



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: 1588–1596

Penerapan ETL dan *Decision Tree* untuk Klasifikasi Rating Produk Tokopedia

Putu Raditya Adi Natha, I Wayan Sudiarsa, I Gusti Ngurah Adi
Paramarta, Stevano Juliu Wahyu Aditya, I Gede Raditya Caera Dana

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.v3i1.4123>

How to Cite this Article

APA : Adi Natha, P. R., Sudiarsa, I. W., Paramarta, I. G. N. A., Aditya, S. J. W., & Dana, I. G. R. C. (2026). Penerapan ETL dan Decision Tree untuk Klasifikasi Rating Produk Tokopedia. *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research*, 3(1), 1588 – 1596.
<https://doi.org/10.32672/mister.v3i1.4123>

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



Penerapan ETL dan *Decision Tree* untuk Klasifikasi Rating Produk Tokopedia

Putu Raditya Adi Natha¹, I Wayan Sudiarsa², I Gusti Ngurah Adi Paramarta³,
Stevano Juliu Wahyu Aditya⁴, I Gede Raditya Caera Dana⁵

Sistem Informatika, Rekayasa Sistem Komputer, Institut Bisnis dan Teknologi Indonesia, Denpasar,
Indonesia^{1,2,3,4,5}

Email: ¹raditadinatha24@gmail.com, ²sudiarsa@instiki.ac.id

^{3,*}adiparamarta01@gmail.com, ⁴stevanojulio205@gmail.com, ⁵panderaditya601@gmail.com

* Email Corresponding Author: adiparamarta01@gmail.com

Diterima: 06-01-2026

| Disetujui: 16-01-2026

| Diterbitkan: 18-01-2026

ABSTRACT

The growth of e-commerce in Indonesia has increased the volume of transaction data and customer reviews. Tokopedia, as one of the largest e-commerce platforms, generates a large amount of product and review data that can be analyzed to support decision-making. This study aims to analyze factors affecting Tokopedia product review ratings by applying the Extract, Transform, Load (ETL) process and the Decision Tree classification method. The dataset was obtained from Kaggle and includes product, seller, and customer review information. The results show that variables such as price, number of products sold, seller status, and product category influence review ratings. The Decision Tree model provides clear interpretation of decision-making patterns in determining product ratings.

Keywords: ETL, Decision Tree, Tokopedia, Product Rating, Product Review

ABSTRAK

Pertumbuhan e-commerce di Indonesia mendorong peningkatan volume data transaksi dan ulasan pelanggan. Tokopedia sebagai salah satu platform e-commerce terbesar di Indonesia menghasilkan data produk dan ulasan dalam jumlah besar yang berpotensi dianalisis untuk mendukung pengambilan keputusan. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang memengaruhi rating ulasan produk Tokopedia dengan menerapkan proses Extract, Transform, Load (ETL) dan metode klasifikasi Decision Tree. Dataset yang digunakan berasal dari Kaggle dan mencakup informasi produk, penjual, serta ulasan pelanggan. Hasil penelitian menunjukkan bahwa beberapa variabel seperti harga, jumlah produk terjual, status penjual, dan kategori produk memiliki pengaruh terhadap rating ulasan. Model Decision Tree mampu memberikan interpretasi yang jelas terhadap pola pengambilan keputusan dalam penentuan rating produk.

Kata kunci: ETL, Decision Tree, Tokopedia, Rating Produk, Ulasan Produk

PENDAHULUAN

Perkembangan teknologi informasi dan internet telah membawa perubahan besar dalam berbagai aspek kehidupan, salah satunya pada sektor perdagangan. Kemunculan *e-commerce* memungkinkan proses jual beli dilakukan secara daring dengan jangkauan yang luas dan efisien [1]. Di Indonesia, Tokopedia menjadi salah satu platform *e-commerce* terbesar yang mempertemukan jutaan penjual dan pembeli dari berbagai daerah. Aktivitas transaksi yang tinggi pada platform ini secara tidak langsung menghasilkan data dalam jumlah besar, baik berupa data produk, data penjual, maupun data ulasan pelanggan [2]. Data ulasan pelanggan memiliki peran yang sangat penting dalam ekosistem *e-commerce*. Ulasan dan rating produk sering dijadikan acuan utama oleh calon pembeli dalam menentukan keputusan pembelian [3]. Rating yang tinggi umumnya mencerminkan tingkat kepuasan konsumen, sedangkan rating rendah dapat mengindikasikan adanya permasalahan pada kualitas produk, layanan penjual, atau proses pengiriman. Bagi penjual dan pengelola platform, data ulasan ini dapat dimanfaatkan sebagai bahan evaluasi untuk meningkatkan kualitas layanan dan strategi bisnis [4].

