Available online at JDR Website: http://journal.unublitar.ac.id/jdr
Journal of Development Research, 10 (1), May 2026, Pages 1-6
DOI: https://doi.org/10.28926/jdr.v10i1.516

Sentiment Analysis of Tokopedia Customer Reviews Using Decision Tree

Dimas Ari Setyawan[1,*], Prasetyo Yekti Utomo[2], Ridho Muarief[3], Deni Nur Fauzi[4], Nur Hanifa Muslima[5]

Politeknik Negeri Madiun, Indonesia

E-mail: dimas.ari@pnm.ac.id[1,*], prasetyo@pnm.ac.id[2], ridho.muarief@pnm.ac.id[3], deninurfauzi@pnm.ac.id[4]

*Corresponding Author

Received: 23 April 2026 | Revised: 11 May 2026 | Accepted: 25 May 2026

Abstract

This research aims to analyze the sentiment of clothing stores on the Tokopedia marketplace. The dataset comprises the top 13 clothing stores and star ratings ranging from 1 to 5, spanning the period from 2019 to 2024. The Decision Tree algorithm was implemented for the sentiment analysis process, employing three attribute splitting methods: Gini Index, Gain Ratio, and Information Gain. A total of 10,490 reviews were utilized and categorized into three sentiment classes: 4,018 reviews with 4 and 5-star ratings were classified as the positive sentiment class; 3,378 reviews with 3-star ratings as the neutral sentiment class; and 3,094 reviews with 1 and 2-star ratings as the negative sentiment class. The results of the Decision Tree modeling demonstrate that the highest performance was achieved using the Information Gain algorithm, yielding an accuracy rate of 44.23%.

Keywords: *decision tree; e-commerce; information gain; sentiment analysis; tokopedia*

Introduction

The development of e-commerce in Indonesia has shown rapid growth in recent years, with Tokopedia emerging as one of the leading marketplaces facilitating transactions between sellers and buyers (Lestari et al., 2025; Utami, 2020). Tokopedia not only dominates the national e-commerce market share with daily transaction volumes reaching millions, but its data also shows that the fashion category, specifically clothing, is one of the segments with the highest sales (Ulhaq & Suprayogi, 2025). This phenomenon highlights the intense competition among online fashion businesses, where customer satisfaction serves as the primary determining factor in maintaining consumer loyalty and building a store's reputation (Warsito & Prahutama, 2020).

In this context, sentiment analysis of buyer reviews emerges as a vital tool for comprehensively measuring customer satisfaction levels (Ilmania & Cahyawijaya, 2018). Reviews left by buyers for clothing stores on Tokopedia often encompass various aspects of the shopping experience, ranging from evaluations of fabric quality, size accuracy, and color consistency relative to the displayed images, to assessments of shipping speed and seller service quality (Warsito & Prahutama, 2020). The unique characteristics of fashion product reviews for instance, "The denim is thick and comfortable, but the back pocket is too small" or "The color is lighter than the photo, but the material is very soft" demonstrate a complexity of sentiment that cannot be measured by star ratings alone. Instead, it requires an in-depth analysis of the review text content (Sachin et al., 2020; Xuanzhen & Xiaohong, 2019).

This research specifically focuses on review data from the 13 most popular clothing stores on Tokopedia, each having a minimum of 1,000 reviews. The data collection period spans from 2019 to 2024 to ensure the relevance of the findings to current market conditions. The analyzed dataset consists of approximately 10,490 reviews, including full review text, star ratings (1-5), and supporting information such as product types and review dates, with a relatively balanced distribution among positive, neutral, and negative sentiments. The selection of the Decision Tree as the analytical method is based on its superior ability to classify specific patterns within text data that has been transformed into structured features, such as the frequency of specific keywords relevant to fashion product evaluations (Setyawan & Fathicah, 2020).

The application of the Decision Tree in this study is expected not only to identify the dominant factors influencing customer sentiment but also to generate intuitive visualizations of consumer evaluation criteria for fashion products (Idris et al., 2025; Kurnianingrum et al., 2025). The findings of this research are projected to provide tangible benefits to three primary stakeholders: online fashion entrepreneurs, who can use the insights to refine product and service quality; marketplace platform managers, in formulating more accurate store rating policies; and consumers, who require reliable references in their purchasing decision-making process. By conducting an in-depth analysis of thousands of clothing store reviews on Tokopedia through a Decision Tree approach, this research is expected to contribute significantly to the development of the fashion e-commerce ecosystem in Indonesia, while enriching the application of text mining techniques within the dynamically evolving context of online retail.

