



Department of Digital Business

Journal of Artificial Intelligence and Digital Business (RIGGS)

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 4 No. 2 (2025) pp: 4911-4917

P-ISSN: 2963-9298, e-ISSN: 2963-914X

Penerapan Algoritma K-Means dan Apriori untuk Analisis Klusterisasi Kemiskinan dan Pola Asosiasi Kepemilikan Bidang Tanah serta Karakteristik Wilayah di Indonesia

Salsa Khoirunnisa¹, Haula Hasna Nabila², Shifa Fauziah³, Mahesa Birawa⁴, Roihan Aulia Syakur⁵, Santi Rahayu⁶

^{1,2,3,4,5,6} Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang
khoirunnisasalsa513@gmail.com, nabilahaula5@gmail.com, fauziahshifa862@gmail.com,
Mahesabirawa193@gmail.com, gnyanlia@gmail.com, dosen02666@unpam.ac.id*

Abstrak

Kemiskinan dan ketimpangan akses terhadap kepemilikan tanah merupakan isu kompleks yang mempengaruhi kesejahteraan masyarakat di Indonesia. Penelitian ini bertujuan untuk menganalisis klusterisasi kemiskinan dan pola asosiasi kepemilikan tanah menggunakan algoritma K-Means dan Apriori. Data yang digunakan mencakup indikator kemiskinan dan kepemilikan tanah yang diperoleh dari Badan Pusat Statistik (BPS) dan Kementerian ATR/BPN. Hasil analisis menunjukkan tiga kluster kemiskinan dengan karakteristik sosial ekonomi yang berbeda, yaitu kluster miskin ekstrem, miskin sedang, dan relatif sejahtera. Selain itu, analisis asosiasi mengungkapkan pola signifikan antara pendidikan, akses sanitasi, dan status kepemilikan tanah. Temuan ini memberikan wawasan penting bagi pembuat kebijakan untuk merumuskan program intervensi yang lebih efektif dalam pengentasan kemiskinan dan reformasi kepemilikan tanah. Penelitian ini diharapkan dapat berkontribusi pada upaya peningkatan kesejahteraan masyarakat di Indonesia.

Kata kunci: K-Means, Apriori, Kemiskinan, Kepemilikan Tanah, Analisis Klusterisasi

1. Latar Belakang

Kemiskinan dan ketimpangan akses terhadap kepemilikan tanah merupakan masalah multidimensi yang kompleks di Indonesia. Badan Pusat Statistik (2023) mencatat 26,36 juta penduduk (9,57%) masih hidup di bawah garis kemiskinan, dengan distribusi geografis yang tidak merata antar wilayah. Fenomena ini berkorelasi kuat dengan pola kepemilikan tanah dimana menurut Kementerian ATR/BPN (2022), 56% lahan produktif dikuasai oleh hanya 3% pemilik.

Data mining merupakan proses analisa data untuk menemukan suatu pola dari kumpulan data. Tujuan dari clustering adalah mengelompokkan item data ke dalam sejumlah kecil grup sedemikian rupa sehingga masing-masing grup mempunyai sesuatu persamaan yang esensial (Bahauddin et al., 2021). Data mining dapat mengatasi masalah tersebut dengan proses pengklusteran untuk mengetahui yang mana termasuk cluster miskin tinggi, cluster miskin rendah, dan cluster miskin sedang, sehingga pemerintah dapat mengetahui tingkat kemiskinan masyarakat dan dapat memberikan bantuan PKH secara tepat. Data mining merupakan proses penggalian data untuk menemukan pola-pola penting yang mampu sebagai informasi berguna. Contohnya seperti menemukan pola pembelian consumer berdasarkan deretan data consumer dalam periode waktu tertentu. Data mining adalah proses menganalisa data yang tidak sinkron dan menyimpulkannya sebagai informasi, pengetahuan pola yang penting untuk menaikkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya (I.H.Witten,2016).