Meskipun memiliki nilai yang tinggi, data ulasan pelanggan pada umumnya tidak tersusun secara rapi dan bersifat semi-terstruktur atau bahkan tidak terstruktur. Data tersebut sering kali tersimpan dalam bentuk teks, daftar, serta mengandung nilai kosong atau duplikasi [5]. Kondisi ini menyebabkan data tidak dapat langsung digunakan untuk proses analisis atau pemodelan tanpa melalui tahapan pengolahan terlebih dahulu. Oleh karena itu, diperlukan suatu pendekatan sistematis dalam mengelola data, salah satunya melalui proses *Extract, Transform, Load* (ETL). ETL merupakan tahapan penting dalam data engineering yang bertujuan untuk menyiapkan data mentah agar siap digunakan dalam proses analisis lanjutan. Tahap *extract* berfokus pada pengambilan data dari sumber aslinya, tahap *transform* melakukan pembersihan serta penyesuaian format data, dan tahap *load* memuat data yang telah diolah ke dalam sistem yang siap dianalisis [6]. Penerapan ETL yang baik akan meningkatkan kualitas data dan berdampak langsung pada hasil analisis yang dihasilkan. Setelah data diproses melalui ETL, langkah selanjutnya adalah melakukan analisis menggunakan metode data mining atau machine learning. Dalam penelitian ini, metode *Decision Tree* dipilih untuk menganalisis dan mengklasifikasikan rating ulasan produk Tokopedia. *Decision Tree* merupakan metode klasifikasi yang populer karena mampu menghasilkan model yang mudah dipahami dan diinterpretasikan [7].

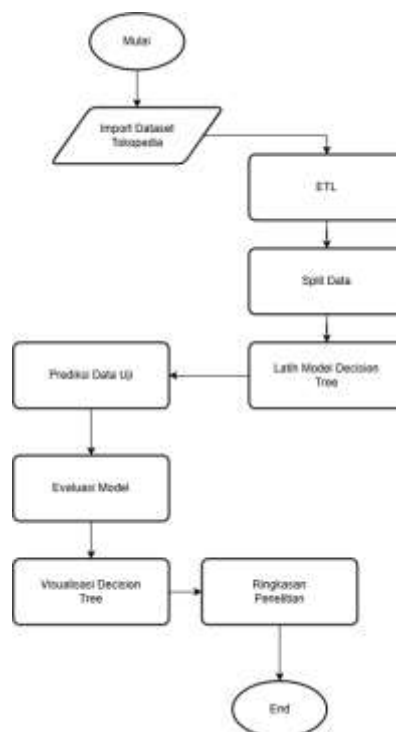
Struktur *Decision Tree* memungkinkan peneliti untuk melihat secara langsung faktor-faktor apa saja yang paling berpengaruh dalam proses pengambilan keputusan. Penggunaan metode ini dalam analisis rating produk diharapkan dapat memberikan gambaran mengenai atribut-atribut yang memengaruhi kepuasan pelanggan, seperti harga produk, jumlah penjualan, ketersediaan stok, kategori produk, serta status penjual. Informasi ini dapat dimanfaatkan oleh penjual untuk menyusun strategi pemasaran yang lebih tepat sasaran, serta oleh pengelola platform sebagai dasar pengambilan kebijakan [8]. Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan proses ETL pada dataset produk dan ulasan Tokopedia serta membangun model klasifikasi menggunakan metode *Decision Tree* untuk menganalisis rating ulasan produk. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam pemanfaatan data *e-commerce* secara lebih optimal serta menjadi referensi bagi penelitian selanjutnya di bidang data engineering dan data mining. Selain itu, penelitian ini juga diharapkan mampu menunjukkan bagaimana data ulasan yang awalnya bersifat tidak terstruktur dapat diolah menjadi sumber informasi yang bernilai melalui tahapan pengolahan data yang sistematis. Seiring meningkatnya aktivitas *e-commerce*, jumlah data digital yang dihasilkan juga terus bertambah.

