



E-ISSN 3032-601X & P-ISSN 3032-7105

Vol. 3, No. 1, 2026

MISTER

**Journal of Multidisciplinary Inquiry in Science,
Technology and Educational Research**

**Jurnal Penelitian Multidisiplin dalam Ilmu
Pengetahuan, Teknologi dan Pendidikan**

**UNIVERSITAS SERAMBI MEKKAH
KOTA BANDA ACEH**

mister@serambimekkah.ac.id

Journal of Multidisciplinary Inquiry in Science Technology
and Educational Research

Journal of MISTER

Vol. 3, No. 1, 2026

Pages: 1515–1525

Klasifikasi Tipe Konsumen Retail Supermarket Menggunakan *Decision Tree* Berdasarkan Data Transaksi

I Gusti Ngurah Adi Prayoga, I Wayan Sudiarsa, Ni Kadek Nila Agustini
Dewi, Kadek Dio Angginantara Putra, Muhammad Iqbal

Institut Bisnis dan Teknologi Indonesia Denpasar

Article in Journal of MISTER

Available at : <https://jurnal.serambimekkah.ac.id/index.php/mister>

DOI : <https://doi.org/10.32672/mister.v3i1.4106>

How to Cite this Article

APA : Adi Prayoga, I. G. N., Sudiarsa, I. W., Dewi, N. K. N. A., Putra, K. D. A., & Iqbal, M. (2026). Klasifikasi Tipe Konsumen Retail Supermarket Menggunakan Decision Tree Berdasarkan Data Transaksi. Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research, 3(1), 1515 – 1525. <https://doi.org/10.32672/mister.v3i1.4106>

Others Visit : <https://jurnal.serambimekkah.ac.id/index.php/mister>

MISTER: Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research is a scholarly journal dedicated to the exploration and dissemination of innovative ideas, trends and research on the various topics include, but not limited to functional areas of Science, Technology, Education, Humanities, Economy, Art, Health and Medicine, Environment and Sustainability or Law and Ethics.

MISTER: Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research is an open-access journal, and users are permitted to read, download, copy, search, or link to the full text of articles or use them for other lawful purposes. Articles on Journal of MISTER have been previewed and authenticated by the Authors before sending for publication. The Journal, Chief Editor, and the editorial board are not entitled or liable to either justify or responsible for inaccurate and misleading data if any. It is the sole responsibility of the Author concerned.



ISSN 3032-7105

ISSN 3032-601X



Klasifikasi Tipe Konsumen Retail Supermarket Menggunakan Decision Tree Berdasarkan Data Transaksi

I Gusti Ngurah Adi Prayoga^{1*}, I Wayan Sudiarsa², Ni Kadek Nila Agustini Dewi³, Kadek
Dio Angginantara Putra⁴, Muhammad Iqbal⁵

Rekayasa Sistem Komputer, Informatika, Institut Bisnis dan Teknologi Indonesia (INSTIKI),
Denpasar/Denpasar Selatan, Indonesia^{1,2,3,4,5}

*Email Korespondensi: sudiarsa@instiki.ac.id

Diterima: 29-12-2025

| Disetujui: 09-01-2026

| Diterbitkan: 11-01-2026

ABSTRACT

The goal of this study is to use a Decision Tree algorithm to group different types of shoppers in a retail supermarket setting. The dataset used in this study came from a public Kaggle repository and has 1,000 transaction records. These records include information about the customers, the types of products they bought, the details of the transactions, and the time of day they took place. We used Python-based libraries like pandas and scikit-learn in Google Colab to process and analyze the data. The research method includes preprocessing the data, creating new features, encoding categorical variables, splitting the data, training the model, and evaluating it. We used accuracy, precision, recall, and F1-score to measure how well the model worked. The experimental results show that the Decision Tree model was about 56.5% accurate when using all features and 61.5% accurate when using only a few key features with the tree depth set to a certain level. These results show that it's possible to group customers based on their transactional and demographic information, but the model's performance could still be better. This study helps us understand how to segment customers in retail analytics and gives us a starting point for future improvements using more advanced machine learning methods.

