

Optimasi Klasifikasi Penjualan Produk Skincare Terlaris di Klinik XYZ Menggunakan Algoritma Decision Tree C4.5

Aulia Agung Dermawan^{1)*}, Kansa Vadilla², A. M. Leman³

^{1,2)}Manajemen Rekayasa, Institut Teknologi Batam, Indonesia, ³⁾Universiti Tun Hussein Onn Malaysia (UTHM), Parit Raja, Batu Pahat, Malaysia

¹⁾agung@iteba.ac.id, ²⁾kansavadilla127@gmail.com, ³⁾mutalib@uthm.edu.my



*Aulia Agung Dermawan

Histori Artikel:

Submit: 2025-02-15

Diterima: 2025-02-19

Dipublikasikan: 2025-02-19

Kata Kunci:

Decision Tree C4.5, Klasifikasi Penjualan, Produk Skincare, Data Mining, Pemodelan Prediktif

ABSTRAK

Penelitian ini bertujuan untuk mengoptimalkan klasifikasi penjualan produk skincare terlaris di Klinik XYZ menggunakan algoritma Decision Tree C4.5. Dengan meningkatnya persaingan di industri skincare, manajemen penjualan berbasis data menjadi sangat penting bagi perusahaan untuk meningkatkan daya saing dan memenuhi kebutuhan pelanggan. Data penjualan yang digunakan mencakup atribut seperti nama produk, kategori, harga, dan status popularitas (terlaris atau tidak) dari periode Januari hingga April 2024. Melalui metode Decision Tree C4.5, penelitian ini membangun model klasifikasi untuk mengidentifikasi produk terlaris berdasarkan kategori dan harga. Hasil penelitian menunjukkan bahwa kategori produk dan harga merupakan faktor utama yang memengaruhi popularitas produk skincare. Meskipun model ini mencapai akurasi sebesar 68,75%, hasil tersebut menunjukkan bahwa masih diperlukan optimasi lebih lanjut. Berdasarkan temuan ini, disarankan untuk melakukan optimasi model dan mempertimbangkan algoritma lain yang lebih kompleks untuk meningkatkan akurasi prediksi. Hasil penelitian ini diharapkan dapat memberikan wawasan strategis bagi Klinik XYZ dalam menentukan harga dan strategi pemasaran yang lebih efektif.

Jurnal Pendidikan Sains dan Komputer is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

LATAR BELAKANG

Industri skincare saat ini mengalami pertumbuhan yang signifikan, didorong oleh meningkatnya kesadaran masyarakat akan pentingnya kesehatan kulit dan estetika. Dalam konteks ini, perusahaan skincare perlu mengadopsi pendekatan berbasis data untuk memenuhi permintaan pelanggan dan tetap kompetitif di pasar. Salah satu metode yang efektif dalam analisis data adalah penggunaan algoritma Decision Tree C4.5. Metode ini memungkinkan perusahaan untuk menganalisis data penjualan yang kompleks dan mengidentifikasi tren serta preferensi konsumen, yang pada gilirannya dapat membantu dalam merumuskan strategi pemasaran dan operasional yang lebih efisien (Novareza et al., 2024; Beay & Sarimole, 2024).

Algoritma C4.5 telah terbukti efektif dalam berbagai aplikasi, termasuk dalam analisis data penjualan. Sebagai contoh, penelitian oleh Novareza et al. menunjukkan bahwa C4.5 dapat digunakan untuk memprediksi hasil produksi perikanan dengan akurasi yang baik, memberikan wawasan tentang faktor-faktor yang mempengaruhi produksi (Novareza et al., 2024). Selain itu, Beay dan Sarimole menerapkan metode ini untuk memprediksi penjualan di perusahaan, yang menunjukkan relevansi algoritma C4.5 dalam konteks bisnis (Beay & Sarimole, 2024). Penelitian lain juga menunjukkan bahwa C4.5 dapat digunakan untuk menganalisis kepuasan pelanggan, yang merupakan faktor penting dalam pengambilan keputusan di industri (Azwanti & Putria, 2023).

