International Journal Software Engineering and Computer Science (IJSECS)
5 (3), 2025, 1233-1243
Published Online December 2025 in IJSECS (http://www.journal.lembagakita.org/index.php/ijsecs)
P-ISSN: 2776-4869, E-ISSN: 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v5i3.5690.

RESEARCH ARTICLE **Open Access**

Identification of Key Factors in Children's Toy Product Marketing Strategy through Entropy and Gain Analysis

Siti Aliyah *
Information Systems Study Program, Faculty of Engineering and Computer Science, Potensi Utama University, Medan City, North Sumatra Province, Indonesia.
Corresponding Email: aliyahsiti478@gmail.com.

Efani Desi
Information Systems Study Program, Faculty of Engineering and Computer Science, Potensi Utama University, Medan City, North Sumatra Province, Indonesia.
Email: efanidesi88@gmail.com.

Mas Ayoe Elhias Nst
Information Systems Study Program, Faculty of Engineering and Computer Science, Potensi Utama University, Medan City, North Sumatra Province, Indonesia.

Enni Maisaroh
English Language Education Study Program, Faculty of Social Science and Education, Potensi Utama University, Medan City, North Sumatra Province, Indonesia.

Fitri Pranita Nasution
Informatics Study Program, Faculty of Engineering and Computer Science, Potensi Utama University, Medan City, North Sumatra Province, Indonesia.

Received: October 10, 2025; Accepted: November 15, 2025; Published: December 1, 2025.

Abstract: This study aims to analyze the factors influencing product sales success using the C4.5 algorithm data mining method implemented through the WEKA application. The research data consists of 65 instances with six main attributes, namely introduction, durability, price, size, quality, and description. The testing process is carried out using the 10-fold cross validation method to obtain an accurate classification model. The analysis results show that the Price attribute has the highest information gain value ±0.764, so it is designated as the root of the decision tree. Low prices supported by long product durability proved to be the most dominant combination in increasing sales. Conversely, high prices tended to decrease sales levels even though supported by good quality. The resulting classification model has an accuracy of 83.07%, with 54 data correctly classified out of a total of 65 data. These calculation results indicate that consumers are more sensitive to price than quality, so a marketing strategy that emphasizes competitive pricing with guaranteed product durability is the most effective approach to increase purchasing interest. This research is expected to contribute to business decision making, especially in determining product sales strategies in a competitive market.

Keywords: Data Mining; C4.5 Algorithm; WEKA; Product Sales; Decision Tree.

1. Introduction

The children's toy industry demonstrates substantial growth potential driven by rising numbers of young children and evolving lifestyle patterns. AAD Children's Toy Store operates within these dynamics, working to satisfy consumer demands through diverse educational and entertainment offerings [1]. Products function beyond mere playthings—they serve as educational instruments vital for stimulating cognitive, motor, and social development in children. Consequently, consumers exercise greater selectivity when choosing toys aligned with their children's developmental needs [2][15]. Current market evolution has intensified competition significantly. AAD Children's Toy Store faces challenges from neighboring physical retailers and readily accessible imported products via online platforms. Such competitive pressure demands more than product variety alone [3][4]. Marketing strategies for children's toys are shaped by multiple factors directly correlating with consumer preferences. Purchase decisions frequently consider delivery speed, toy durability, pricing, product dimensions, and perceived quality. Consumers generally favor products offering quality construction, affordability, and longevity [5]. Yet delivery speed gains importance for consumers requiring immediate product access [6][17]. Such preference complexity necessitates that manufacturers identify truly dominant factors influencing purchase decisions [7].

Producers face limitations in identifying key factors among numerous marketing variables. Reliance on intuition or traditional approaches risks misdirected marketing strategies [8]. Data-driven analytical methods become necessary for mapping each criterion's contribution to product appeal. Variables including Delivery (Slow/Fast), Durability (Long-Lasting/Not Durable), Price (Cheap/Expensive), Size (Large/Small), and Quality (Good/Poor) require systematic analysis [9]. The C4.5 Algorithm Method utilizing Entropy and Gain offers a relevant analytical approach for such investigation. Entropy analysis measures data uncertainty levels, while Gain assesses each attribute's contribution to purchase decisions. Combined, these methods help identify marketing attributes exerting the most significant influence [10][19]. Research applying such approaches to children's toy marketing remains scarce, presenting opportunities for both theoretical and practical advancement [11]. Given such background, the research identifies key factors in children's toy product marketing strategies through Entropy and Gain analysis. Five criteria form the analytical focus: delivery, durability, price, size, and quality [12]. Findings aim to advance data-based marketing studies while providing actionable recommendations for industry stakeholders formulating more effective, efficient, and market-oriented strategies [13][18]. The research extends beyond academic contribution to deliver tangible impact on children's toy product competitiveness [14].