Keywords: consumer classification 1; decision tree 2; retail analytics 3; machine learning 4; data mining 5

ABSTRAK

Tujuan dari penelitian ini adalah untuk menggunakan algoritma Pohon Keputusan untuk mengelompokkan berbagai jenis pelanggan di lingkungan ritel supermarket. Dataset yang digunakan berasal dari repositori publik Kaggle yang dipublikasikan oleh Fares Ashraf, dan berisi 1.000 data transaksi yang mencakup informasi demografi pelanggan, kategori produk, detail transaksi, dan informasi waktu. Google Colab, bersama dengan pustaka Python seperti pandas dan scikit-learn, digunakan untuk memproses dan menganalisis data. Metodologi penelitian mencakup pra-pemrosesan data, rekayasa fitur, pengkodean variabel kategorikal, pemisahan data, pelatihan model, dan evaluasi model. Kinerja model diukur berdasarkan akurasi, presisi, recall, dan skor F1-nya. Hasil eksperimen menunjukkan bahwa model Decision Tree menghasilkan akurasi sekitar 56,5% pada skenario fitur lengkap dan meningkat menjadi 61,5% ketika menggunakan fitur-fitur utama dengan kedalaman pohon yang dibatasi. Hasilnya menunjukkan bahwa mengklasifikasikan jenis konsumen berdasarkan atribut transaksi dan demografi adalah layak, meskipun kinerja model masih dapat ditingkatkan. Studi ini membantu kita memahami cara melakukan segmentasi pelanggan dalam analisis ritel.

Katakunci klasifikasi konsumen 1; decision tree 2; analitik ritel 3; machine learning 4; data mining 5

PENDAHULUAN

Analisis data kini menjadi alat penting di banyak bidang, terutama ritel. Hal ini karena teknologi informasi berkembang sangat pesat dan cara menyimpan serta menangani data semakin baik. Setiap hari, supermarket mendapatkan banyak informasi tentang pembelian yang dilakukan pelanggan. Informasi ini tidak hanya melacak penjualan, tetapi juga menunjukkan bagaimana perilaku pelanggan, produk apa yang mereka sukai, seberapa sering mereka membelinya, dan bagaimana mereka menggunakannya. Dengan melihat informasi ini, bisnis dapat membuat pilihan yang lebih baik berdasarkan data.

Dalam industri ritel yang penuh persaingan, pemahaman mengenai motivasi pelanggan untuk melakukan pembelian menjadi sangat krusial. Untuk mencapai hal ini, analisis ritel sering dilakukan dengan metode pengelompokan pelanggan ke dalam kategori berdasarkan karakteristik dan perilaku serupa. Melalui segmentasi ini, bisnis dapat mengembangkan strategi pemasaran yang lebih terarah, yang dapat meningkatkan kepuasan dan loyalitas pelanggan, serta efisiensi operasional secara keseluruhan. Biasanya, pelanggan dikelompokkan berdasarkan demografi, kebiasaan pembelian, serta nilai transaksi. Hal ini tidak hanya membantu dalam memahami pola konsumsi, tetapi juga memberikan gambaran seberapa besar kontribusi masing-masing pelanggan terhadap pendapatan bisnis (Almagribi 2025)

Dalam konteks analisis pelanggan, berbagai penelitian telah mengaplikasikan penambahan data dan pembelajaran mesin untuk pengelompokan pelanggan ritel. Algoritma yang umum digunakan mencakup Naive Bayes, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), dan Pohon Keputusan. Di antara algoritma-algoritma ini, Pohon Keputusan dianggap populer karena kemudahan pemahaman dan kemampuannya untuk merepresentasikan hubungan yang kompleks dalam data tanpa memerlukan asumsi yang ketat (Mukhlisin & Nugroho, 2025). Hal ini memungkinkan manajemen untuk lebih memahami faktor-faktor yang mempengaruhi hasil klasifikasi dan pengambilan keputusan dalam konteks bisnis.

Banyak peneliti baru-baru ini mulai melihat bagaimana atribut temporal, seperti hari, bulan, dan periode waktu tertentu, dapat memberikan wawasan tambahan yang berguna dalam memahami perilaku pelanggan. Namun, ada juga tantangan dalam menggunakan semua fitur yang ada. Model yang didasarkan pada fitur yang terlalu banyak tanpa strategi seleksi yang tepat dapat menjadi sulit untuk dipahami. Ini menjadi perhatian utama, terutama ketika hasil dari analisis tersebut perlu disajikan secara jelas kepada pemangku kepentingan (Sousa et al., 2023).

