Intechno Journal :
Information Technology Journal
Vol. 8 No. 1 July 2026
Homepage: https://jurnal.amikom.ac.id/index.php/intechno
INTECHNO Journal
INFORMATION TECHNOLOGY JOURNAL
e-ISSN: 2655-1438
p-ISSN: 2655-1632

Web Based Smartphone Recommendation System Using C4.5 Decision Tree Algorithm for User Need Classification

Resca Amelina[1], Usman[2], Bayu Rianto[3]
[1,2,3]Department of Information Systems, Faculty of Engineering and Computer Science, Universitas Islam Indragiri, Riau, Indonesia

Article Info

Article history:

Received: 28 June 2026
Revised: 13 July 2026
Accepted: 16 July 2026
Published: 31 July 2026

Keywords:

Decision Tree C4.5
Smartphone Recommendation
Classification
Machine Learning
Web Based System

ABSTRACT

The rapid growth of the smartphone market has created significant difficulty for retail staff in providing consistent, objective purchase recommendations, particularly for customers with limited technical knowledge. This study presents SmartPick, a recommendation system that applies the C4.5 Decision Tree algorithm to address the inefficiency of manual recommendation at Sahabat Ponsel Tembilahan retail store. Purpose: To classify users into smartphone usage categories Gaming, Photography/Videography, Student, and General User based on budget, usage intent, and technical preferences, replacing subjective salesperson judgment with a consistent, rule-based decision process. Methods: A balanced dataset of 500 smartphones was used to train and evaluate the C4.5 model, chosen for its interpretability and ability to represent decision logic as human-readable rules. Results: The model achieved 96.00% testing accuracy, outperforming CART, Random Forest, Naive Bayes, and K-NN. Novelty: Unlike prior recommendation approaches relying on collaborative or content-based filtering, which require extensive historical transaction data, SmartPick integrates technical specifications and budget constraints directly into a single interpretable Decision Tree model, offering a transparent and scalable alternative for retail-based smartphone recommendation that remains explainable to non-technical users.

Corresponding Author:

Resca Amelina
Email: rescaamelina@gmail.com

1. INTRODUCTION

The rapid advancement of information technology has transformed consumer electronics retail, especially the smartphone market. In 2023, Indonesia recorded over 352 million mobile subscribers, reflecting the rapid penetration of smartphones as the primary communication device [1]. Hundreds of new models are launched annually with diverse specifications and prices, creating substantial consumer confusion. Sahabat Ponsel Tembilahan faces challenges delivering accurate recommendations through manual salesperson consultations. Staff must simultaneously weigh RAM, storage, processor, battery capacity, camera quality, intended usage, and budget a process that is time-consuming and inconsistent. Customers outside Tembilahan cannot receive recommendations without visiting the store in person. Classification-based recommendation systems offer an objective, scalable solution. Most prior smartphone recommendation studies rely on collaborative or content-based filtering, which require large historical transaction data and do not integrate technical specifications, usage intent, and budget into a single interpretable model [2], [3]. Decision Tree has also been applied to recommendation tasks in other domains [4]. For instance, Amperawansyah & Putri[5] achieved 95% accuracy using Forward Chaining and Decision Tree for smartphone selection, but limited their scope to the 1–3 million IDR price range only. Ramadhan et al. [6] applied Decision Tree to classify smartphone prices with an average

accuracy of 81%, yet focused solely on price without incorporating usage intent or user-need categories. Zuhdiansyah & Luthfiarta [2] proposed a collaborative filtering approach using K-Means and SVD, which is vulnerable to cold-start problems when new devices are added to the catalog.