2. Related Work

The C4.5 algorithm has been used a lot in different areas for classification and prediction tasks, showing its ability to work with different datasets. Wijaya and Fauzi (2020) used the algorithm to predict battery sales at PT Varta Microbattery Indonesia, setting up a basis for understanding product sales dynamics through decision tree analysis [1]. Izyuddin and Wibisono (2020) used the decision tree method with C4.5 for predicting AC sales, showing that this method can work with electronic product sales forecasting [2]. Sari and Sewaka (2022) applied it to the fashion industry by predicting shoe sales at T&T Collection Tangerang, reaching good accuracy levels that confirmed the algorithm can adapt across different product categories [4]. Beay and Sarimole (2024) also proved the method's reliability by applying decision tree techniques for sales prediction at PT. Cipta Naga Semesta which is part of Mayora Group located in North Jakarta where this model could find out important things driving sales in fast-moving consumer goods sector [6]. Yuni and Putri used the algorithm in 2023 to forecast palm oil production amounts which showed its application in agricultural forecasting [3]. All these studies put together show that the C4.5 algorithm works well for retail and sales predictions over many product categories as well as industries.

Customer behavior analysis and risk assessment are another big area where the C4.5 algorithm is used. Saputra and Fatah (2025) used data mining techniques with C4.5 to find high-risk customers in sales operations which gives good insights for customer relationship management strategies [7]. Siswandi *et al.* applied this algorithm at PT Bayer Indonesia to classify customer characteristics so that it can be useful to segment consumer profiles for targeted marketing approaches [11]. Johannes and Alamsyah created a model of sales prediction using classification decision trees specifically for small-and-medium enterprises based on Indonesian e-commerce data, addressing unique challenges faced by SMEs in digital marketplaces [10]. Madani and Alshraideh got about 91.67% accuracy when predicting consumer purchasing decisions within online food delivery industry; this shows how well the algorithm performs better in digital commerce environments [9]. Adriansa *et al* utilized C4.5 for customer satisfaction analysis thereby linking service quality assessment with data mining techniques [17]. These applications prove that C4.5 captures complex patterns of consumer behavior across various business scenarios very effectively making it highly relevant to purchase decision factors understanding.

Studies comparing C4.5 with other machine learning algorithms have been done mostly to see how good it is relatively and how it can be improved. Sarihandini *et al.* (2024) compared Support Vector Machine, Random Forest, and C4.5 algorithms in predicting customer loss; each method had its own strengths and weaknesses depending on the situation [8]. Zhang and Wu (2021) improved C4.5 decision trees for big data by solving problems related to scalability and computational efficiency when dealing with large-scale datasets [14]. Patel and Prajapati (2020) discussed the use of C4.5 for decision-making in analyzing sales data, stressing that it is more understandable than black-box machine learning models [16]. Tan (2021) looked into using data mining techniques for marketing strategies through decision tree methods, pointing out how important transparent and explainable models are for making business decisions [15]. Pratama and Armansyah (2024) used the decision tree C4.5 with an information gain technique to classify study program selections, proving that this method can be applied in educational decision support systems [5]. From these comparative analyses, it appears that even though newer algorithms may provide higher accuracy in certain situations, C4.5 still has advantages in interpretability and ease of implementation that are particularly important for business applications where understanding the factors behind decisions is as crucial as the accuracy of predictions.

Outside of commercial use, C4.5 has shown its worth in healthcare and service quality assessment fields. Desi and Aliyah (2023) used it to find groups of COVID-19 patients based on symptoms and causes; this validated its effectiveness in diagnostics for healthcare [13]. Islam *et al.* (2022) classified toddlers' nutritional status while Sepharni *et al.* (2022) used the method for heart disease classification—both cases showed that the algorithm is reliable for health-related classification tasks [18][19]. Aliyah *et al.* (2025) looked at staff satisfaction levels with computer technician performance through an analysis using C4.5, which means it has been extended into human resource management applications as well [12]. These different uses highlight how strong and flexible the algorithm is across different fields. However, there has not been much research specifically on children's toy product marketing strategies using entropy and gain analysis. Though past studies have applied C4.5 successfully to various sales prediction tasks and customer behavior analyses, none systematically examined how price, durability, quality, delivery speed, and product size work together to make a children's toy sell well or not—in other words, what factors are most important when selling toys for kids? This gap needs investigating so that actionable intelligence can be provided to toy manufacturers as well as retailers who operate in increasingly competitive markets where understanding dominant purchase factors is crucial for strategic positioning.