Penelitian ini bertujuan untuk memanfaatkan algoritma Pohon Keputusan untuk mengklasifikasikan pelanggan supermarket ke dalam dua kategori: anggota dan pelanggan tetap. Pendekatan ini dianggap signifikan karena mengintegrasikan berbagai atribut transaksi yang komprehensif, termasuk elemen temporal yang diambil dari stempel waktu transaksi. Selain itu, penelitian ini menerapkan dua metodologi pemodelan: yang pertama menggunakan semua fitur untuk menangkap sebanyak mungkin informasi, sedangkan yang kedua melakukan seleksi fitur untuk menciptakan model yang lebih mudah dipahami dan diinterpretasikan (Cant & Toit, 2012; Mukhlisin & Nugroho, 2025).

METODE PENELITIAN

Gambaran Umum Metode Penelitian :

Penelitian ini menggunakan pendekatan klasifikasi data dengan algoritma Decision Tree untuk mengelompokkan tipe konsumen retail supermarket ke dalam dua kelas, yaitu member dan normal. Pendekatan ini dipilih karena Decision Tree mampu menghasilkan aturan keputusan yang mudah dipahami dan diinterpretasikan, sehingga sesuai untuk analisis perilaku konsumen ritel.

Secara umum, tahapan penelitian dimulai dari pengambilan data transaksi supermarket, dilanjutkan dengan pra-pemrosesan data, rekayasa fitur, pembagian data latih dan data uji, pemodelan menggunakan algoritma Decision Tree, serta evaluasi dan interpretasi hasil klasifikasi. Alur metodologi penelitian secara keseluruhan ditunjukkan pada Gambar 1.

Sumber Data :

Dataset yang digunakan dalam penelitian ini adalah dataset “Supermarket Sales” yang dipublikasikan oleh Fares Ashraf dan tersedia secara publik melalui platform Kaggle pada tautan berikut: <https://www.kaggle.com/datasets/faresashraf1001/supermarket-sales/data>

Dataset ini berisi 1.000 data transaksi penjualan supermarket dengan 18 atribut yang merepresentasikan informasi transaksi, demografi pelanggan, serta waktu pembelian. Atribut-atribut utama dalam dataset meliputi:

- Invoice ID
- Branch
- City
- Customer type
- Gender
- Product line
- Unit price
- Quantity
- Tax 5%
- Sales
- Date
- Time
- Payment
- COGS
- Gross margin percentage
- Gross income

Atribut Customer type digunakan sebagai label kelas (target) dalam proses klasifikasi, dengan dua kategori utama yaitu member dan normal. Pemilihan dataset ini didasarkan pada kelengkapan atribut transaksi dan demografi yang relevan untuk analisis perilaku konsumen ritel, serta sifat dataset yang terbuka sehingga memungkinkan replikasi penelitian oleh peneliti lain.

Flowchart Metodologi Penelitian :



Gambar 1. Flowchart Metodologi Penelitian Klasifikasi Tipe Konsumen Retail Supermarket
(Sumber: Hasil penelitian, 2026)

Penjelasan Tahapan Metodologi Penelitian :

A. Akuisisi Data :

Tahap akuisisi data dilakukan dengan mengunduh dataset Supermarket Sales dari platform Kaggle. Dataset yang diperoleh digunakan sebagai data mentah untuk seluruh proses analisis dan pemodelan dalam penelitian ini.

B. Pra-pemrosesan Data :

Dataset memiliki struktur tabel dengan 1.000 baris data (record) dan 18 kolom atribut. Berdasarkan jenis datanya, atribut-atribut dalam dataset dikelompokkan sebagai berikut:

A. Atribut kategorikal:

Branch, City, Customer type, Gender, Product line, dan Payment

B. Atribut numerik:

Unit price, Quantity, Tax 5%, Sales, COGS, dan Gross income

C. Atribut temporal:

Date dan Time

Hasil pemeriksaan awal menunjukkan bahwa dataset tidak memiliki nilai kosong (missing value) dan tidak terdapat data duplikat. Atribut Invoice ID dihapus karena bersifat unik dan tidak berkontribusi terhadap proses klasifikasi. Variabel kategorikal dikonversi ke dalam bentuk numerik menggunakan teknik one-hot encoding agar dapat diproses oleh algoritma Decision Tree.

C. Rekayasa Fitur

Rekayasa fitur dilakukan dengan mengekstraksi informasi dari atribut Date dan Time. Dari atribut tersebut dihasilkan fitur baru berupa jam transaksi (hour), hari dalam minggu (day_of_week), dan bulan transaksi (month). Penambahan fitur ini bertujuan untuk menangkap pola perilaku konsumen berdasarkan waktu pembelian.