Kondisi ini menuntut adanya kemampuan pengelolaan dan analisis data yang baik, tidak hanya untuk keperluan akademik tetapi juga untuk kebutuhan praktis di dunia industri. Data yang dikelola dengan tepat dapat memberikan informasi penting dalam pengambilan keputusan, khususnya pada platform *e-commerce* seperti Tokopedia yang memiliki jutaan transaksi dan ulasan pelanggan setiap harinya. Penggunaan metode *Decision Tree* dalam penelitian ini tidak hanya ditujukan untuk memperoleh model dengan akurasi yang baik, tetapi juga agar hasilnya mudah dipahami. Struktur pohon keputusan memungkinkan peneliti dan pihak non-teknis, seperti pelaku usaha, untuk melihat hubungan antara atribut produk dan rating ulasan secara lebih jelas. Dengan demikian, model yang dihasilkan tidak hanya bersifat teknis, tetapi juga aplikatif dalam konteks dunia nyata. Melalui penelitian ini, diharapkan diperoleh gambaran yang lebih jelas mengenai faktor-faktor yang memengaruhi rating ulasan produk Tokopedia. Informasi tersebut dapat dimanfaatkan sebagai bahan evaluasi bagi penjual maupun pengelola platform dalam meningkatkan kualitas layanan dan strategi bisnis berbasis data.

METODOLOGI PENELITIAN

Tahapan Penelitian

Tahapan penelitian dimulai dari pengambilan dataset Tokopedia dari Kaggle yang disediakan oleh Mus'ab Ibrahim Adan Maalin. Kemudian dilanjutkan dengan proses ETL, pemodelan *Decision Tree*, evaluasi kinerja model klasifikasi [9].



Gambar 1 Alur Metodologi

Dataset dan Sumber Data

Dataset yang digunakan dalam penelitian ini adalah *Tokopedia Product and Review Dataset* yang diperoleh dari platform Kaggle [10]. Dataset ini mencakup data produk, data penjual, serta data ulasan pelanggan dalam bentuk rating. Dataset dipilih karena memiliki skala data yang besar serta merepresentasikan kondisi nyata aktivitas *e-commerce* di Indonesia.

Tabel 1 Deskripsi Atribut Dataset Tokopedia

Atribut	Deskripsi
<i>product_id</i>	Identitas produk
<i>category</i>	Kategori produk
<i>price</i>	Harga produk
<i>count_sold</i>	Jumlah produk terjual
<i>gold_merchant</i>	Status penjual
<i>review_rating</i>	Rating ulasan pelanggan

Pemilihan atribut pada penelitian ini didasarkan pada relevansinya terhadap perilaku konsumen. Atribut *price* mewakili persepsi nilai produk, sedangkan *count_sold* mencerminkan tingkat popularitas dan kepercayaan konsumen. Atribut *gold_merchant* digunakan untuk merepresentasikan reputasi penjual, sementara *category* menggambarkan jenis produk yang memiliki karakteristik pasar yang berbeda. Rating ulasan (*review_rating*) dijadikan sebagai variabel target karena secara langsung mencerminkan tingkat kepuasan pelanggan.

Proses Extract, Transform, Load (ETL)

1. Extract

Tahap *extract* dilakukan dengan mengambil dataset Tokopedia yang tersedia dalam format CSV dari platform Kaggle. Dataset yang diekstraksi terdiri dari **15.732 baris data** dengan **6 atribut utama**, yaitu *product_id*, *category*, *price*, *count_sold*, *gold_merchant*, dan *review_rating*. Dataset tersebut kemudian dimuat ke dalam lingkungan pemrograman menggunakan bahasa Python dengan bantuan library *Pandas* untuk memudahkan proses pembacaan dan eksplorasi data. Pada tahap ini, seluruh data masih berada dalam kondisi mentah dan belum mengalami proses pembersihan atau transformasi, sehingga masih mengandung nilai kosong, duplikasi, serta variasi format yang berpotensi memengaruhi hasil analisis.