Methods

This research utilizes data from the 13 most popular clothing stores on Tokopedia. Each store with a collection period spanning from 2019 to 2024 to ensure the relevance of the findings to current market conditions. The dataset to be analyzed consists of approximately 10,490 reviews, including full review text and star ratings (1-5). Reviews with 1 and 2 stars are categorized into the negative sentiment class, 3-star reviews into the neutral sentiment class, and 4 and 5-star reviews into the positive sentiment class. The total dataset for this research consists of 10,490 reviews from 13 online stores on the Tokopedia marketplace, with detailed store information is provided in Table 1.

Table 1. Number of reviews per sentiment

No	Sentiment	Reviews
1	Positive	4018
2	Neutral	3378
3	Negative	3094

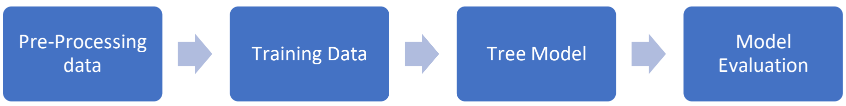

Figure 1. Decision Tree Design Workflow

The flowchart for the Decision Tree classification method design consists of several stages, as illustrated in Figure 1. The first stage is data preprocessing, where the raw data is transformed to meet the specific requirements of the research. In the second stage, the training data is processed using the Decision Tree method. The third stage results in the formation of the tree model. In the final stage, the generated Decision Tree model is evaluated, and a rule base for the classification is established.

A. The data preprocessing stage involves cleaning the review data through the following steps:
   1. Data Cleaning: Removing reviews that contain website links (URLs), special characters, and duplicate entries.
   2. Tokenization: Breaking down the text into a collection of tokens, variables, or individual words.
   3. Case Transformation (Lowercasing): Converting all text into lowercase letters.
   4. Stopword Removal: Filtering out common words or those with no significant meaning using the stopwordbahasa.csv dictionary.
   5. Token Filtering: Selecting only tokens that consist of a minimum of 4 to a maximum of 25 characters.
   6. Data Storage: Finally, the processed data is saved into a single dataset for the subsequent stages of analysis.

B. Training Data

   Following the data preprocessing stage, the data is divided into several subsets, specifically into training data and testing data. This stage ensures that the dataset is fully prepared for the decision tree model construction process. Next, the data are evaluated using a cross-validation method.

C. Tree Models

   The Decision Tree algorithm involves an attribute splitting process to construct the tree structure. This research utilizes three attribute splitting methods, as follows:
   1. Gini Index
   2. Gain Ratio
   3. Information Gain

D. Model Evaluations

   The next stage involves evaluating the three constructed Decision Tree models using a cross-validation technique with k values of 3, 5, and 7. Additionally, evaluation is conducted by applying pruning and non-pruning techniques to the generated tree models

The results of the data preprocessing stage are divided into 3,320 positive sentiment reviews, 2,949 neutral sentiment reviews, and 2,841 negative sentiment reviews. Each sentiment class possesses distinct linguistic characteristics, each displaying specific word frequency statistics. Subsequently, the processes of tokenization, case folding (lowercasing), stopword removal (using an Indonesian dictionary), stemming (Indonesian), and token filtering (minimum of 4 characters) were applied. The TF-IDF (Term Frequency-Inverse Document Frequency) method was utilized to extract features, tokens, or words from each document, as detailed in Table 2. Figure 2 displays the word cloud of the total dataset, where the token 'bagus' appears as the largest, indicating that it is the most frequently occurring token across all documents.

Table 2. Total data statistics

No	word	in documents	total	in class (Negative)	in class (Neutral)	in class (Positive)
1	*bagus*	1931	2046	304	696	1046
2	*sesuai*	1268	1312	426	438	448
3	*bahan*	1113	1152	308	384	460

4	*bahannya*	880	912	210	271	431
5	*banget*	769	835	190	236	409
6	*ukuran*	693	777	279	266	232
7	*barang*	694	758	309	225	224
8	*warna*	579	708	290	250	168
9	*beli*	590	654	257	191	206
10	*tipis*	437	448	252	153	43
11	*harga*	414	435	157	136	142
12	*nyaman*	408	413	37	110	266
13	*adem*	398	399	33	98	268
14	*size*	325	393	151	136	106
15	*baju*	347	392	168	153	71
16	*pengiriman*	365	374	86	120	168
17	*cepat*	322	328	25	93	210
18	*suka*	318	327	33	113	181
19	*kualitas*	316	324	96	96	132
20	*rapi*	306	318	59	116	143