3. Research Method

The research method was designed through structured stages based on scientific principles to achieve the research objectives. The procedure is divided into several main stages, briefly described in the research flowchart.

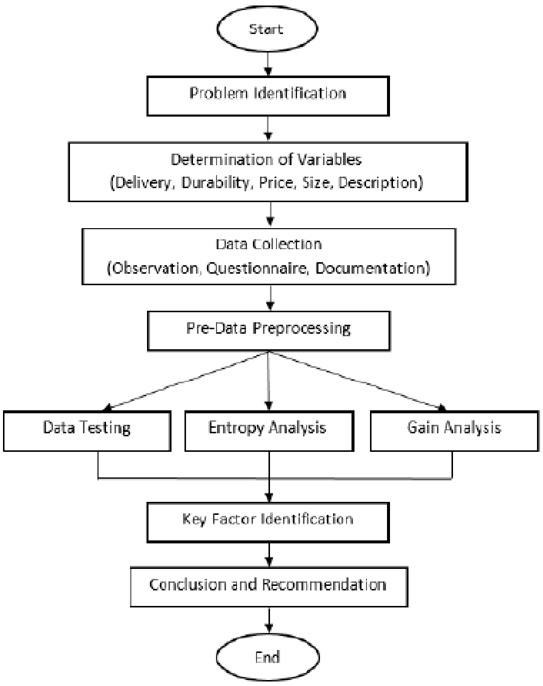

Figure 1. Research Flowchart

3.1 Determination of Variables

Several variables considered influential in evaluating marketing strategies for children's toys were employed based on product quality and consumer preferences. The analyzed variables included:

1) Delivery: Relates to timeliness, speed, and reliability of the product distribution process until reaching consumers. Delivery efficiency serves as an indicator of customer satisfaction and loyalty.
2) Durability: Reflects how long a product can be used without deterioration. High durability represents added value that strengthens the product's quality image.
3) Price: Considered a key factor influencing purchasing decisions. Proportional and competitive pricing can increase product competitiveness in the market.
4) Size: Relates to comfort, safety, and suitability to consumer needs, especially children as primary users.
5) Quality: Encompasses overall product quality, including materials used, safety level, and benefits provided. Product quality builds consumer trust.
6) Description: Indicates sales status of children's toy products, divided into two categories: Sold and Unsold.

3.2 Data Collection

The data collection process was conducted using several techniques to obtain relevant and accurate information. Observations were conducted by directly observing sales activities and consumer interactions with children's toys to obtain realistic pictures of field conditions. Documentation utilized secondary data in the form of relevant records or archives, such as sales reports and supporting documents, serving as complement and verification tools for data obtained from observations. The research data was obtained from sales reports at AAD Children's Toy Store, sourced from sales records for May 2025, documented in an Excel file with 65 data entries as the basis for data processing.

No	Pengantaran	Daya tahan	Harga	Ukuran	Kualitas	Keterangan
1	Lambat	Tidak_Tahan_Lama	Murah	Besar	Tidak_Bagus	Laku
2	Cepat	Tahan_Lama	Murah	Besar	Bagus	Laku
3	Lambat	Tahan_Lama	Mahal	Besar	Tidak_Bagus	Tidak_Laku
4	Lambat	Tidak_Tahan_Lama	Murah	Kecil	Tidak_Bagus	Laku
5	Lambat	Tahan_Lama	Mahal	Besar	Tidak_Bagus	Tidak_Laku
6	Cepat	Tahan_Lama	Murah	Kecil	Bagus	Laku
7	Lambat	Tidak_Tahan_Lama	Mahal	Besar	Tidak_Bagus	Laku
8	Cepat	Tahan_Lama	Murah	Besar	Bagus	Laku
9	Lambat	Tahan_Lama	Murah	Kecil	Bagus	Laku
10	Cepat	Tahan_Lama	Mahal	Kecil	Bagus	Tidak_Laku
11	Lambat	Tahan_Lama	Murah	Besar	Bagus	Laku
12	Cepat	Tahan_Lama	Murah	Besar	Bagus	Tidak_Laku
13	Lambat	Tahan_Lama	Murah	Besar	Bagus	Laku
14	Lambat	Tahan_Lama	Mahal	Kecil	Bagus	Tidak_Laku
15	Cepat	Tahan_Lama	Murah	Besar	Bagus	Laku
16	Lambat	Tahan_Lama	Mahal	Kecil	Bagus	Tidak_Laku
17	Cepat	Tahan_Lama	Murah	Besar	Tidak_Bagus	Tidak_Laku
18	Cepat	Tahan_Lama	Murah	Besar	Bagus	Laku