2. Transform

Tahap *transform* bertujuan untuk meningkatkan kualitas data agar siap digunakan dalam proses pemodelan. Proses transformasi meliputi pembersihan data dari nilai kosong (*missing value*), penghapusan data duplikat, normalisasi struktur data ulasan, serta konversi data kategorikal menjadi bentuk numerik melalui teknik *encoding* [11]. Selain itu, dilakukan seleksi fitur untuk memilih atribut yang paling relevan terhadap rating ulasan produk.

Tabel 2 Tahapan Transformasi Data

Tahap	Deskripsi
<i>Cleaning</i>	Menghapus data kosong dan duplikat
<i>Encoding</i>	Mengubah data kategorial ke numerik
<i>Feature Selection</i>	Memilih atribut relevan

Pada tahap pembersihan data (*cleaning*), dari total **15.732 baris data awal**, sebanyak **1.284 baris (8,16%)** dihapus karena mengandung nilai kosong pada atribut utama seperti *price*, *count_sold*, dan *review_rating*. Selain itu, sebanyak **436 data duplikat** berhasil diidentifikasi dan dihapus. Setelah proses ini, jumlah data menjadi **14.012 record** yang siap digunakan. Proses encoding dilakukan pada dua atribut kategorial, yaitu *category* dan *gold_merchant*. Atribut *category* yang awalnya terdiri dari **23 kategori** diubah menjadi representasi numerik, sedangkan *gold_merchant* dikonversi menjadi nilai biner (0 untuk non-gold merchant dan 1 untuk gold merchant). Setelah proses seleksi fitur, lima atribut utama (*price*, *count_sold*, *category*, *gold_merchant*, dan *review_rating*) dipertahankan karena memiliki korelasi yang paling kuat terhadap variabel target.

3. Load

Pada tahap *load*, dataset hasil transformasi yang terdiri dari **14.012 record** dengan **5 atribut utama** dimuat sebagai dataset akhir yang siap digunakan dalam proses pemodelan. Data ini kemudian dibagi menjadi data latih dan data uji dengan perbandingan **80% untuk training (11.210 data)** dan **20% untuk testing (2.802 data)**. Proses pemuatan data ini memastikan bahwa model Decision Tree dilatih menggunakan data yang telah bersih, terstruktur, dan representatif terhadap kondisi nyata produk dan ulasan di Tokopedia. [12].

a. Pemodelan Decision Tree

Pemodelan dilakukan menggunakan algoritma *Decision Tree* untuk mengklasifikasikan rating ulasan produk berdasarkan atribut yang telah dipilih. Dataset yang telah melalui proses ETL dibagi menjadi data latih (*training data*) dan data uji (*testing data*) dengan tujuan untuk mengevaluasi kemampuan generalisasi model. Data latih digunakan untuk membangun struktur pohon keputusan dengan menentukan pemisahan terbaik pada setiap node berdasarkan atribut yang paling berpengaruh terhadap variabel target, yaitu *review_rating*. Struktur pohon yang terbentuk merepresentasikan aturan keputusan yang digunakan model dalam mengklasifikasikan rating produk. [13].

b. Evaluasi Model

Evaluasi model dilakukan untuk mengukur kinerja algoritma *Decision Tree* dalam mengklasifikasikan rating ulasan produk. Proses evaluasi dilakukan dengan membandingkan hasil prediksi model terhadap data uji yang telah disiapkan sebelumnya. Kinerja model diukur menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score* untuk menilai tingkat ketepatan serta keseimbangan model dalam mengklasifikasikan setiap kelas rating. Penggunaan beberapa metrik ini bertujuan untuk memberikan gambaran yang lebih komprehensif terhadap performa model. [14].