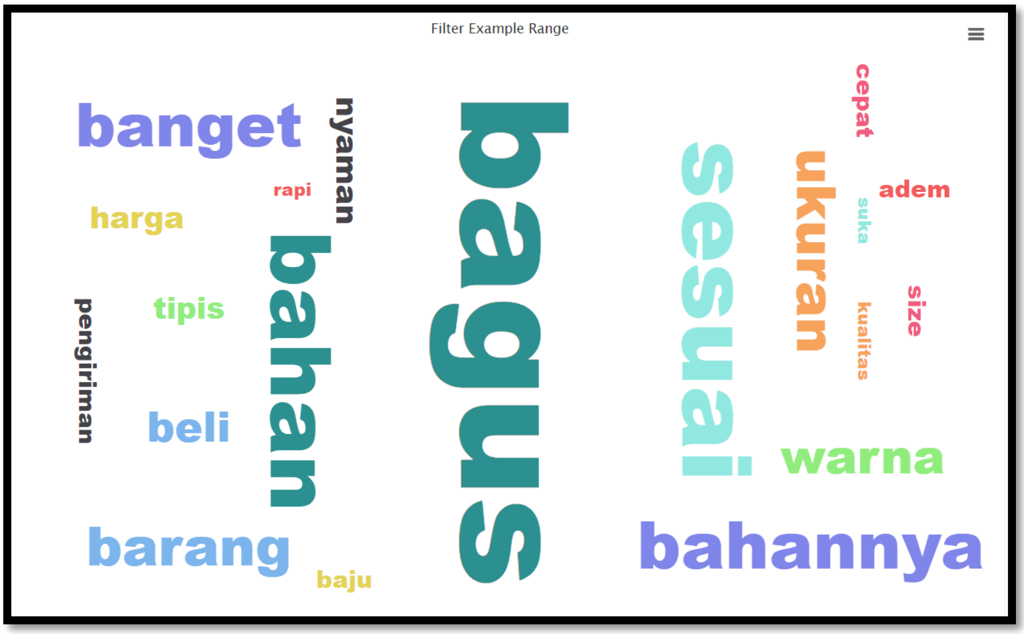

Figure 2. Word Cloud

Results and Discussion

A comparison of the Decision Tree methods before pruning reveals that the Information Gain model achieved the highest accuracy at 44.23%. In the cross-validation stage with k=3, the highest accuracy was also produced by the Information Gain model at 44.2%. For k=5, the Information Gain model maintained the lead with an accuracy of 44.07%. Finally, in the k=7 cross-validation stage, the Information Gain model reached an accuracy of 43.88%. Consequently, as illustrated in Figure 3, the Information Gain model consistently demonstrates the highest accuracy overall. Figure 4 illustrates the resulting decision tree model, with the term "bagus" positioned as the root node, indicating that it is the most influential feature.

The low accuracy score is attributed to certain terms frequently appearing across all sentiment classes, such as the terms "bagus", "sesuai", and "bahan". These terms represent the top three most frequently occurring words in the positive, negative, and neutral sentiments. For the negative and neutral sentiments, the term "bagus" is preceded by the word "tidak", forming a distinct phrase: "tidak bagus". This phrasing pattern also applies to the term "tidak sesuai".