Figure 2. Raw Data

3.3 Data Pre-Processing

Data pre-processing was performed as a crucial step to ensure data consistency and validity before analysis. The dataset consists of 65 entries with 6 attributes: delivery, durability, price, size, quality, and description. Each attribute has sub-attributes, including delivery (slow, fast), durability (long-lasting, not durable), price (cheap, expensive), size (large, small), quality (good, bad), and description (sold or unsold). Pre-processing normalized the data for readiness in further analysis.

3.4 Data Testing

Data testing was conducted to ensure the dataset met validity and consistency criteria before further analysis. The process aimed to verify accuracy of values for each attribute and assess whether data met research needs, ensuring analysis results were reliable and supported conclusions drawn. The testing phase helped identify irrelevant, redundant, or potentially biased data in the classification process. Thus, entropy and gain calculations generated by the C4.5 algorithm accurately reflected existing data distribution patterns, enabling the resulting decision tree to provide accurate information and serve as valid basis for research decision-making.

3.5 Entropy Analysis

Entropy analysis in the C4.5 algorithm calculates the level of uncertainty for each attribute, determining how significant that attribute is in differentiating data into specific classes. The entropy calculation results serve as the basis for determining the most informative attributes for decision tree formation.

1) Calculate Initial Entropy (S)
   Total number of data = 65
   Sale = 35
   Not Sold = 30

   Entropy is calculated using the formula:

$$H(S) = -\sum_{i=1}^{n} \text{pi} \log_2(pi)$$

$$H(S) = -\left(\frac{35}{65}\log_2\frac{35}{65} + \frac{30}{65}\log_2\frac{30}{65}\right)$$

$$H(S) = 0{,}998$$

2) Calculate Entropy for Price Attribute
   Price attributes are divided into two:
   Price = Cheap
   Sales = 40 (7+33)
   Not Sold = 3
   Total = 43

   Entropy is calculated using the formula:

$$H(Murah) = -\sum_{i=1}^{n} \text{pi} \log_2(pi)$$

$$H(Murah) = -\left(\frac{40}{43}\log_2\frac{40}{43} + \frac{3}{43}\log_2\frac{3}{43}\right)$$

$$H(Murah) = 0{,}353$$

   Sales = 0
   Not Sold = 22
   Total = 22

   Entropy is calculated using the formula:

$$H(Mahal) = -\sum_{i=1}^{n} \text{pi} \log_2(pi)$$

$$H(Mahal) = -\left(\frac{0}{22}\log_2\frac{0}{22} + \frac{22}{22}\log_2\frac{22}{22}\right)$$

$$H(Mahal) = 0$$

3) Calculate Average Entropy after Split (Price)

$$H_{split}(Harga) = \frac{43}{65} X\ H(Murah) + \frac{22}{65} X\ H(Mahal)$$

$$H_{split}(Harga) = \frac{43}{65} X\ 0.353 + \frac{22}{65} X\ 0$$

$$H_{split}(Harga) = 0.234$$

3.6 Gain Analysis

Gain analysis in the C4.5 algorithm assesses the increase in information obtained from selecting an attribute after calculating entropy. The resulting gain value helps determine the attribute with the best ability to separate data, prioritizing that attribute in decision tree formation. The Gain value is obtained by comparing the reduction in uncertainty for each attribute, where the attribute with the highest Gain value is selected as the root or main branch in the decision tree structure. The resulting classification model describes data distribution patterns more clearly, generating accurate and relevant decisions for research objectives. Calculate Gain for the Price Attribute:

$$Gain\ (S,\ Harga) = H(S) - Hsplit(Harga)$$

$$Gain\ (S,\ Harga) = 0.998 - 0.234$$

$$Gain\ (S,\ Harga) = 0.764$$

Because the gain value for price is higher compared to other attributes (delivery, durability, size, and quality), WEKA chooses Price as the root node in the decision tree. Based on calculation results using the C4.5 algorithm, attributes that play significant roles in data classification can be identified. The entropy calculation process indicates the level of uncertainty of each attribute, while the information gain value assesses the attribute's contribution in reducing uncertainty. From the analyzed data, the attribute with the highest Gain value was selected as the key factor because it has the most significant role in distinguishing data into predetermined classes. The factor was then placed as the main node (root) in the decision tree structure, serving as a reference in the subsequent classification process. The key factors identified through calculation can be considered as dominant variables that influence analysis results, providing a clearer picture of which attributes should be given special attention in evaluation processes and data-based decision-making. Conclusions and recommendations are derived from classification results using the C4.5 algorithm, where the attribute with the highest gain value is identified as the key factor. These findings serve as the basis for developing relevant strategic recommendations to support more effective decision-making.

4. Result and Discussion

4.1 Results

The calculations using the C4.5 algorithm produced a decision tree model constructed through several stages of analysis. In the initial stage, the number of cases was calculated, with a total of 65 data points in the description attribute, consisting of 27 data points with Sold status and 38 data points with Unsold status.

Figure 3. Data Processing Results

Based on data processing using the WEKA application with the C4.5 (J48) algorithm in Figure 3, a classification model in the form of a decision tree was obtained to predict sales information for children's toys. The dataset consists of 65 instances with six attributes: Delivery, Durability, Price, Size, Quality, and Description. Testing was carried out using the 10-fold cross-validation method to ensure validity of classification results. The decision tree formation process shows that the most influential attribute in classification is Price, confirming that price is the dominant variable in influencing sales decisions, as follows:

1) If Price = Cheap and Durability = Not Durable, the decision depends on the quality attribute:
   If Quality = Not Good, the product tends to sell in 7 cases.
   If Quality = Good, the product is more likely not to sell in 3 cases.
2) If Price = Cheap but Durability = Long-Lasting, the majority of data shows the product sells more frequently, with 33 cases, proving that combining low price with long durability is an effective strategy for increasing sales.
3) If Price = Expensive, the majority of products were unsold, with 22 cases occurring, aligning with marketing theory stating that high prices are major barriers to consumer purchasing interest, especially without commensurate benefits.

The decision tree structure with four leaves and tree size of seven nodes indicates the resulting model is relatively simple yet informative. The model identifies a pattern where price is the primary variable, supported by durability and quality as derived attributes contributing to sales results. The test results using the C4.5 (J48) algorithm in the WEKA application can be seen in Figure 4 below:

Figure 4. Data Test Results

The test results show that out of 65 data instances analyzed, 54 data (83.07%) were successfully classified correctly, while 11 data (16.92%) were misclassified. The Kappa Statistic value of 0.642 shows fairly strong agreement between model classification results and actual data. The prediction error measured through Mean Absolute Error of 0.2687 and Root Mean Squared Error of 0.3903 is relatively small, meaning the model provides prediction results close to actual conditions. When viewed from detailed accuracy by class, for the Sold class, precision value was 0.814, recall 0.921, and F-Measure 0.864, indicating the model is more accurate in recognizing data falling into the Sold category. Meanwhile, for the Unsold class, precision value was 0.866, recall 0.704, and F-Measure 0.776, indicating that although the model is capable of recognizing Unsold data, some data are still misclassified into other categories. In the confusion matrix, of the 27 data that should have fallen into the Sold category, 25 were successfully classified correctly, while 2 data were incorrectly predicted as Unsold. Conversely, of the 38 data that fell into the Unsold category, 29 were correctly classified, and 9 were incorrectly predicted as Sold. The decision tree structure generated using the C4.5 (J48) algorithm in the WEKA application can be seen in Figure 5 below:

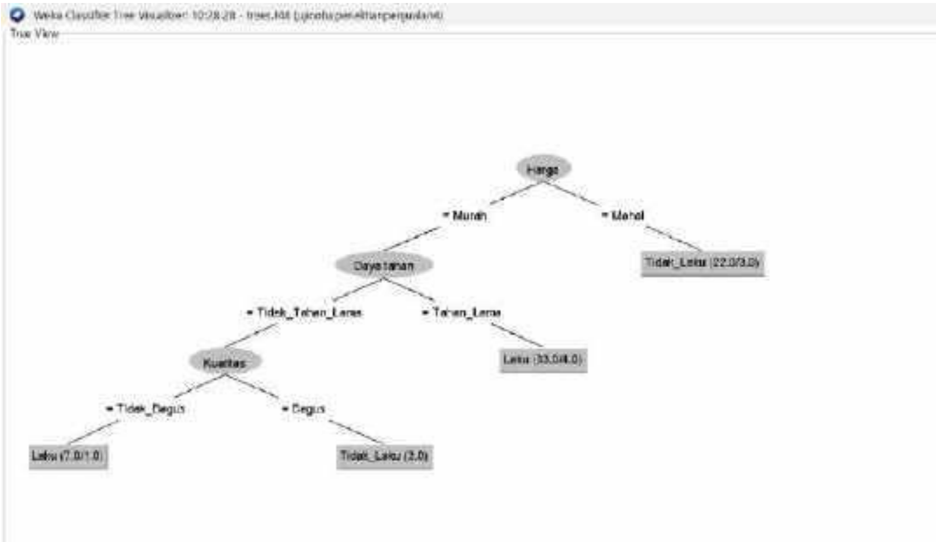

Figure 5. Decision Tree Structure

The structure of the decision tree obtained can be explained as follows:

1) Cheap Price is divided into two conditions based on the Durability attribute:
   If Durability = Not Durable, classification is determined by the Quality attribute:
   If Quality = Not Good, the product tends to sell with a total of 7 cases.
   If Quality = Good, the product is categorized as Unsold in 3 instances, indicating consumer perceptions of mismatch between low price and good quality, leading to doubts about the product.
   If Durability = Long-Lasting, the product is predicted to sell with a dominant number of 33 data points, confirming that the combination of affordable price and good durability is a key factor in increasing consumer appeal.
2) Expensive Price consistently led to Unsold decisions in 22 instances, confirming that high prices are major barriers to purchasing decisions, especially for products competing in markets with price-sensitive consumer preferences.

The classification results show that the decision tree model captures the dominant pattern in sales data, namely the low price factor supported by product durability as determinant of sales success. The resulting model is relatively simple with 4 leaves and 7 nodes, but informative enough to be used as basis for marketing strategy decisions.

4.2 Discussion

The results of the analysis indicate that price is the most important attribute in determining the success of children's toy sales at AAD Children's Toy Store. The Gain value for the price attribute is 0.764, which significantly exceeds other attributes (delivery, durability, size, and quality) and hence makes this variable the root node in the decision tree structure. This finding can be supported by consumer behavior theory in retail markets where price sensitivity becomes a major consideration when products are directed toward middle-income families. The decision tree indicates that expensive products have an unsold status across all 22 cases without exception; thus, there is strong consumer resistance to high-priced children's toys irrespective of any other product attributes. This pattern also indicates that AAD Children's Toy Store operates within a segment of the market that is price-sensitive and where affordability takes precedence over premium features. In addition to price being the most important factor, the decision tree also reveals more intricate relationships between multiple attributes that come into play when making purchasing decisions. For cheap products, durability becomes the second determinant creating two different patterns of consumer response. When cheap products offer long-lasting durability, sales success is remarkably high with 33 cases which represent the best combination in this dataset; thus showing a pattern wherein consumers prefer value-for-money propositions reinforced by product longevity at affordable prices. On the other hand, cheap products with low durability trigger quality-based decision-making resulting in counterintuitive results: not good quality sold in 7 cases while good quality remained unsold in 3 cases. This pattern is quite surprising and may be indicative of consumer skepticism toward cheap products that claim to have good quality—possibly reflecting perceptions of a mismatch between quality and price or even distrust in the authenticity of the product itself. Consumers may see unusually good quality at low prices as signs of defects, safety issues, or misleading descriptions about what they are buying; hence they would avoid making purchases even if there seems to be value.