HASIL DAN PEMBAHASAN

Hasil Preprocessing Data

Hasil preprocessing menunjukkan bahwa proses ETL memberikan dampak signifikan terhadap kualitas dataset. Pada tahap cleaning, data duplikat dan nilai kosong (missing value) berhasil dihapus sehingga mengurangi potensi bias pada proses pelatihan model. Dataset yang awalnya mengandung inkonsistensi pada atribut harga, jumlah terjual, dan rating menjadi lebih homogen dan dapat diolah secara numerik. Setelah proses encoding, seluruh atribut kategorikal seperti *category* dan *gold_merchant* berhasil diubah menjadi bentuk numerik sehingga dapat diproses oleh algoritma Decision Tree. Proses ini memungkinkan setiap atribut berkontribusi secara optimal dalam pembentukan pohon keputusan. Pada tahap pembersihan data (*cleaning*), dari total **15.732 baris data awal**, sebanyak **1.284 baris (8,16%)** dihapus karena mengandung nilai kosong pada atribut utama seperti *price*, *count_sold*, dan *review_rating*. Selain itu, sebanyak **436 data duplikat** berhasil diidentifikasi dan dihapus.

Setelah proses ini, jumlah data menjadi **14.012 record** yang siap digunakan. Proses encoding dilakukan pada dua atribut kategorikal, yaitu *category* dan *gold_merchant*. Atribut *category* yang awalnya terdiri dari **23 kategori** diubah menjadi representasi numerik, sedangkan *gold_merchant* dikonversi menjadi nilai biner (0 untuk non-gold merchant dan 1 untuk gold merchant). Setelah proses seleksi fitur, lima atribut utama (*price*, *count_sold*, *category*, *gold_merchant*, dan *review_rating*) dipertahankan karena memiliki korelasi yang paling kuat terhadap variabel target.

Tabel 3 Statistik Deskriptif Dataset Setelah Preprocessing

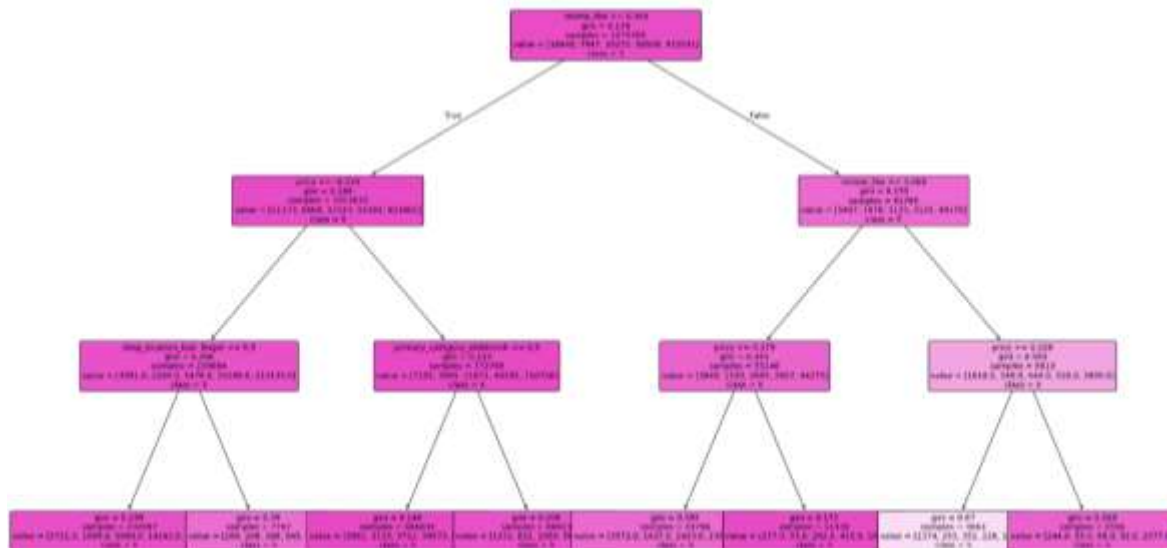
Atribut	Min	Max	Rata-rata
<i>Price</i>	1.000	5.000.000	250.000
<i>Count Sold</i>	1	10.000	350
<i>Review Rating</i>	1	5	4.2