D. Pembagian Data

Dataset yang telah melalui tahapan pra-pemrosesan dan rekayasa fitur dibagi menjadi 80% data latih dan 20% data uji. Pembagian ini bertujuan untuk mengukur kemampuan generalisasi model dan menghindari terjadinya overfitting.

E. Pemodelan Decision Tree

Pemodelan dilakukan menggunakan algoritma Decision Tree dengan dua skenario, yaitu model dengan seluruh fitur dan model dengan fitur terpilih serta pembatasan kedalaman pohon. Pembatasan kedalaman pohon dilakukan untuk menghasilkan model yang lebih sederhana dan mudah diinterpretasikan.

F. Evaluasi Model

Evaluasi kinerja model dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score, serta confusion matrix untuk menganalisis kesalahan klasifikasi pada masing-masing kelas.

HASIL DAN PEMBAHASAN

Kondisi Awal Dataset dan Pra-pemrosesan Data :

Hasil analisis awal menunjukkan dataset penjualan supermarket yang digunakan dalam penelitian ini memiliki kualitas data yang baik, sebagaimana tercermin dari tidak adanya nilai kosong dan data duplikat. Hal ini sejalan dengan makalah yang mencatat bahwa dalam validasi data, identifikasi dan penghapusan nilai tidak valid adalah langkah kritis dalam memastikan efektivitas model analisis data, terutama dalam konteks dataset besar yang kompleks (Yasodha, 2025). Selain itu, temuan bahwa dataset ini terkomposisi dari 1.000 transaksi dengan atribut numerik, kategorikal, dan temporal mencerminkan keberagaman data yang dapat memberikan insight yang berharga mengenai perilaku konsumen (Sinaga et al, 2023).

algoritma Decision Tree. Variabel kategorikal dikonversi ke dalam format yang sesuai, sedangkan atribut yang tidak relevan atau bersifat redundan dihapus dari dataset. Selain itu, dilakukan rekayasa fitur dengan mengekstraksi informasi waktu dari atribut tanggal transaksi, yaitu jam transaksi, hari dalam minggu, dan bulan transaksi. Langkah ini bertujuan untuk menangkap pola perilaku konsumen berdasarkan waktu pembelian dan meningkatkan representasi data sebelum proses pemodelan.

Hasil Klasifikasi Menggunakan Decision Tree dengan Fitur Lengkap :

Model Decision Tree pertama dibangun menggunakan seluruh fitur hasil pra-pemrosesan dan rekayasa fitur, termasuk atribut transaksi, demografis, dan fitur temporal. Dataset dibagi menjadi data latih dan data uji dengan proporsi 80% untuk data latih dan 20% untuk data uji. Evaluasi kinerja model dilakukan pada data uji menggunakan metrik akurasi, presisi, recall, dan F1-score.

Hasil evaluasi menunjukkan bahwa model Decision Tree dengan fitur lengkap menghasilkan tingkat akurasi sebesar **56,5%**. Nilai presisi, recall, dan F1-score berada pada kisaran yang relatif seimbang. Hal ini menunjukkan bahwa model mampu mengenali pola dasar perbedaan antara pelanggan bertipe *member* dan *normal*. Namun demikian, tingkat akurasi tersebut juga mengindikasikan bahwa model masih mengalami kesulitan dalam membedakan beberapa kasus, terutama pada data dengan pola transaksi yang saling tumpang tindih.

Kinerja model yang belum optimal pada skenario ini diduga disebabkan oleh kompleksitas model akibat jumlah fitur yang cukup banyak serta adanya fitur kategorikal hasil one-hot encoding yang berpotensi menambah noise. Tanpa pembatasan kompleksitas, Decision Tree cenderung membentuk aturan yang terlalu spesifik terhadap data latih, sehingga kemampuan generalisasi model menjadi terbatas.