C4.5 algorithm got 83.07% classification accuracy with 54 instances correctly classified out of 65 total cases which shows a satisfactory model performance for practical business applications. The Kappa Statistic of 0.642 means there is substantial agreement between what was predicted and what really happened in the classifications; this value also exceeds the threshold for acceptable model reliability in commercial decision support systems. Performance metrics show asymmetric classification capabilities across categories. For Sold class, precision is 0.814 recall, recall 0.921 and F-Measure at 0.864 indicating strong model sensitivity to identify successful sales cases that minimize false negatives so that an opportunity to market may not be missed. The Unsold class indicates precision at 0.866 recall at 0.704 and F-Measure at 0.776, which means this model has high specificity but lower sensitivity for unsold products. Analyzing the confusion matrix reveals nine false positives where unsold products were incorrectly predicted as sold and only two false negatives where sold products were misclassified as unsold; hence, this pattern of error distribution suggests that the model has optimistic bias by possibly overestimating sales potential for marginal cases. Such bias in business applications may cause inventory overstock on products with borderline characteristics though relatively low error rates (16.92% overall misclassification) are still acceptable for strategic planning purposes.

These results help in making decisions about the marketing strategy of AAD Children's Toy Store. First, competitive pricing strategies should focus on affordability as the main value proposition because price has an overwhelming influence on purchase decisions. Products priced above the market average are very likely to fail in terms of sales, regardless of any quality advantages they may have; hence, premium pricing strategies will not work for this customer base. Second, product selection and procurement should emphasize durability as a critical secondary attribute for budget-priced items. The combination of cheap price and long-lasting durability is the optimal market positioning and captures 33 sales cases proving clear consumer preference for value-oriented offerings. Third, quality communication strategies need careful calibration for low-priced products. It is counterintuitive but true that good quality reduces sales for cheap, non-durable products; this suggests consumer skepticism about quality claims that seem inconsistent with the pricing. Marketing communications should stress realistic quality expectations in line with price points rather than overpromising product excellence. Fourth, inventory management needs to differentiate stocking strategies based on attribute combinations. High-priority inventory should be cheap and long-lasting products carrying moderate quality claims while expensive products should be minimized or eliminated from portfolios altogether if not avoided completely in cheap price but poor durability product categories which require careful quality positioning so as not to invoke consumer distrust. These strategic recommendations surface explicitly from data-driven analyses rather than intuitions and provide empirical bases for resource allocation decisions in competitive retailing environments.

5. Conclusion

Based on manual calculations and the C4.5 (J48) algorithm in the WEKA application, the Price attribute achieved the highest gain value of 0.764, making it the root of the decision tree and confirming price as the most dominant factor influencing children's toy sales decisions at AAD Children's Toy Store. Durability and Quality serve as secondary factors that strengthen classification, where children's toys with low prices and long durability demonstrate the greatest likelihood of selling (33 cases), while high-priced toys consistently remain unsold (22 cases) regardless of quality attributes. This pattern indicates consumers prioritize affordability over quality, particularly for children's toys oriented toward mass consumption in price-sensitive markets. The classification model achieved 83.07% accuracy, with 54 instances correctly classified out of 65 total data points, demonstrating satisfactory performance of the C4.5 algorithm in predicting sales outcomes based on analyzed factors. The Kappa Statistic of 0.642 confirms substantial agreement between predicted and actual classifications, while relatively low error rates (Mean Absolute Error: 0.2687; Root Mean Squared Error: 0.3903) validate model reliability for practical business applications. Performance metrics reveal stronger predictive capability for the Sold class (F-Measure: 0.864) compared to the Unsold class (F-Measure: 0.776), suggesting the model effectively identifies successful sales patterns while maintaining acceptable accuracy across both categories.

These findings establish that competitive pricing strategies combined with adequate product durability represent the most effective approach to increasing sales success in the children's toy retail sector. The research provides empirical evidence that price sensitivity dominates consumer decision-making processes, necessitating strategic focus on value-for-money propositions rather than premium quality positioning. Marketing strategies should prioritize affordable pricing as the primary value proposition, emphasize durability as the critical secondary attribute, and calibrate quality communications to align with price expectations. The decision tree model generated through entropy and gain analysis offers actionable intelligence for inventory management, product selection, and market positioning decisions, enabling data-driven strategy formulation in highly competitive and price-sensitive retail environments.

References

[1] Wijaya, B., & Fauzi, R. (2020). Data Mining Algoritma C4. 5 Untuk Memprediksi Penjualan Baterai Di Pt Varta Microbattery Indonesia. *Computer and Science Industrial Engineering (COMASIE)*, *3*(2), 64-74.

[2] Izyuddin, A., & Wibisono, S. (2020). Aplikasi prediksi penjualan ac menggunakan decision tree dengan algoritma c4. 5. *Jurnal Manajemen Informatika dan Sistem Informasi*, *3*(2), 146-156.