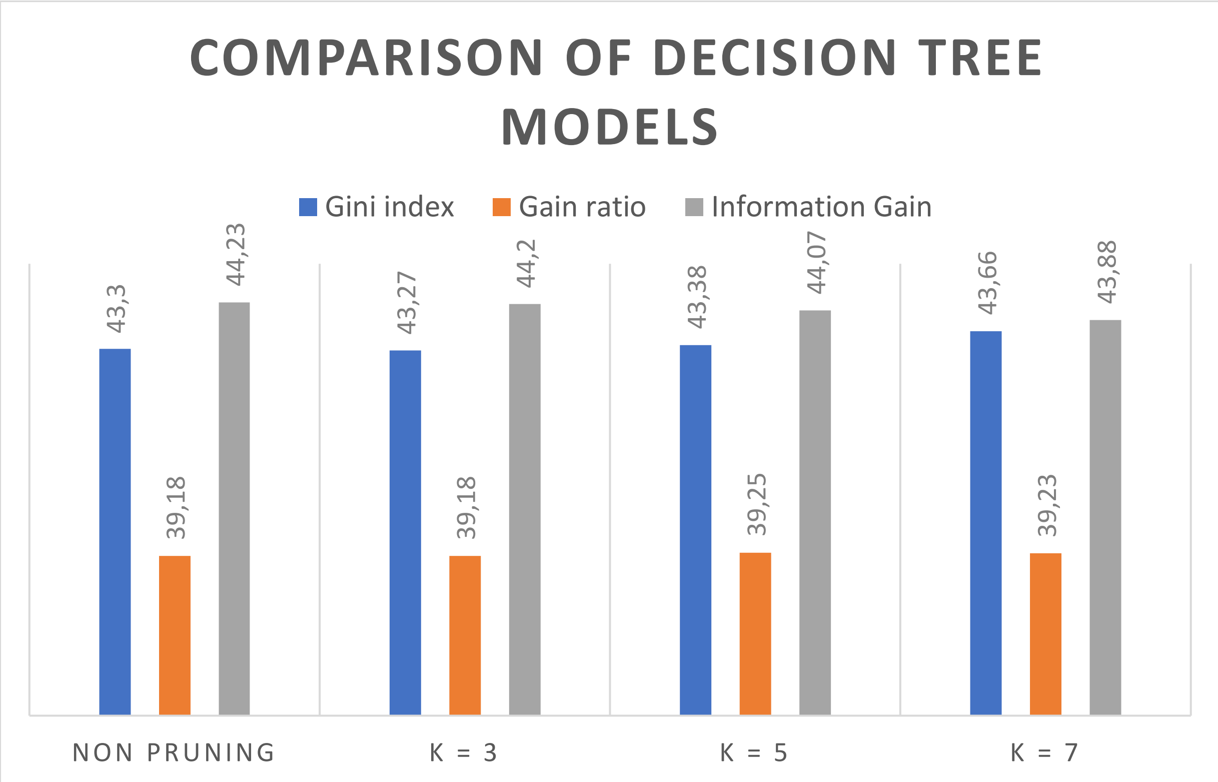

Figure 3. Comparison of Decision Tree models

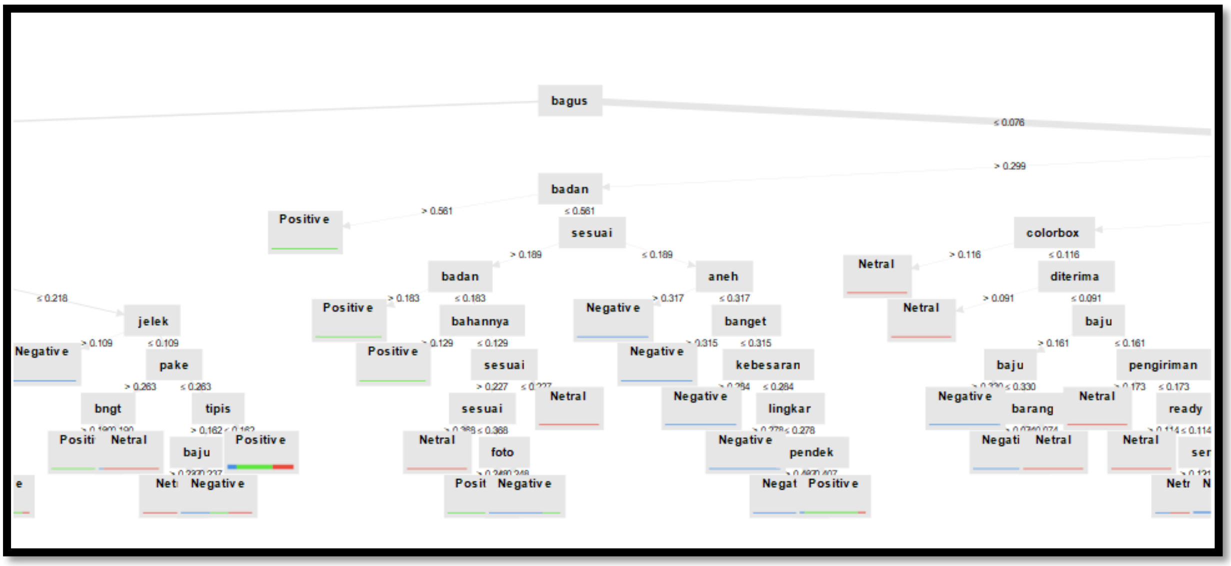

Figure 4. Decison Tree (k=3)