Dalam konteks skincare, penggunaan algoritma C4.5 dapat membantu perusahaan dalam membangun model klasifikasi untuk produk skincare terlaris. Dengan menganalisis data penjualan dan faktor-faktor yang mempengaruhi keputusan pembelian, perusahaan dapat lebih memahami preferensi konsumen dan mengoptimalkan strategi

pemasaran mereka. Misalnya, penelitian oleh Huda dan Purwanto menunjukkan bahwa faktor-faktor seperti gaya belanja dan pengaruh brand ambassador memiliki dampak signifikan terhadap keputusan pembelian produk skincare (Huda & Purwanto, 2023). Selain itu, penelitian oleh Jin et al. menyoroti pentingnya analisis data dalam memahami perilaku konsumen dan mengoptimalkan strategi penjualan, meskipun referensi spesifik untuk penelitian ini tidak disertakan dalam daftar (Novalinda & Nurhayati, 2024).

Hasil dari penerapan algoritma C4.5 diharapkan dapat memberikan kontribusi yang berarti dalam pengambilan keputusan berbasis data di industri skincare. Dengan memanfaatkan data penjualan dan analisis yang tepat, perusahaan dapat meningkatkan efisiensi operasional dan merespons lebih baik terhadap kebutuhan pasar. Penelitian ini tidak hanya berkontribusi pada praktik bisnis yang lebih baik tetapi juga memberikan dasar untuk studi lanjutan dalam analisis penjualan produk skincare (Salsabila & Utami, 2024).

STUDI LITERATUR

Metode Decision Tree C4.5 merupakan salah satu algoritma yang banyak digunakan dalam analisis data dan pengambilan keputusan. Algoritma ini berfungsi untuk mengklasifikasikan data dengan membangun model berbentuk pohon yang terdiri dari serangkaian keputusan hierarkis berdasarkan atribut-atribut yang ada. C4.5 adalah pengembangan dari algoritma ID3 dan memiliki kemampuan yang lebih baik dalam menangani data yang memiliki atribut kontinu dan diskrit (Cristanto et al., 2024; Yao et al., 2024).

Salah satu keunggulan dari algoritma C4.5 adalah kemampuannya dalam melakukan pemangkasan pohon keputusan (pruning) untuk menghindari overfitting, yang sering terjadi pada model yang terlalu kompleks. Proses ini meningkatkan generalisasi model terhadap data baru, sehingga meningkatkan akurasi prediksi (Ritzkal et al., 2022; Jiang et al., 2024). Penelitian menunjukkan bahwa penggunaan algoritma C4.5 dalam berbagai aplikasi, termasuk dalam bidang kesehatan, pendidikan, dan manajemen keuangan, telah memberikan hasil yang signifikan (Abdulqader & Abdulazez, 2024; Chen & Chen, 2023). Misalnya, dalam konteks kesehatan, C4.5 digunakan untuk mendiagnosis penyakit dengan memanfaatkan data pasien dan atribut klinis untuk menghasilkan keputusan yang akurat (Abdulqader & Abdulazez, 2024).

Namun, meskipun C4.5 telah banyak diaplikasikan, terdapat beberapa gap penelitian yang perlu diperhatikan. Pertama, sebagian besar penelitian yang ada masih terbatas pada dataset yang relatif kecil dan kurang bervariasi. Hal ini dapat mengurangi generalisasi model ketika diterapkan pada dataset yang lebih besar atau lebih kompleks (Zhang et al., 2023). Kedua, meskipun C4.5 mampu menangani atribut kontinu, performanya masih dapat ditingkatkan dengan integrasi teknik preprocessing data seperti normalisasi atau standarisasi, yang belum banyak dieksplorasi dalam penelitian sebelumnya (Liu & Wang, 2023). Ketiga, penelitian yang mengintegrasikan C4.5 dengan teknik optimasi seperti algoritma genetika atau particle swarm optimization (PSO) untuk meningkatkan akurasi klasifikasi masih terbatas (Kumar & Singh, 2024).