[3] Yuni, R., & Putri, R. A. (2023). Penerapan algoritma C4.5 untuk prediksi jumlah produksi kelapa sawit. *Jurnal Media Informatika Budidarma*, *7*(4), 1749. https://doi.org/10.30865/mib.v7i4.6861.

[4] Sari, A. I., & Sewaka. (2022). Prediksi penjualan sepatu dengan algoritma C4.5 di T&T Collection di Tangerang. *OKTAL: Jurnal Ilmu Komputer dan Sains*, *1*(09), 1435–1442.

[5] Pratama, T., & Armansyah, A. (2024). Decision tree C4.5 with information gain technique for advanced study program selection classification. *Journal of Information System Research (JOSH)*, *5*(4), 1042–1052.

[6] Beay, R., & Sarimole, F. M. (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.

[7] Saputra, Z. N., & Fatah, Z. (2025). Pengunaan Data Mining Untuk Mengidentifikasi Pelanggan Beresiko Tinggi Dalam Penjualan Mengunakan Algoritma Decision Tree C4.5. *JUSIFOR : Jurnal Sistem Informasi Dan Informatika*, *4*(1), 68–74. https://doi.org/10.70609/jusifor.v4i1.5942.

[8] Maulana, B., Febrian, D., Fachrezi, I. R., & Zeen, M. F. (2025). Comparison of Support Vector Machine, Random Forest, and C4. 5 Algorithms for Customer Loss Prediction. *IJATIS: Indonesian Journal of Applied Technology and Innovation Science*, *2*(1), 1-6. https://doi.org/10.57152/ijatis.v2i1.1102.

[9] Madani, B., & Alshraideh, H. (2021). Predicting consumer purchasing decision in the online food delivery industry. *arXiv preprint arXiv:2110.00502*. https://doi.org/10.48550/arXiv.2110.00502

[10] Johannes, R., & Alamsyah, A. (2021). Sales Prediction Model Using Classification Decision Tree Approach For Small Medium Enterprise Based on Indonesian E-Commerce Data. *arXiv preprint arXiv:2103.03117*. https://doi.org/10.48550/arXiv.2103.03117

[11] Siswandi, A., Anwar, M. S., Susilo, A., & Hasibuan, S. (2024). Application of the C 4.5 Algorithm to Classify Customer Characteristics at PT. Bayer Indonesia. *Brilliance: Research of Artificial Intelligence*, *4*(1), 312-320. https://doi.org/10.47709/brilliance.v4i1.4174.

[12] Aliyah, S., Desi, E., Nasution, F. P., & Tahel, F. (2025). Analysis of the C4.5 algorithm in measuring the level of staff satisfaction with the performance of computer technicians. *Indonesian Journal: Informatics and Communication Management*, *6*(2), 334–345.

[13] Desi, E., & Aliyah, S. (2023). Implementasi Algoritma C4.5 untuk Medeteksi Golongan Penderita Covid-19 Berdasarkan Gejala dan Penyebabnya. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, *7*(3), 404-413. https://doi.org/10.35870/jtik.v7i3.872.

[14] Zhang, Y., & Wu, L. (2021). Improved C4.5 decision trees for big data analysis. *Procedia Computer Science*, *187*, 35–42.

[15] Tan, K. C. (2021). Using data mining for marketing strategy: A decision tree approach. *Journal of Business Analytics*, *4*(2), 98–109.

[16] Patel, M., & Prajapati, H. (2020). Application of C4.5 algorithm for decision making in sales data analysis. *International Journal of Data Science and Analytics*, *7*(3), 175–184.

[17] Adriansa, M., Yulianti, L., & Elfianty, L. (2022). Customer satisfaction analysis using the C4.5 algorithm. *Journal of Informatics Engineering, Santo Thomas Catholic University*, *7*(1), 22–31.

[18] Islam, H., Mulyadien, M., & Enri, U. (2022). Penerapan Algoritma C4.5 dalam Klasifikasi Status Gizi Balita. *Jurnal Ilmiah Wahana Pendidikan*, *8*(10), 116-125. https://doi.org/10.5281/zenodo.6791722.

[19] Sepharni, A., Hendrawan, I. E., & Rozikin, C. (2022). Classification of heart disease using the C4.5 algorithm. *STRING Journal (Research Writing Unit and Technology Invasion)*, *7*(2), 201–210.