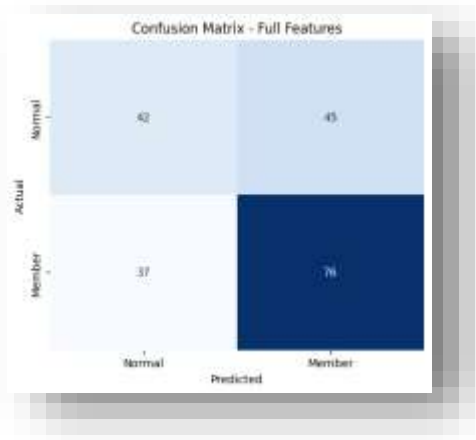
Hasil Klasifikasi dengan Pemilihan Fitur dan Pembatasan Kedalaman Pohon :

Pada skenario kedua, dilakukan pemilihan fitur utama yang dianggap paling relevan terhadap klasifikasi tipe konsumen, yaitu jenis kelamin, kategori produk, jumlah pembelian, dan total penjualan. Selain itu, diterapkan pembatasan kedalaman pohon keputusan untuk mengurangi kompleksitas model dan risiko overfitting.

Hasil evaluasi menunjukkan bahwa akurasi model meningkat menjadi **61,5%**, lebih tinggi dibandingkan skenario fitur lengkap. Peningkatan ini menunjukkan bahwa pengurangan kompleksitas model dan fokus pada fitur-fitur yang paling informatif dapat meningkatkan kemampuan generalisasi model terhadap data uji. Model pada skenario ini juga menghasilkan struktur pohon keputusan yang lebih sederhana dan mudah diinterpretasikan.

Analisis lebih lanjut terhadap laporan klasifikasi menunjukkan bahwa model memiliki performa yang lebih baik dalam mengidentifikasi pelanggan bertipe *member* dibandingkan pelanggan *normal*. Hal ini mengindikasikan bahwa pola transaksi pelanggan *member* cenderung lebih consistent dan lebih mudah dikenali oleh model berdasarkan atribut jumlah pembelian dan total penjualan.

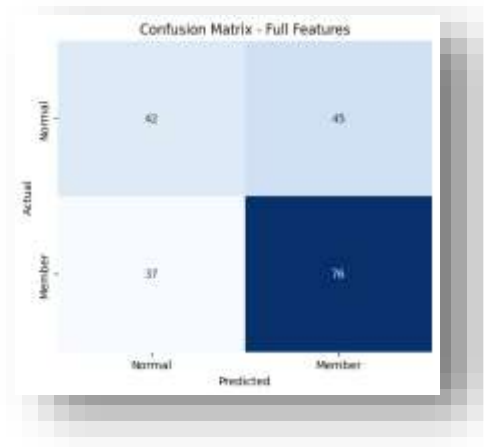
Hasil Klasifikasi Menggunakan Seluruh Fitur :



Gambar 2. Confusion Matrix Model Decision Tree Menggunakan Seluruh Fitur
(Sumber: Hasil pengolahan data, 2026)

Berdasarkan confusion matrix, model berhasil mengklasifikasikan 76 konsumen member dengan benar, namun 45 konsumen normal salah diklasifikasikan sebagai member. Hal ini menunjukkan bahwa model memiliki kecenderungan bias terhadap kelas member sehingga akurasi keseluruhan hanya mencapai 56,5%.

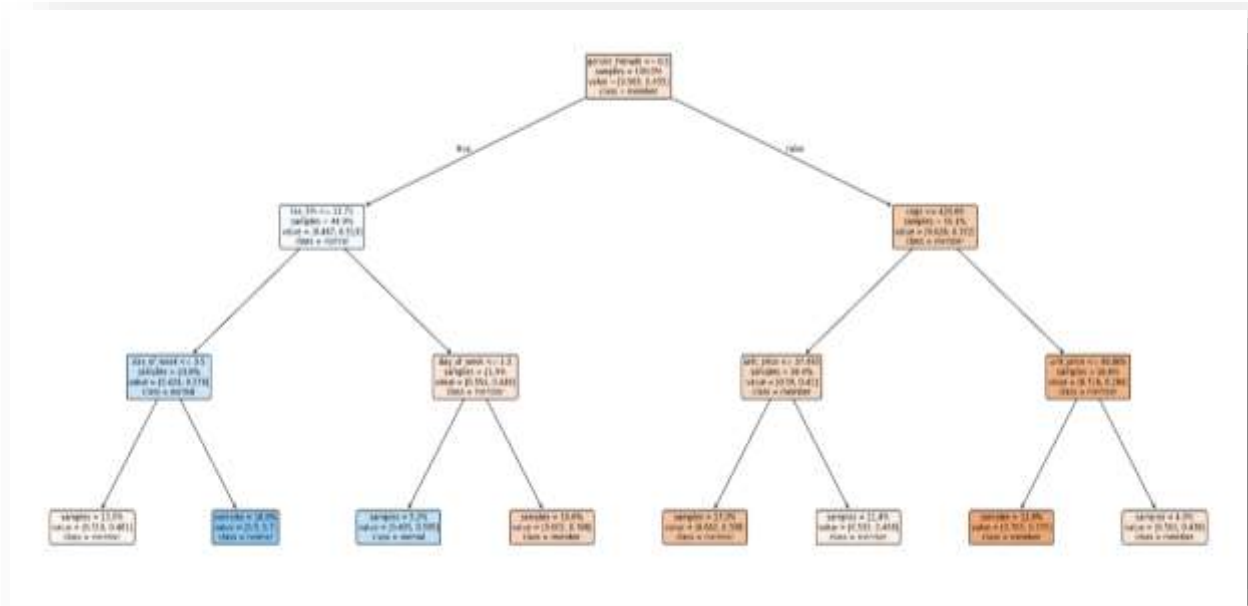
Hasil Klasifikasi Menggunakan Fitur Terpilih :