The C4.5 Decision Tree algorithm is well-suited for datasets with mixed attribute types. Unlike ID3, C4.5 uses Gain Ratio to avoid bias toward high-cardinality attributes and supports cost-complexity pruning to prevent overfitting [7], [8]. C4.5 was selected over alternative classifiers for this task for three reasons. First, unlike ensemble methods such as Random Forest, C4.5 produces a single tree whose if-then decision path can be directly inspected and explained to non-technical retail staff and customers, which is essential for a recommendation context where users expect a transparent justification. Second, unlike distance-based methods such as K-NN or probabilistic methods such as Naive Bayes both of which assume feature independence or scale-sensitive distances, C4.5 natively handles the mixed continuous and ordinal-encoded categorical attributes in the dataset (e.g., price, RAM, chipset tier) without requiring feature scaling. Third, C4.5's Gain Ratio splitting criterion and cost-complexity pruning give it a built-in mechanism to control overfitting on a moderately sized dataset, in contrast to CART's Gini-based splits, which can favor high-cardinality attributes. These properties are validated empirically in Section 3.6, where C4.5 is compared against CART, Random Forest, Naive Bayes, and K-NN on identical data splits. To fill this gap, this paper proposes SmartPick, a C4.5-based smartphone recommendation model that classifies user needs into four usage categories: Gaming, Photography/Videography, Student, and General User. The novelty of this work lies in integrating multi-variable C4.5 classification (14 engineered features across 500 balanced devices from 16 brands) into a single interpretable model, making the recommendation decision process fully transparent to end users a feature absent from prior smartphone recommendation systems that rely on opaque filtering or ensemble methods. The proposed model was implemented as a web-based prototype, SmartPick, to validate its real-world feasibility in a retail setting.

2. METHOD

2.1 Research Framework

Research followed a structured pipeline: (1) Literature Study, (2) Pre-research Analysis, (3) Data Collection, (4) Data Preprocessing, (5) C4.5 Model Training and Evaluation, (6) System Design, (7) System Implementation, (8) System Testing, and (9) Conclusion. The research framework is illustrated in Figure 1.

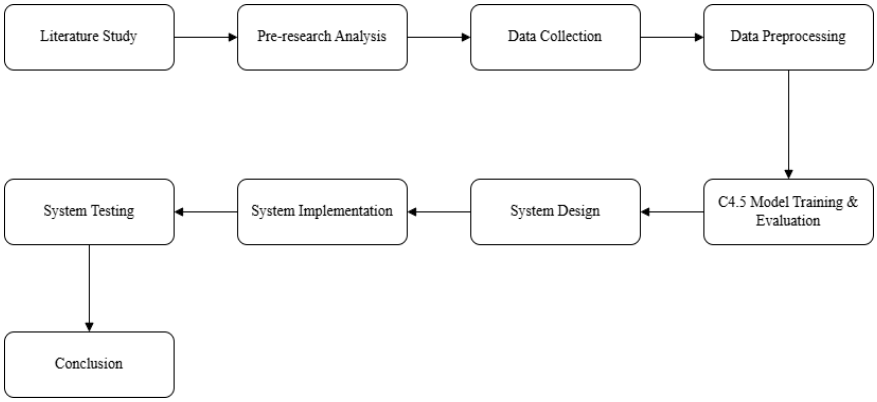

Figure 1. Research Framework

2.2 Dataset

The dataset comprises 500 smartphones from the Sahabat Ponsel Tembilahan catalog and curated public datasets (Google, Kaggle and, GSMArena). Four target classes were defined with equal distribution (125 records each), as shown in Table 1.

Table 1. Dataset Distribution

No	Class	Count	Percentage
1	Gaming	125	25%
2	Photography/Videography	125	25%
3	Student	125	25%
4	General User	125	25%
	Total	**500**	**100%**

