dengan demikian, Modelling bukan sekadar penerapan algoritma, tetapi keputusan strategis berbasis bukti untuk menghasilkan model akurat dan andal (Brzowska et al., 2023).

5. *Evaluation*

Tahap *Evaluation* berfungsi sebagai proses peninjauan menyeluruh untuk memastikan bahwa model yang dihasilkan benar-benar sejalan dengan sasaran bisnis maupun tujuan penelitian. Langkah ini meliputi

pengujian terhadap tingkat akurasi, relevansi, kemudahan interpretasi, serta kegunaan model dalam praktik, sambil melakukan pemeriksaan ulang pada seluruh fase sebelumnya agar tidak ada faktor penting yang terabaikan, seperti variabel krusial atau potensi bias pada data. Proses pengujian memanfaatkan metode seperti cross-validation serta indikator kinerja yang tepat, kemudian dilanjutkan dengan analisis kelayakan penerapan yang mempertimbangkan efisiensi, kemudahan integrasi, dan potensi risiko. Temuan dari evaluasi ini akan menjadi acuan untuk memutuskan apakah model tersebut layak diimplementasikan, memerlukan penyesuaian, atau sebaiknya diganti dengan alternatif yang lebih efektif (Brzowska *et al.*, 2023).

6. Deployment

Tahap *deployment* merupakan kelanjutan setelah pembuatan model, menegaskan bahwa pengembangan model bukanlah akhir proyek. Meski tujuannya hanya memperdalam pemahaman data, wawasan yang dihasilkan harus diolah dan disajikan agar berguna bagi pengguna akhir. Hasil tersebut dapat berupa dokumentasi, visualisasi, atau laporan terstruktur yang mendukung pengambilan keputusan. Dalam banyak kasus, model langsung diintegrasikan ke alur kerja organisasi, seperti personalisasi halaman web, rekomendasi produk, atau evaluasi basis data pemasaran (Brzowska *et al.*, 2023).

Kompleksitas implementasi bervariasi, dari laporan sederhana hingga sistem *data mining* terintegrasi yang digunakan berulang di seluruh unit organisasi. Bentuk implementasi bergantung pada kebutuhan, infrastruktur, dan sumber daya yang tersedia. Biasanya pelaksanaan dilakukan oleh pihak klien, sementara analisis data berperan melalui konsultasi teknis untuk memastikan model diintegrasikan secara efektif dan memberikan manfaat optimal sesuai tujuan awal (Brzowska *et al.*, 2023).