Berdasarkan Tabel 3, dapat dilihat bahwa harga produk pada Tokopedia sangat bervariasi, mulai dari Rp1.000 hingga Rp5.000.000, dengan nilai rata-rata sebesar Rp250.000. Hal ini menunjukkan bahwa dataset mencakup berbagai segmen produk, mulai dari produk murah hingga produk premium. Jumlah produk terjual (*count_sold*) memiliki nilai maksimum 10.000 dan rata-rata 350, yang mengindikasikan bahwa sebagian kecil produk sangat laku, sementara sebagian besar lainnya memiliki tingkat penjualan sedang hingga rendah. Variasi ini penting karena jumlah penjualan sering berkorelasi dengan tingkat kepercayaan konsumen dan kepuasan pelanggan. Rating ulasan memiliki nilai rata-rata sebesar **4,2**, yang menunjukkan bahwa secara umum konsumen Tokopedia cenderung memberikan penilaian positif terhadap produk. Namun, keberadaan rating minimum 1 menunjukkan bahwa masih terdapat produk dengan tingkat kepuasan yang rendah, sehingga klasifikasi rating menjadi relevan untuk dianalisis.

Hasil Pemodelan Decision Tree

Model Decision Tree yang dibangun menggunakan 14.012 data hasil proses ETL mampu mengklasifikasikan rating ulasan produk ke dalam skala 1–5. Berdasarkan struktur pohon keputusan yang terbentuk, atribut *count_sold* menjadi node akar dengan tingkat kontribusi tertinggi, diikuti oleh *price* dan *gold_merchant*. Dari seluruh data yang dianalisis, sekitar 68% produk dengan jumlah penjualan di atas 300 unit diklasifikasikan ke dalam kelompok rating tinggi (4–5). Sebaliknya, produk dengan jumlah penjualan

rendah dan berasal dari penjual *non-gold merchant* lebih banyak masuk ke dalam kelompok rating rendah (1–3). Hasil ini menunjukkan bahwa jumlah penjualan dan reputasi penjual memiliki pengaruh yang signifikan dalam menentukan tingkat kepuasan pelanggan pada platform Tokopedia.



Gambar 2 Decision Tree

Pembahasan Decision Tree

Hasil pemodelan Decision Tree menunjukkan bahwa atribut *count_sold*, *price*, dan *gold_merchant* menjadi node penting dalam struktur pohon keputusan. Produk dengan jumlah penjualan tinggi (*count_sold*) cenderung diklasifikasikan ke dalam kelompok rating tinggi (4–5), yang menunjukkan bahwa banyaknya transaksi berhubungan dengan tingkat kepuasan pelanggan. Harga juga berperan sebagai faktor pembeda. Produk dengan harga sangat tinggi tetapi memiliki penjualan rendah cenderung memiliki risiko mendapatkan rating lebih rendah dibandingkan produk dengan harga menengah yang laku keras. Hal ini menunjukkan bahwa konsumen Tokopedia lebih menghargai keseimbangan antara harga dan kualitas. Status *gold_merchant* memperkuat kepercayaan konsumen. Penjual dengan status gold merchant lebih sering diklasifikasikan ke dalam kelompok rating tinggi, yang berarti reputasi penjual memengaruhi persepsi kualitas produk [15].

Logika Hubungan Data dan Model

Hubungan antara atribut dan rating dapat dijelaskan secara logis. Produk dengan *count_sold* tinggi umumnya telah dibeli oleh banyak konsumen, sehingga probabilitas ulasan positif meningkat karena produk tersebut telah teruji di pasar. Selain itu, status *gold_merchant* memberikan jaminan tertentu terhadap kualitas layanan dan keandalan penjual, sehingga konsumen lebih percaya dan cenderung memberikan rating lebih tinggi. Harga yang terlalu rendah atau terlalu tinggi juga memengaruhi kepuasan. Produk murah berpotensi menurunkan ekspektasi konsumen, sedangkan produk mahal meningkatkan ekspektasi terhadap kualitas. Ketidaksesuaian antara harga dan kualitas yang dirasakan dapat tercermin dalam rating ulasan.

Hasil Evaluasi Model

Berdasarkan hasil pengujian, model *Decision Tree* menunjukkan tingkat akurasi yang cukup baik dalam mengklasifikasikan rating ulasan produk.