Gambar 3. Confusion Matrix Model Decision Tree Menggunakan Fitur Terpilih
(Sumber: Hasil pengolahan data, 2026)

- Model dengan fitur terpilih menjadi lebih sederhana dan mudah diinterpretasikan, karena hanya menggunakan fitur yang dianggap paling relevan.
- Confusion matrix menunjukkan bahwa prediksi benar dan salah masih cukup seimbang, sehingga performa model belum optimal.
- Masih terdapat kesalahan klasifikasi yang tinggi, baik:
 - data Normal yang diprediksi sebagai Member, maupun
 - data Member yang diprediksi sebagai Normal.
- Hal ini mengindikasikan bahwa fitur terpilih belum sepenuhnya mampu membedakan antara kelas Normal dan Member.
- Meskipun kompleksitas model berkurang, penyederhanaan fitur berdampak pada penurunan akurasi.
- Dengan demikian, terdapat trade-off antara interpretabilitas dan performa, di mana model lebih mudah dijelaskan tetapi masih perlu peningkatan ketepatan klasifikasi.

Visualisasi dan Interpretasi Pohon Keputusan :



Gambar 4. Visualisasi Pohon Keputusan Klasifikasi Tipe Konsumen Retail Supermarket
(Sumber: Hasil pengolahan data, 2026)

Pohon keputusan merupakan salah satu alat yang efektif dalam analisis data untuk mengidentifikasi hubungan antar variabel dan klasifikasi tipe konsumen. Berdasarkan analisis pohon keputusan yang dilakukan, fitur **gender_Female** berfungsi sebagai pemisah utama pada node akar. Hal ini menunjukkan bahwa jenis kelamin adalah atribut dominan dalam membedakan antara tipe konsumen member dan normal. Penelitian sebelumnya menunjukkan bahwa fitur demografis dapat berfungsi sebagai pembeda yang signifikan dalam berbagai konteks, termasuk dalam analisis perilaku (Walsh et al., 2018).

Cabang Pelanggan :

Analisis Pelanggan Laki-laki:

Pada cabang yang menggambarkan pelanggan laki-laki ($gender_{Female} \leq 0.5$), *model menggunakan atribut tax5%* sebagai indikator transaksi. Kemudian, atribut *dayofweek* digunakan untuk menangkap pola pembelian, yang relevan dengan kebiasaan konsumen berdasar waktu dalam seminggu. Beberapa node daun pada cabang ini memperlihatkan dominasi kelas normal, di mana proporsi pelanggan normal mencapai sekitar 70%. Hal ini sejalan dengan studi yang menyoroti pentingnya pengelompokan berdasarkan atribut transaksi untuk memaksimalkan strategi pemasaran dan meningkatkan kepuasan pelanggan (Muneiah & Rao, 2019).

Analisis Pelanggan Perempuan :

Di sisi lain, cabang pelanggan perempuan ($\text{genderFemale} > 0.5$) menunjukkan bahwa atribut *cogs* (*cost of goods sold*) dan *unitprice* menjadi faktor utama dalam klasifikasi. Mayoritas node daun di cabang ini diprediksi sebagai konsumen member. Temuan ini menunjukkan bahwa pelanggan wanita lebih cenderung menjadi member jika faktor-faktor transaksi tertentu terpenuhi, yang dapat dipahami dengan pendekatan analitis yang telah diterapkan dalam berbagai studi klasifikasi (Muneiah & Rao, 2019).