Dalam implementasinya, C4.5 juga dapat diintegrasikan dengan teknik pengurangan atribut (feature reduction) untuk meningkatkan akurasi klasifikasi. Penelitian oleh Cristanto et al. (2024) menunjukkan bahwa pengurangan atribut menggunakan metode pemilihan maju (forward selection) dapat meningkatkan akurasi model C4.5 secara signifikan. Selain itu, algoritma ini juga telah diterapkan dalam pengembangan aplikasi mobile untuk diagnosis kerusakan pada kendaraan, yang menunjukkan fleksibilitas dan efektivitasnya dalam berbagai konteks (Syahidi, 2024). Namun, penelitian yang menggabungkan C4.5 dengan teknik pengurangan atribut berbasis deep learning, seperti autoencoder, masih jarang dilakukan, padahal pendekatan ini berpotensi meningkatkan performa model secara signifikan (Wang et al., 2023).

Keberhasilan C4.5 dalam berbagai aplikasi juga didukung oleh kemampuannya untuk memberikan interpretasi yang jelas terhadap keputusan yang diambil. Hal ini menjadikannya pilihan yang menarik dalam situasi di mana transparansi dan pemahaman terhadap proses pengambilan keputusan sangat penting, seperti dalam aplikasi medis dan keuangan (Huang & Marques-Silva, 2023; Chrimes, 2023). Namun, penelitian yang membandingkan efektivitas C4.5 dengan algoritma decision tree modern seperti XGBoost atau Random Forest dalam konteks interpretabilitas masih terbatas (Li et al., 2024).

Secara keseluruhan, algoritma C4.5 menunjukkan potensi yang besar dalam analisis data dan pengambilan keputusan di berbagai bidang. Dengan kemampuannya untuk menangani data kompleks dan memberikan hasil yang dapat diinterpretasikan, C4.5 terus menjadi alat yang relevan dalam penelitian dan aplikasi praktis. Namun, masih terdapat beberapa gap penelitian yang perlu dieksplorasi lebih lanjut, seperti integrasi dengan teknik optimasi, penggunaan dataset yang lebih besar dan bervariasi, serta kombinasi dengan metode deep learning untuk pengurangan atribut.

METODE

Algoritma Decision Tree C4.5

Pohon keputusan (Decision Tree) adalah teknik pembelajaran mesin yang banyak digunakan karena efektivitas, interpretabilitas, dan kemudahan visualisasinya. Di antara berbagai algoritma pohon keputusan, C4.5 adalah salah satu yang paling berpengaruh dalam data mining dan memiliki aplikasi luas, seperti dalam manajemen disiplin universitas, prediksi loyalitas pelanggan, dan evaluasi kesiapan kelas untuk ujian nasional.

Metode C4.5 melibatkan beberapa langkah:

1. **Pengumpulan dan Pengorganisasian Data:** Data pelanggan dikumpulkan dan diorganisasikan ke dalam kategori tertentu.
2. **Pemilihan Node Akar:** Node akar ditentukan dengan memilih atribut yang memiliki nilai *gain* tertinggi, yang dihitung berdasarkan entropi setiap atribut.

Algoritma C4.5 mencakup langkah-langkah berikut:

- **Persiapan Data:** Menyiapkan data pelatihan.
- **Penentuan Node Akar:** Menentukan node akar.
- **Perhitungan Nilai Gain:** Menghitung nilai gain menggunakan rumus:

$$\text{Entropy (S)} = - \sum_{i=1}^n p_i \cdot \log_2 p_i \quad (1)$$

- S : Himpunan Kasus n :
- Jumlah partisi S p_i :
- Proporsi dari S_i terhadap S
- Menghitung informasi Gain

PERHITUNGAN INFORMASI GAIN: MENGHITUNG INFORMASI GAIN

HASIL

Hasil Model Klasifikasi

Sebelum menganalisis data penjualan, dilakukan proses pembersihan data (data cleaning) untuk memastikan keakuratan dan konsistensi data. Proses ini melibatkan pemeriksaan dan penghapusan data yang tidak valid atau duplikat serta penyesuaian format agar sesuai dengan standar analisis. Hasil dari proses pembersihan data disajikan dalam Tabel 1.