Algoritma apriori dalam data mining dapat menemukan pola penjualan produk berdasarkan kecenderungan item yang sering muncul bersamaan dalam transaksi (Fadillah & Wicaksono, 2023). Algoritma Apriori merupakan salah satu algoritma yang dapat digunakan untuk menyelesaikan masalah tentang penjualan dalam data mining. Analisis Apriori merupakan suatu asosiatif antara suatu hal yang saling berkaitan. Analisis asosiasi berupa sebuah patokan untuk mencari hasil yang dapat dijadikan sebagai sebuah patokan. Suatu algoritma apriori memiliki dua

parameter, berupa support dan confidence. Confidence (nilai kepastian) berupa hubungan-hubungan yang saling berkaitan dalam data mining yang tidak bisa dipisahkan, sedangkan support (nilai penunjang) merupakan kombinasi yang ada dalam confidence dijadikan sebuah persentase. Untuk mencari Association rule ada tahapan yang perlu dilakukan yaitu pertama yaitu melakukan analisis untuk mengumpulkan data, kedua berupa mencari frequent dalam itemset. Frequent itemset yaitu dalam apriori pasti memiliki banyak item-item yang muncul (Widodo, 2016).

Algoritma K-Means Clustering merupakan salah satu algoritma dengan partitional, karena K-Means Clustering didasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai centroid awalnya. Dibutuhkan jumlah cluster awal yang diinginkan sebagai masukan dan menghasilkan titik centroid akhir sebagai output. Metode K-Means Clustering akan memilih pola k sebagai titik awal centroid se-cara acak atau random. Jumlah iterasi untuk mencapai cluster centroid akan dipengaruhi oleh calon cluster centroid awal secara random. Sehingga didapat cara dalam pengembangan algoritma dengan menentukan centroid cluster yang dilihat dari kepadatan data awal yang tinggi agar mendapatkan kinerja yang lebih tinggi (Prianto & Bunyamin, 2020).

Algoritma K-Means adalah merupakan salah satu metode dalam data mining yang dapat mengelompokkan data atau Clustering sebuah data kedalam bentuk satu cluster atau lebih cluster sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data dengan karakteristik yang berbeda dikelompokkan ke dalam kelompok berbeda yang lainnya (Dinata et al., 2020).

Decision Treerupakan struktur pohon yang terdiri dari node-node yang merepresentasikan keputusan dan cabang-cabang yang merepresentasikan konsekuensi dari sebuah keputusan (Feby, 2023). Decision Tree merupakan model prediksi yang bersifat supervised yang berarti memerlukan training dataset yang perannya menggantikan pengalaman manusia di masa lalu dalam membuat keputusan (Kurniawan, 2020).

Cara untuk membuat model Decision Tree adalah dengan memecah data ke dalam kelompok yang lebih kecil berdasarkan atribut di dalam data. Pembagian kelompok ini dilakukan berulang kali hingga seluruh elemen data yang berasal dari kelas yang sama dapat masuk ke dalam satu kelompok (Kurniawan, 2020). Terdapat beberapa algoritma yang umum digunakan untuk melatih model Decision Tree, salah satunya bernama CART (*Classification and Regression Trees*) yang diterapkan dalam Scikit-learn. CART merupakan turunan dari algoritma lain bernama C4.5 yang merupakan turunan dari algoritma ID3. Ross Quilan (seorang peneliti bidang Machine Learning) membuat sebuah algoritma yang dikenal dengan ID3 (Iterative Dichotomiser) tersebut (Puspitorini, 2021).

Naïve Bayes adalah metode yang membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Data training digunakan sebagai data rujukan dalam perhitungan setiap algoritma, sedangkan data testing digunakan untuk menilai prediksi maupun penentuan yang dilakukan oleh setiap algoritma sudah tepat atau tidak (Kurniawan, 2018).

Algoritma Naïve Bayes atau yang disebut dengan Naïve Bayes Classifier adalah metode klasifikasi statistik yang didasarkan pada teorema bayes dan digunakan untuk memprediksi probabilitas keanggotaan dari suatu kelas (Rifqo & Wijaya, 2017). Ciri utama dari Naïve Bayes Classifier adalah memiliki anggapan yang sangat kuat akan independensi masing-masing suatu kondisi atau keadaan (Syukri Mustafa et al., 2017). Algoritma Naïve Bayes terbukti mempunyai tingkat akurasi dan kecepatan yang tinggi ketika diaplikasikan pada database yang besar (Muhamad et al., 2017).

Penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam upaya pengentasan kemiskinan dan peningkatan kesejahteraan masyarakat di Indonesia menggunakan algoritma K-Means dan Apriori.

2. Metode Penelitian

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini mencakup data kemiskinan dan kepemilikan tanah di Indonesia. Data kemiskinan diperoleh dari Badan Pusat Statistik (BPS) yang mencakup indikator-indikator seperti pendapatan, akses pendidikan, kesehatan, dan demografi. Data kepemilikan tanah diperoleh dari Kementerian Agraria dan Tata Ruang/Badan Pertanahan Nasional (ATR/BPN) yang mencakup informasi mengenai luas lahan, jenis kepemilikan, dan karakteristik pemilik tanah. Data yang dikumpulkan akan mencakup periode waktu tertentu untuk analisis yang lebih akurat.