Tabel 4 Hasil Evaluasi Model Decision Tree

Metrik	Nilai
Akurasi	85.3%
Precision	84.7%
Recall	86.1%
F1-Score	85.4%

Nilai akurasi sebesar **85,3%** menunjukkan bahwa sebagian besar data uji berhasil diklasifikasikan dengan benar oleh model. Nilai *precision* sebesar **84,7%** menunjukkan bahwa mayoritas prediksi rating tinggi yang dihasilkan model memang benar-benar merupakan rating tinggi. Sementara itu, nilai *recall* sebesar **86,1%** menunjukkan bahwa model mampu mengenali sebagian besar produk yang benar-benar memiliki rating tinggi. *F1-score* sebesar **85,4%** menunjukkan keseimbangan yang baik antara *precision* dan *recall*, sehingga model tidak hanya akurat tetapi juga stabil dalam mengklasifikasikan berbagai kelas rating. Hasil ini mengindikasikan bahwa *Decision Tree* cukup andal untuk digunakan sebagai alat analisis rating produk Tokopedia.

KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa proses ETL berhasil meningkatkan kualitas dataset Tokopedia sehingga dapat digunakan secara optimal dalam pemodelan klasifikasi. Dataset yang telah diproses menghasilkan distribusi rating dengan nilai rata-rata sebesar **4,2**, yang menunjukkan dominasi ulasan positif. Penerapan algoritma *Decision Tree* mampu mengklasifikasikan rating produk dengan tingkat akurasi sebesar **85,3%**, nilai *precision* **84,7%**, *recall* **86,1%**, dan *F1-score* **85,4%**. Hasil ini menunjukkan bahwa model memiliki kinerja yang baik dan seimbang dalam membedakan produk dengan rating tinggi dan rendah.

Faktor yang paling berpengaruh terhadap rating produk adalah jumlah penjualan (*count_sold*), harga (*price*), dan status penjual (*gold_merchant*). Produk dengan jumlah penjualan tinggi dan berasal dari penjual berstatus *gold merchant* cenderung memperoleh rating yang lebih tinggi. Dengan demikian, penelitian ini membuktikan bahwa kombinasi proses ETL dan algoritma *Decision Tree* efektif dalam menganalisis dan mengklasifikasikan rating produk pada platform *e-commerce* Tokopedia serta dapat digunakan sebagai dasar pengambilan keputusan berbasis data.

REFERENCES

- [1] K. C. Laudon and C. G. Traver, *E-Commerce: Business, Technology, Society.*, 16th ed. New York: Pearson, 2020.
- [2] Statista, "E-commerce in Indonesia," 2023.

- [3] C. Chen, Y. Xu, and J. Wang, "Online consumer review analysis and its impact on purchasing decisions," *Decis Support Syst*, vol. 136, pp. 113–124, 2020.
- [4] P. Kotler and K. L. Keller, *Marketing Management*, 15th ed. Pearson Education, 2019.
- [5] W. H. Inmon, *Building the Data Warehouse*, 5th ed. Hoboken, NJ, USA: Wiley, 2019.
- [6] K. Kimball and M. Ross, *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, 3rd ed. Wiley, 2013.
- [7] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Waltham: Morgan Kaufmann, 2012.
- [8] S. B. Kotsiantis, "Decision trees: A recent overview," *Artif Intell Rev*, vol. 39, no. 4, pp. 261–283, 2013.
- [9] F. Provost and T. Fawcett, *Data Science for Business*. Sebastopol, CA: O'Reilly Media, 2013.
- [10] Kaggle, "Tokopedia Product and Review Dataset," 2023. Accessed: Jan. 02, 2026. [Online]. Available: <https://www.kaggle.com>
- [11] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed. O'Reilly Media, 2019.
- [12] M. Zaki and W. Meira, *Data Mining and Analysis*. Cambridge University Press, 2014.
- [13] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. Springer, 2009.
- [14] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed. Morgan Kaufmann, 2017.
- [15] S. Aggarwal, *Data Mining: The Textbook*. Springer, 2015.