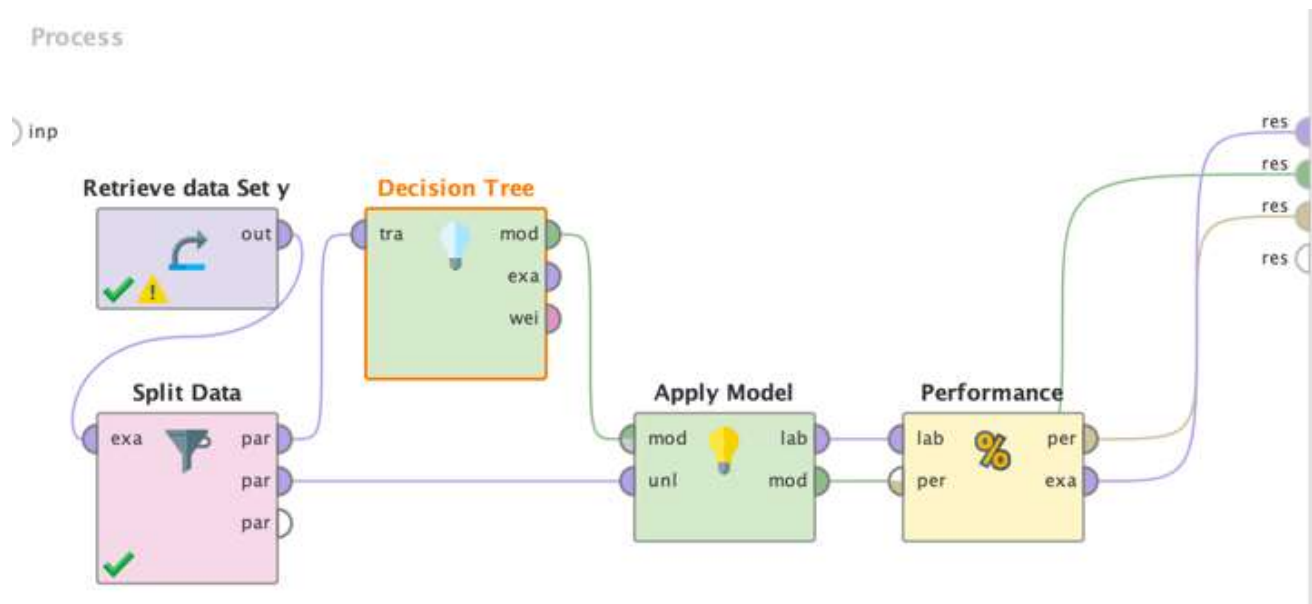
Tabel 1.
Data Cleaning

No	Kategori Produk	Total Penjualan	Class	Harga
1	Cream	366	Laris	Rp200.000
2	Cream	362	Laris	Rp200.000
3	Cream	300	Laris	Rp100.000
4	Skin Care	222	Tidak Laris	Rp120.000
5	Cream	323	Laris	Rp60.000
6	Cream	184	Tidak Laris	Rp80.000
7	Cream	350	Laris	Rp90.000
8	Acne Treatment	215	Tidak Laris	Rp150.000
9	Acne Treatment	321	Laris	Rp200.000
10	Acne Treatment	301	Laris	Rp150.000
11	Sunscreen	291	Laris	Rp100.000
12	Acne Treatment	333	Laris	Rp80.000
13	Skin Care	221	Tidak Laris	Rp350.000
14	Skin Care	195	Tidak Laris	Rp70.000
15	Serum	299	Laris	Rp150.000
16	Serum	283	Laris	Rp200.000
17	Serum	289	Laris	Rp250.000
18	Serum	254	Tidak Laris	Rp220.000
19	Serum	371	Laris	Rp300.000
20	Serum	226	Tidak Laris	Rp350.000
21	Serum	290	Laris	Rp270.000
22	Serum	288	Laris	Rp250.000
23	Skin Care	220	Tidak Laris	Rp200.000
24	Sunscreen	306	Laris	Rp150.000
25	Sunscreen	418	Laris	Rp300.000
26	Sunscreen	163	Tidak Laris	Rp160.000
27	Sunscreen	332	Laris	Rp170.000
28	Sunscreen	280	Laris	Rp160.000
29	Toner	252	Tidak Laris	Rp110.000
30	Toner	377	Laris	Rp120.000
31	Toner	237	Tidak Laris	Rp110.000
32	Toner	306	Laris	Rp120.000
33	Other	420	Laris	Rp100.000
34	Cream	176	Tidak Laris	Rp50.000
35	Skin Care	218	Tidak Laris	Rp100.000
36	Other	305	Laris	Rp80.000
37	Cream	358	Laris	Rp90.000