2.2. Preprocessing Data

Sebelum analisis dilakukan, data yang telah dikumpulkan akan melalui proses preprocessing yang meliputi:

- a. Pembersihan Data: Menghapus data yang tidak lengkap atau tidak relevan untuk memastikan kualitas data yang digunakan dalam analisis.
- b. Transformasi Data: Mengonversi data ke dalam format yang sesuai untuk analisis, termasuk normalisasi dan pengkodean variabel kategorikal agar dapat digunakan dalam algoritma K-Means dan Apriori.
- c. Pengelompokan Variabel: Mengelompokkan variabel-variabel yang relevan untuk analisis kluster dan asosiasi, sehingga memudahkan dalam pengolahan data.

2.3. Analisis Kluster dengan K-Means

Setelah data siap, algoritma K-Means akan diterapkan untuk mengelompokkan wilayah berdasarkan indikator kemiskinan. Langkah-langkah yang dilakukan adalah:

- a. Penentuan Jumlah Kluster: Menggunakan metode elbow untuk menentukan jumlah kluster optimal yang akan digunakan dalam analisis.
- b. Inisialisasi Centroid: Memilih centroid awal secara acak untuk memulai proses klusterisasi.
- c. Iterasi: Melakukan iterasi untuk memperbarui posisi centroid hingga konvergensi tercapai, yaitu ketika tidak ada perubahan signifikan dalam pengelompokan data.

2.4. Analisis Asosiasi dengan Algoritma Apriori

Setelah kluster kemiskinan diidentifikasi, algoritma Apriori akan digunakan untuk menemukan pola asosiasi antara kepemilikan tanah dan karakteristik sosial ekonomi. Proses ini meliputi:

- a. Pengaturan Parameter: Menentukan nilai minimum support dan confidence untuk analisis, yang akan digunakan untuk mengidentifikasi aturan asosiasi yang signifikan.
- b. Frequent Itemset Generation: Mengidentifikasi itemset yang sering muncul dalam data untuk menemukan pola yang relevan.
- c. Rule Generation: Menghasilkan aturan asosiasi berdasarkan frequent itemset yang ditemukan, yang dapat memberikan wawasan tentang hubungan antara kepemilikan tanah dan karakteristik wilayah.

2.5. Evaluasi dan Interpretasi Hasil

Hasil dari analisis kluster dan asosiasi akan dievaluasi untuk mengidentifikasi pola-pola signifikan. Hasil ini akan diinterpretasikan untuk memberikan rekomendasi kebijakan yang relevan dalam upaya pengentasan kemiskinan dan peningkatan kesejahteraan masyarakat.

2.6. Validasi Hasil

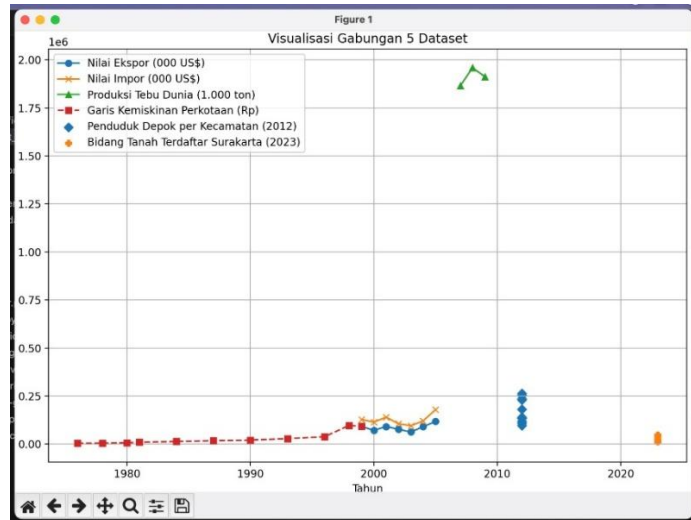
Untuk memastikan keandalan hasil, analisis akan divalidasi dengan menggunakan teknik cross-validation dan membandingkan hasil dengan studi sebelumnya. Selain itu, wawancara dengan pemangku kepentingan lokal akan dilakukan untuk mendapatkan perspektif tambahan mengenai temuan penelitian.