Hubungan Antara Atribut

Gabungan analisis menarik tersebut menekankan bahwa kombinasi faktor demografis (gender), indikator nilai transaksi (*tax5%*, *cogs*, *unitprice*), dan faktor temporal (*dayofweek*) semuanya berkontribusi dalam membedakan tipe konsumen. Penggunaan pohon keputusan memberikan wawasan yang mendalam tentang faktor-faktor yang mempengaruhi perilaku konsumen, di mana tiap cabang dari pohon keputusan menyediakan informasi yang terstruktur dan intuitif untuk pengambilan keputusan bisnis (Yu et al., 2010).

KESIMPULAN

Studi ini menunjukkan bahwa algoritma Pohon Keputusan dapat berhasil digunakan untuk mengklasifikasikan pelanggan supermarket ritel berdasarkan data transaksi mereka. Dataset untuk penelitian ini diperoleh dari Kaggle, di mana mereka memastikan data berkualitas tinggi. Hal ini dibuktikan dengan tidak adanya nilai yang hilang atau data duplikat selama tahap pra-pemrosesan, yang sejalan dengan praktik terbaik dalam penggunaan data untuk penelitian (Safavian & Landgrebe, 1991).

Hasilnya menunjukkan bahwa model Pohon Keputusan awal memiliki akurasi 56,5%. Setelah memilih fitur penting dan membatasi kedalaman pohon keputusan, akurasi meningkat menjadi 61,5%. Peningkatan akurasi ini menyoroti pentingnya menyederhanakan model dan berfokus pada fitur yang paling signifikan. Ini akan membantu model melakukan generalisasi lebih baik dan mengurangi risiko kelebihan pas (*overfitting*), tantangan utama dalam pembelajaran mesin dan analisis data (Ahmed et al., 2018). Pemilihan fitur secara langsung memengaruhi kemampuan algoritma untuk mengidentifikasi pola penting dalam data dan menunjukkan seberapa baik model selaras dengan tujuan penelitian (Jia & Cristianini, 2015).

Analisis visualisasi pohon keputusan mengungkapkan bahwa jenis kelamin merupakan faktor signifikan dalam menentukan jenis konsumen. Proses klasifikasi ini menunjukkan efektivitas Pohon Keputusan dalam mengurutkan berdasarkan jenis kelamin, yang mengindikasikan hubungan kuat yang memerlukan pemantauan terhadap bias dalam distribusi data. Dalam kasus ini, temuan ini didukung oleh penelitian yang menunjukkan bias gender dalam analisis data. Ini menyiratkan bahwa efek demografis harus dipertimbangkan saat mengembangkan model lebih lanjut (Chang et al., 2015; Dixon et al., 2018).

Alasan penting lainnya untuk menggunakan algoritma Pohon Keputusan sebagai alat pendukung keputusan dalam analisis ritel adalah karena mudah dipahami. Pemangku kepentingan dapat membuat keputusan berdasarkan hasil analisis yang jelas dan akuntabel ketika modelnya jelas dan mudah dipahami (Ahmed et al., 2018). Untuk penelitian di masa depan, disarankan untuk menggunakan kumpulan data yang berbeda dan berbagai metode untuk mengoptimalkan parameter. Disarankan juga untuk membandingkan hasil Pohon Keputusan dengan metode ensemble seperti Hutan Acak dan Peningkatan Gradien. Studi ini memberikan arah yang jelas untuk mengembangkan metode klasifikasi yang lebih baik dalam menganalisis data ritel dan konsumen (Wen et al., 2024).