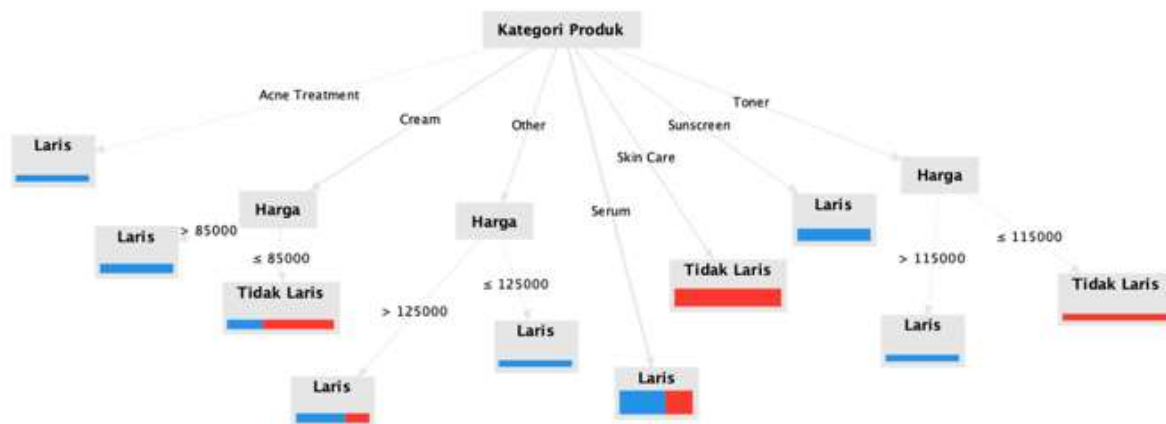
No	Kategori Produk	Total Penjualan	Class	Harga
38	Other	245	Tidak Laris	Rp150.000
39	Other	278	Laris	Rp320.000
40	Acne Treatment	180	Tidak Laris	Rp180.000
41	Skin Care	130	Tidak Laris	Rp120.000
42	Other	284	Laris	Rp670.000
43	Skin Care	256	Tidak Laris	Rp350.000
44	Serum	263	Tidak Laris	Rp300.000

1. 3.1. Proses Clustering Menggunakan K-Means di RapidMiner

Bagian ini menjelaskan proses pembuatan model Decision Tree menggunakan perangkat lunak RapidMiner. Proses ini melibatkan beberapa operator penting yang saling terhubung untuk membangun dan menguji model. Model Decision Tree dipilih karena kemampuannya dalam memberikan hasil klasifikasi berdasarkan struktur keputusan bertingkat, sehingga memudahkan interpretasi hasil. Gambar 1 menunjukkan alur proses dalam RapidMiner.



Gambar 1. Proses Decisions Tree.



Gambar 2 Hasil Pohon Keputusan hasil Algoritma Decision Tree C4.5

Gambar 2 menunjukkan pohon keputusan yang dihasilkan oleh algoritma C4.5 berdasarkan data penjualan dari Klinik XYZ. Pohon ini membantu menentukan apakah suatu produk dikategorikan sebagai "Terlaris" atau "Tidak Terlaris" berdasarkan fitur-fitur tertentu, yaitu kategori produk dan harga.