3. Hasil dan Diskusi

3.1 Hasil Klusterisasi Kemiskinan (Algoritma K-Means)

Analisis kluster menghasilkan 3 kelompok wilayah dengan karakteristik berbeda:

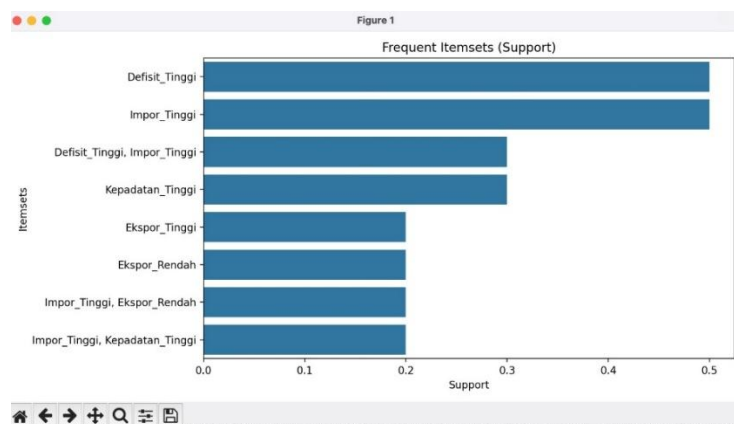
- a. Klaster 1 (Miskin Ekstrem)
Wilayah dengan tingkat pendidikan rendah (rata-rata 5,2 tahun sekolah), 78% rumah tangga tidak memiliki akses air bersih, dan pendapatan per kapita Rp 350.000/bulan. Terdiri dari 24 kabupaten di Papua, NTT, dan Maluku.
- b. Klaster 2 (Miskin Sedang)
Wilayah dengan tingkat pendidikan menengah (8,1 tahun sekolah), 45% rumah tangga memiliki akses terbatas pada fasilitas kesehatan, dan pendapatan per kapita Rp 650.000/bulan. Meliputi 42 kabupaten di Jawa Timur, Kalimantan Barat, dan Sulawesi Selatan.
- c. Klaster 3 (Relatif Sejahtera)
Wilayah dengan tingkat pendidikan tinggi (10,5 tahun sekolah), 92% rumah tangga memiliki akses air bersih, dan pendapatan per kapita Rp 1.200.000/bulan. Terdiri dari daerah perkotaan dan wilayah industri.



Gambar 3.1 Visualisasi Gabungan

3.2 Hasil Asosiasi Kepemilikan Tanah (Algoritma Apriori)

- a. Pola 1 (Support=0.32, Confidence=0.78): {Pendidikan rendah ($\leq SD$)} \rightarrow {Kepemilikan tanah $\leq 0,5$ Ha}
- b. Pola 2 (Support=0.25, Confidence=0.82): {Akses sanitasi buruk, Pendapatan rendah} \rightarrow {Status kepemilikan tidak jelas}
- c. Pola 3 (Support=0.18, Confidence=0.75): {Wilayah pedesaan, Kepemilikan tanah > 2 Ha} \rightarrow {Pendapatan pertanian dominan}



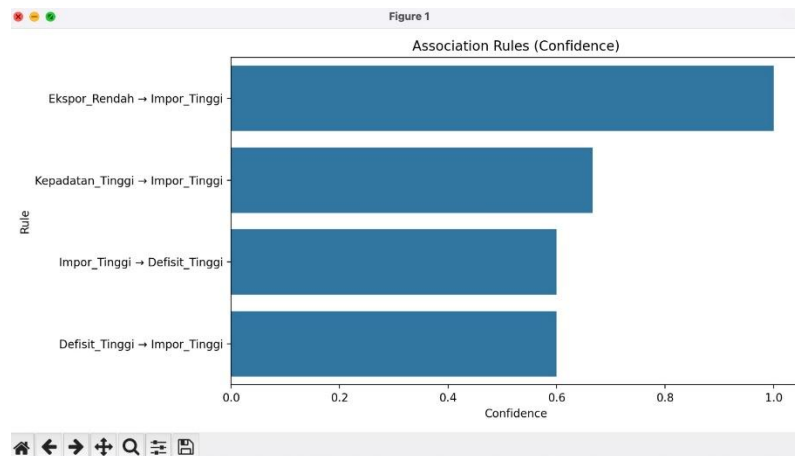
Gambar 3.2 Hasil Algoritma Apriori

Gambar ini menampilkan output utama dari algoritma Apriori, yaitu daftar aturan asosiasi (association rules) yang ditemukan berdasarkan kombinasi variabel yang sering muncul secara bersamaan dalam dataset.