DAFTAR PUSTAKA

- Ahmed, A., Rızaner, A., & Ulusoy, A. (2018). A novel decision tree classification based on post-pruning with Bayes minimum risk. *PLOS ONE*, 13(4), e0194168. <https://doi.org/10.1371/journal.pone.0194168>
- Almagribi, A. (2025). Clustering and classification of retail sales data: A big data and data mining analysis. *Journal of Innovations in Computer Science*, 4(2), 242–253. <https://doi.org/10.56347/jics.v4i2.303>
- Cant, M., & Toit, M. (2012). Identifying the factors that influence retail customer loyalty and capitalising them. *International Business & Economics Research Journal (IBER)*, 11(11), 1223. <https://doi.org/10.19030/iber.v11i11.7370>
- Chang, J., Travaglione, A., & O'Neill, G. (2015). How can gender signal employee qualities in retailing? *Journal of Retailing and Consumer Services*, 27, 24–30. <https://doi.org/10.1016/j.jretconser.2015.07.004>
- Chen, J. (2024). Advanced analytics for retail inventory and demand forecasting. *TEBMR*, 10, 113–119. <https://doi.org/10.62051/jme9b319>
- Dixon, L., Li, J., Sorensen, J., Thain, N., & Vasserman, L. (2018). Measuring and mitigating unintended bias in text classification. *Proceedings of the ACM Conference*, 67–73. <https://doi.org/10.1145/3278721.3278729>
- Gunawan, I., & Setiawan, T. (2023). Analisis regresi linier dalam memprediksi data penjualan supermarket. *Jurnal Saintikom (Jurnal Sains Manajemen Informatika dan Komputer)*, 22(1), 198. <https://doi.org/10.53513/jis.v22i1.7556>
- He, W., & Zeng, Q. (2021). Research on sales forecast based on XGBoost-LSTM algorithm model. *Journal of Physics: Conference Series*, 1754(1), 012191. <https://doi.org/10.1088/1742-6596/1754/1/012191>
- Jia, S., & Cristianini, N. (2015). Learning to classify gender from four million images. *Pattern Recognition Letters*, 58, 35–41. <https://doi.org/10.1016/j.patrec.2015.02.006>
- Liu, H. (2024). Comparative analysis of machine learning algorithms for sales forecasting in the Russian toy retail sector. *Advances in Economics, Management and Political Sciences*, 128(1), 180–187. <https://doi.org/10.54254/2754-1169/2024.18672>
- Mansur, S., Sattar, K., Hosseini, S., Pervez, S., Ahmad, I., Saleem, K., & Elhendi, A. (2025). Sales forecasting for retail stores using hybrid neural networks and sales-affecting variables. *PeerJ Computer Science*, 11, e3058. <https://doi.org/10.7717/peerj-cs.3058>
- Mukhlisin, M., & Nugroho, H. (2025). Customer loyalty classification using KNN and decision tree for sales strategy development. *Sinkron*, 9(3), 1159–1166. <https://doi.org/10.33395/sinkron.v9i3.15110>
- Mühlbacher, T., Linhardt, L., Möller, T., & Piringer, H. (2018). TreePOD: Sensitivity-aware selection of Pareto-optimal decision trees. *IEEE Transactions on Visualization and Computer Graphics*, 24(1), 174–183. <https://doi.org/10.1109/TVCG.2017.2745158>
- Phillips, N., Neth, H., Woike, J., & Gaissmaier, W. (2017). FFTrees: A toolbox to create, visualize, and evaluate fast-and-frugal decision trees. *Judgment and Decision Making*, 12(4), 344–368. <https://doi.org/10.1017/S1930297500006239>
- Safavian, S. R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(3), 660–674. <https://doi.org/10.1109/21.97458>
- Sousa, A., Moro, S., & Pereira, R. (2023). Cluster-based approaches toward developing a customer loyalty

- program in a private security company. *Applied Sciences*, 14(1), 78. <https://doi.org/10.3390/app14010078>
- Vera-Salmerón, E., Domínguez-Nogueira, C., Sáez, J., Romero-Béjar, J., & Mota-Romero, E. (2024). Differentiating pressure ulcer risk levels through interpretable classification models based on readily measurable indicators. *Healthcare*, 12(9), 913. <https://doi.org/10.3390/healthcare12090913>
- Wen, K., Joseph, M., & Sivakumar, V. (2024). Big Mart sales prediction using machine learning. *EAI Endorsed Transactions on Internet of Things*, 10. <https://doi.org/10.4108/eetiot.6453>
- Muneiah, J., & Rao, C. (2019). Customer's class transformation for profit maximization in multi-class setting of telecom industry using probability estimation decision trees. *Journal of Intelligent & Fuzzy Systems*, 37(6), 8167–8197. <https://doi.org/10.3233/JIFS-190628>
- Walsh, J., Heazlewood, I., & Climstein, M. (2018). Application of gradient boosted trees to gender prediction based on motivations of masters athletes. *Model Assisted Statistics and Applications*, 13(3), 235–252. <https://doi.org/10.3233/MAS-180438>
- Yu, Z., Haghighat, F., Fung, B., & Yoshino, H. (2010). A decision tree method for building energy demand modeling. *Energy and Buildings*, 42(10), 1637–1646. <https://doi.org/10.1016/j.enbuild.2010.04.006>