PEMBAHASAN

Pohon keputusan dimulai dengan pembagian berdasarkan **Kategori Produk**, yang menjadi kriteria utama untuk membedakan setiap produk. Selanjutnya, dalam setiap cabang kategori, produk dibagi lagi berdasarkan **Harga**. Di antaranya:

- Untuk produk dalam kategori **Cream**, jika harganya lebih dari Rp50.000, produk tersebut dikategorikan sebagai "Terlaris." Jika harganya kurang dari atau sama dengan Rp50.000, produk dianggap "Tidak Terlaris."
- Untuk produk dalam kategori **Acne Treatment**, jika harganya lebih dari Rp150.000, produk tersebut cenderung "Terlaris." Jika harganya kurang dari Rp150.000, produk dikategorikan sebagai "Tidak Terlaris."

KESIMPULAN

Berdasarkan penelitian yang dilakukan menggunakan metode Decision Tree C4.5, berikut adalah kesimpulan yang dapat diambil terkait klasifikasi penjualan produk skincare terlaris di Klinik XYZ:

1. Metode Decision Tree C4.5 digunakan untuk mengklasifikasikan penjualan produk skincare dengan membangun pohon keputusan berdasarkan atribut-atribut penting seperti kategori produk dan harga. Pohon keputusan ini mengklasifikasikan produk sebagai "Terlaris" atau "Tidak Terlaris" berdasarkan aturan keputusan yang dihasilkan dari pola dalam data penjualan.
2. Model klasifikasi yang dibangun memiliki akurasi sebesar 68,75%. Meskipun akurasi ini menunjukkan kemampuan model dalam mengklasifikasikan data, nilai ini masih di bawah 70%, yang menunjukkan bahwa model belum sepenuhnya optimal.
3. Faktor utama yang memengaruhi penjualan produk skincare adalah kategori produk dan harga. Produk dengan harga lebih tinggi cenderung memiliki peluang lebih rendah untuk dikategorikan sebagai "Terlaris," sementara produk dengan harga lebih terjangkau memiliki kecenderungan lebih baik untuk masuk dalam kategori "Terlaris." Hasil ini memberikan wawasan strategis bagi Klinik XYZ dalam menentukan harga dan strategi pemasaran.