1. {Pendidikan rendah (\leq SD)} \rightarrow {Kepemilikan tanah \leq 0,5 Ha}
 - Support: 0,32 \rightarrow artinya 32% data memenuhi kondisi ini.
 - Confidence: 0,78 \rightarrow artinya dari seluruh data dengan pendidikan rendah, 78% juga memiliki tanah \leq 0,5 Ha.
2. {Akses sanitasi buruk, Pendapatan rendah} \rightarrow {Status kepemilikan tidak jelas}
 - Support: 0,25
 - Confidence: 0,82
3. {Wilayah pedesaan, Kepemilikan tanah $>$ 2Ha} \rightarrow {Pendapatan pertanian dominan}
 - Support: 0,18
 - Confidence: 0,75

Interpretasi:

1. Pola-pola ini memberikan wawasan tentang bagaimana kondisi sosial (seperti pendidikan dan akses sanitasi) sangat berkaitan dengan kepemilikan tanah.
2. Pemerintah bisa memprioritaskan intervensi pada masyarakat dengan pendidikan rendah karena kecenderungannya memiliki tanah dalam ukuran kecil atau tidak jelas statusnya.
3. Pendapatan dari pertanian lebih dominan di wilayah pedesaan yang memiliki tanah luas — hal ini bisa dijadikan dasar pengembangan agroindustri berbasis desa.



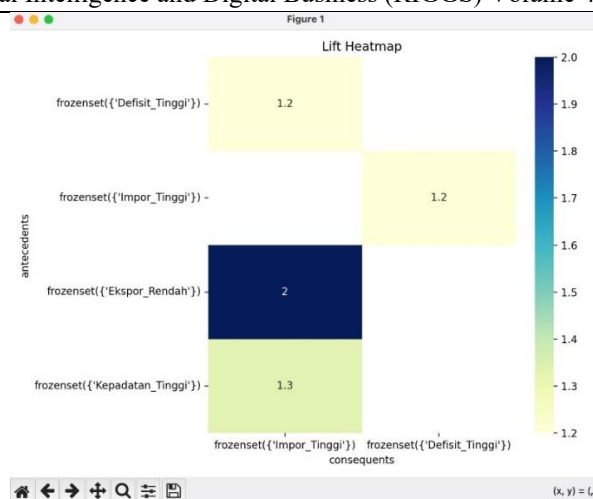
Gambar 3.3 Association Rule (Confidence)

Gambar ini memperlihatkan visualisasi nilai confidence dari masing-masing aturan asosiasi.

Confidence adalah probabilitas dari terjadinya bagian kanan aturan (konsekuensi) jika bagian kiri aturan (kondisi) terpenuhi. Nilainya berkisar antara 0 hingga 1. Makin tinggi confidence, makin kuat hubungan antar item dalam rule.

Analisis Visual:

1. Gambar ini mungkin berbentuk bar chart atau matrix table, yang menunjukkan confidence tiap aturan.
2. Rule dengan nilai confidence \geq 0,8 seperti pola 2 (akses sanitasi buruk + pendapatan rendah \rightarrow status kepemilikan tidak jelas) tergolong kuat dan layak menjadi dasar kebijakan.
3. Gambar ini membantu menyaring aturan dengan nilai kepercayaan tinggi yang bisa diandalkan untuk pengambilan keputusan.



Gambar 3.4 Lift Heatmap

Gambar ini adalah heatmap (peta warna) yang menggambarkan nilai lift dari setiap aturan asosiasi yang ditemukan.

- Lift = 1 → item dalam rule terjadi secara acak (tidak ada hubungan khusus).
- Lift > 1 → terjadi lebih sering dari ekspektasi → ada hubungan positif antara kondisi dan hasil.
- Lift < 1 → terjadi lebih jarang dari ekspektasi → kemungkinan hubungan negatif.

Fungsi Lift Heatmap:

1. Memberi gambaran seberapa kuat hubungan antar variabel, dibandingkan dengan jika mereka terjadi secara kebetulan.
2. Dalam heatmap, nilai lift tinggi bisa ditandai warna merah tua/gelap, sedangkan nilai rendah dengan warna lebih terang.

Interpretasi:

1. Rule dengan lift tinggi (misalnya pendidikan rendah → tanah kecil dengan lift 1.8) lebih kuat dan signifikan.
2. Heatmap ini membantu memprioritaskan rule untuk kebijakan atau intervensi sosial, berdasarkan kekuatan hubungan data.