Conclusion

The dataset comprises 10,490 reviews from 13 clothing stores on the Tokopedia marketplace, categorized into 4,018 positive sentiment reviews, 3,378 neutral sentiment reviews, and 3,094 negative sentiment reviews. Following the implementation of the Decision Tree model, the results indicate that the Gini Index algorithm achieved a peak accuracy of

43.66%, the Gain Ratio algorithm reached 39.25%, and the Information Gain algorithm yielded the highest accuracy of 44.23%. Consequently, among the three algorithms tested, Information Gain proved to be the attribute splitting method with the superior accuracy performance.

Since the highest accuracy achieved by the Decision Tree model is only 44.23%, the system is not yet reliable enough to serve as an automated review classifier, meaning sellers still heavily rely on manual curation to accurately gauge consumer sentiment. This low performance is primarily due to the limitation of using solely unigram-based feature extraction, which causes the model to fail in capturing the context of negation phrases (such as "tidak bagus" or "tidak sesuai") and struggle to separate classes due to the high overlap of dominant terms across all sentiments. To address these limitations, future research should implement N-Gram feature extraction (such as bigrams or trigrams) and negation handling techniques during the preprocessing stage, as well as explore more robust algorithms for text data—such as Support Vector Machine (SVM), Ensemble Methods, or deep learning approaches like IndoBERT—to improve overall accuracy performance.

References

Idris, M., Rifai, A., & Tania, K. D. (2025). Sentiment Analysis of Tokopedia App Reviews Using Machine Learning and Word Embeddings. *Sinkron, 9*(1), 210–219. https://doi.org/10.33395/sinkron.v9i1.14278

Ilmania, A., & Cahyawijaya, S. (2018). Aspect Detection and Sentiment Classification Using Deep Neural Network for Indonesian Aspect-Based Sentiment Analysis. In *2018 International Conference on Asian Language Processing (IALP)* (pp. 62–67). IEEE. https://doi.org/10.1109/IALP.2018.8629181

Kurnianingrum, D., Jumbri, I. A., & Utama, I. D. (2025). Comparative Sentiment Analysis of Shopee and Tokopedia in Indonesia: Insights from Twitter Data. *Journal of Information Systems and Informatics, 7*(1), 1–12.

Lestari, A. A., Faqih, A., & Dwilestari, G. (2025). Improving Sentiment Analysis Performance of Tokopedia Reviews Using Principal Component Analysis and Naïve Bayes Algorithm. *Journal of Artificial Intelligence and Engineering Applications, 4*(2), 1–10. https://doi.org/10.59934/jaiea.v4i2.743

Sachin, S., Tripathi, A., Mahajan, N., Aggarwal, S., & Nagrath, P. (2020). Sentiment Analysis Using Gated Recurrent Neural Networks. *SN Computer Science, 1*(3), 1–8. https://doi.org/10.1007/s42979-020-0076-y

Setyawan, D. A., & Fathicah, C. (2020). Pengembangan Metode Decision Tree dengan Diskritisasi Data dan Splitting Atribut Menggunakan Hierarchical Clustering dan Dispersion Ratio. *JUTI: Jurnal Ilmiah Teknologi Informasi, 18*(2), 179–187. https://doi.org/10.12962/j24068535.v18i2.a1005

Ulhaq, D. A. L., & Suprayogi. (2025). Comparing Machine Learning Models for Sentiment Analysis of Tokopedia Reviews. *Journal of Applied Informatics and Computing, 9*(6), 3642–3647. https://doi.org/10.30871/jaic.v9i6.11239

Rohman, A. N., Musyarofah, R. L., Utami, E., & Raharjo, S. (2020). Natural Language Processing on Marketplace Product Review Sentiment Analysis. In *2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICORIS50180.2020.9320827

Warsito, B., & Prahutama, A. (2020). Sentiment Analysis on Tokopedia Product Online Reviews Using Random Forest Method. *E3S Web of Conferences, 202*, 16006. https://doi.org/10.1051/e3sconf/202020216006

Xuanzhen, F., & Xiaohong, L. (2019). Sentiment Classification of Reviews Based on BiGRU Neural Network and Fine-Grained Attention. *Journal of Physics: Conference Series, 1229*(1), 012064. https://doi.org/10.1088/1742-6596/1229/1/012064