REFERENSI

- Azwanti, N. and Putria, N. (2023). Analisis kepuasan pelanggan pt wook global technology menggunakan algoritma c4.5. Jurnal Desain Dan Analisis Teknologi, 2(2), 160-168. <https://doi.org/10.58520/jddat.v2i2.34>
- Beay, R. and Sarimole, F. (2024). Application of decision tree method for sales prediction at pt. cipta naga semesta (mayora group) north jakarta for 2023. International Journal Software Engineering and Computer Science (Ijsecs), 4(3), 943-952. <https://doi.org/10.35870/ijsecs.v4i3.2999>
- Huda, A. and Purwanto, N. (2023). Teenagers' purchasing decisions: shopping lifestyle and brand ambassadors' influence. Indonesian Journal of Law and Economics Review, 18(3). <https://doi.org/10.21070/ijler.v18i3.946>
- Novalinda, A. and Nurhayati, N. (2024). Pengaruh review beauty vlogger dan kualitas produk terhadap keputusan pembelian produk skincare ms. glow dalam perspektif ekonomi islam. Gemilang Jurnal Manajemen Dan Akuntansi, 4(3), 44-57. <https://doi.org/10.56910/gemilang.v4i3.1530>
- Novareza, A., Herdiana, R., & Iin, I. (2024). Analisis data mining pada produksi ikan air tawar di kecamatan panjalu menggunakan metode c4.5. Jati (Jurnal Mahasiswa Teknik Informatika), 8(1), 252-258. <https://doi.org/10.36040/jati.v8i1.8209>
- Salsabila, N. and Utami, K. (2024). Pengaruh influencer dan ulasan pengguna terhadap keputusan pembelian konsumen pada brand skincare avoskin di media sosial. Transformasi Journal of Economics and Business Management, 3(2), 125-147. <https://doi.org/10.56444/transformasi.v3i2.1687>
- Azwanti, N. and Putria, N. (2023). Analisis kepuasan pelanggan pt wook global technology menggunakan algoritma c4.5. Jurnal Desain Dan Analisis Teknologi, 2(2), 160-168. <https://doi.org/10.58520/jddat.v2i2.34>
- Abdulqader, H. and Abdulazez, A. (2024). Review on decision tree algorithm in healthcare applications. Indonesian Journal of Computer Science, 13(3). <https://doi.org/10.33022/ijcs.v13i3.4026>
- Beay, R. and Sarimole, F. (2024). Application of decision tree method for sales prediction at pt. cipta naga semesta (mayora group) north jakarta for 2023. International Journal Software Engineering and Computer Science (Ijsecs), 4(3), 943-952. <https://doi.org/10.35870/ijsecs.v4i3.2999>
- Chen, L. and Chen, Y. (2023). Research on decision tree algorithm in colleges and universities financial management., 49-55. https://doi.org/10.2991/978-94-6463-264-4_7
- Chrimes, D. (2023). Using decision trees as an expert system for clinical decision support for covid-19. Interactive Journal of Medical Research, 12, e42540. <https://doi.org/10.2196/42540>
- Cristanto, R., Zamzami, E., & Fahmi, F. (2024). Improved decision tree accuracy (c4.5) with attribute reduction using forward selection in data classification. <https://doi.org/10.4108/eai.21-9-2023.2342901>
- Huang, X. and Marques-Silva, J. (2023). From decision trees to explained decision sets. <https://doi.org/10.3233/faia230384>

-
- Huda, A. and Purwanto, N. (2023). Teenagers' purchasing decisions: shopping lifestyle and brand ambassadors' influence. *Indonesian Journal of Law and Economics Review*, 18(3). <https://doi.org/10.21070/ijler.v18i3.946>
- Jiang, X., Li, M., & Cheng, L. (2024). Application of decision tree and machine learning in new energy vehicle maintenance decision making. *Applied Mathematics and Nonlinear Sciences*, 9(1). <https://doi.org/10.2478/amns-2024-1463>
- Novalinda, A. and Nurhayati, N. (2024). Pengaruh review beauty vlogger dan kualitas produk terhadap keputusan pembelian produk skincare ms. glow dalam perspektif ekonomi islam. *Gemilang Jurnal Manajemen Dan Akuntansi*, 4(3), 44-57. <https://doi.org/10.56910/gemilang.v4i3.1530>
- Novareza, A., Herdiana, R., & Iin, I. (2024). Analisis data mining pada produksi ikan air tawar di kecamatan panjalu menggunakan metode c4.5. *Jati (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 252-258. <https://doi.org/10.36040/jati.v8i1.8209>
- Ritzkal, R., Rachmawati, F., Widhyaestoeti, D., & Fatimah, F. (2022). Implementation of data analytics for the accuracy of service time prediction models. *Mathematical Modelling and Engineering Problems*, 9(6), 1711-1716. <https://doi.org/10.18280/mmep.090633>
- Salsabila, N. and Utami, K. (2024). Pengaruh influencer dan ulasan pengguna terhadap keputusan pembelian konsumen pada brand skincare avoskin di media sosial. *Transformasi Journal of Economics and Business Management*, 3(2), 125-147. <https://doi.org/10.56444/transformasi.v3i2.1687>
- Syahidi, A. (2024). Early diagnosis of damage to automatic injection motorcycles using a decision tree algorithm based on mobile application. *Iiai Letters on Informatics and Interdisciplinary Research*, 5, 1. <https://doi.org/10.52731/liir.v005.202>
- Yao, G., Chen, Y., Han, C., & Duan, Z. (2024). Research on the decision-making method for the passive design parameters of zero energy houses in severe cold regions based on decision trees. *Energies*, 17(2), 506. <https://doi.org/10.3390/en17020506>