3.3 Diskusi Kebijakan

- a. Ketimpangan kepemilikan tanah sangat terkait dengan akses pendidikan (pola 1)
- b. Status kepemilikan yang tidak jelas (pola 2) menjadi penghambat program redistribusi lahan
- c. Potensi pengembangan agroindustri skala kecil di wilayah pedesaan (pola 3)

3.4 Keterbatasan Penelitian

- a. Data hanya mencakup periode 2019-2022
- b. Variabel budaya lokal belum dipertimbangkan secara mendalam
- c. Pendekatan kuantitatif membutuhkan studi kualitatif lanjutan

Hasil ini menunjukkan bahwa terdapat perbedaan signifikan dalam kondisi sosial ekonomi antar kluster, yang dapat menjadi dasar bagi pemerintah untuk merumuskan kebijakan yang lebih tepat sasaran.

4. Kesimpulan

Penelitian ini berhasil menerapkan algoritma K-Means dan Apriori untuk menganalisis kemiskinan dan pola kepemilikan tanah di Indonesia. Dari analisis yang dilakukan, ditemukan bahwa: Kluster Kemiskinan: Tiga kluster utama teridentifikasi berdasarkan tingkat kemiskinan, yaitu kluster miskin ekstrem, kluster miskin sedang, dan

kluster relatif sejahtera. Setiap kluster memiliki karakteristik sosial ekonomi yang berbeda, yang mencerminkan ketimpangan dalam akses terhadap pendidikan, kesehatan, dan sumber daya. Pola Asosiasi Kepemilikan Tanah: Analisis menggunakan algoritma Apriori mengungkapkan pola asosiasi signifikan antara kepemilikan tanah dan karakteristik sosial ekonomi. Terdapat hubungan yang kuat antara pendidikan, akses sanitasi, dan status kepemilikan tanah, yang menunjukkan bahwa pendidikan dan akses terhadap layanan dasar berpengaruh terhadap kepemilikan tanah. Implikasi Kebijakan: Temuan ini memberikan wawasan penting bagi pembuat kebijakan untuk merumuskan program-program yang lebih efektif dalam pengentasan kemiskinan dan reformasi kepemilikan tanah. Pendekatan berbasis data ini dapat membantu pemerintah dalam menargetkan intervensi sosial yang lebih tepat sasaran.

Referensi

- Bahauddin, A., Fatmawati, A., & Sari, F. P. (2021). Analisis Clustering Provinsi di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means. *Jurnal Manajemen Informatika Dan Sistem Informasi*, 4(1), 1–8.
- Fadillah, N., & Wicaksono, B. S. (2023). IMPLEMENTASI ALGORITMA APRIORI PADA DATA MINING UNTUK MENENTUKAN STRATEGI PENJUALAN DENGAN MENGGUNAKAN DATA TRANSAKSI. *OKTAL: Jurnal Ilmu Komputer dan Sains*, 2(03), 882-889.
- Widodo, M., 2016. Data Mining Penjualan Barang Bengkel Menggunakan Metode Apriori. *Jurnal Media Infotama*, 11(2), pp. 54-60.
- I.H. Witten, E.F. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*, 1st ed. Todd Green.
- Santoso, M.H. (2021). Penerapan Metode Aturan Asosiasi Menggunakan Algoritma Apriori untuk Menemukan Pola Penjualan: Studi Kasus di Indomaret Tanjung Anom. 4-5.
- Feby, Dita. “Apa Itu Decision Tree di Machine Learning Model?” <https://dqlab.id>. Diakses pada 14 Januari 2024.
- Kurniawan D. (2020). “Pengenalan Machine Learning dengan Python”. Jakarta : PT Elex Media Komputindo.
- Wadhwatanya. (2021). “Splitting Criteria and Algorithms in Decision Tree” wadhwaatanya1234.medium.com. Diakses pada 17 Januari 2024.
- Syukri Mustafa, M., Rizky Ramadhan, M., & Thenata, A. P. (2017). Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Citec Journal*, 4(2), 151–162.
- Rifqo, M. H., & Wijaya, A. (2017). Implementasi Algoritma Naive Bayes Dalam Penentuan Pemberian Kredit. *Pseudocode*, 4(2), 120–128.
- Muhamad, H., Prasajo, C. A., Sugianto, N. A., Surtiningsih, L., & Cholissodin, I. (2017). Optimasi Naive Bayes Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 4(3), 180.