Gaming requires flagship chipsets, RAM ≥ 8 GB, and fast charging ≥ 65 W (price Rp 3–28 million). Photography/Videography prioritizes high-resolution multi-lens cameras (avg. 89.5 MP) and OLED display (price Rp 2–32 million). Student emphasizes affordability (price < Rp 2.1 million, avg. Rp 1.6 million) with entry-level chipsets. General User covers balanced mid-range devices (RAM 4–12 GB, price Rp 1.2–10 million). A sample size of 500 (125 per class) was constrained by the number of devices available across the combined Sahabat Ponsel Tembilahan, Kaggle, and GSMArena sources, and by the manual labeling process required to assign each device to a usage category guided by published reviews discussing device suitability for gaming, photography, or general use, and cross-checked with Sahabat Ponsel Tembilahan staff. Despite this constraint, stratified sampling preserved equal class representation, and the narrow 5-Fold Cross-Validation spread (94.82% ± 1.76%) together with the small train-test accuracy gap (2.00%) indicate that the model is not simply overfitting to a small sample. Nonetheless, 500 records remain modest relative to the full diversity of commercially available smartphones, so the reported accuracies should be interpreted as evidence of internal validity rather than population-level generalization. Evaluating the C4.5 model on progressively larger datasets (e.g., 1,000 and 2,000 samples) to characterize scalability and overfitting risk as sample size grows is identified as important future work.
Dataset splitting used stratified sampling, as shown in Table 2.

Table 2. Stratified Dataset Split (70:15:15)

No	Class	Training (70%)	Validation (15%)	Testing (15%)	Total
1	Gaming	88	19	18	125
2	Photography/Videography	87	19	19	125
3	Student	88	18	19	125
4	General User	87	19	19	125
	Total	**350**	**75**	**75**	**500**

2.3 Data Preprocessing

Preprocessing steps Missing value check no missing values or duplicates found, Categorical encoding Processor Chipset mapped to a 5-level performance score (1=Entry to 5=Flagship Ultra) and Screen Type to a display quality score (1–5), Feature engineering four derived features: Camera_Total (front+rear MP), Storage_RAM_Ratio, Price_per_RAM, and Price_Category (tier 1–5). The final 14 input features are listed in Table 3.

Table 3. Input Features Used in C4.5 Algorithm

No	Feature	Description
1	Processor_Score	Chipset performance score (1=Entry to 5=Flagship Ultra)
2	Screen Size (inch)	Display size in inches
3	Screen_Score	Display panel quality score (1–5)
4	RAM (GB)	RAM capacity in Gigabytes
5	Storage (GB)	Internal storage in Gigabytes

6	Battery (mAh)	Battery capacity in mAh
7	Fast Charging (W)	Fast-charging wattage
8	Rear Camera (MP)	Rear camera resolution in Megapixels
9	Front Camera (MP)	Front camera resolution in Megapixels
10	Price (IDR)	Smartphone price in Rupiah
11	Camera_Total	Front + rear camera sum (engineered)
12	Storage_RAM_Ratio	Storage ÷ RAM (engineered)
13	Price_per_RAM	Price per GB of RAM (engineered)
14	Price_Category	Price tier 1–5 (engineered)

2.4 C4.5 Algorithm

C4.5 has been widely applied across various classification domains including student performance prediction [9], [10]employee evaluation , athlete classification [11], and customer satisfaction analysis [12]. C4.5 selects the optimal splitting attribute using Gain Ratio (GR):

$$GR(S,A) = Gain(S,A) / SplitInfo(S,A) \quad (1)$$

where Information Gain is $Gain(S,A) = H(S) - \Sigma(|S_v|/|S|)\times H(S_v)$ and Entropy $H(S) = -\Sigma\, p_i \log_2(p_i)$. The initial entropy of the balanced 4-class dataset was 2.0000 bits. Cost-complexity pruning (ccp_alpha) was applied post-training.

Hyperparameter tuning via grid search yielded the optimal configuration in Table 4.

Table 4. Best Hyperparameters (Grid Search)

Parameter	Value
criterion	entropy
max_depth	5
min_samples_split	2
min_samples_leaf	3
ccp_alpha	0.005
random_state	42

2.5 Evaluation Metrics

Performance was evaluated using Accuracy, Precision, Recall, F1-Score (per class and macro-average), Confusion Matrix, ROC-AUC (One-vs-Rest), Cohen's Kappa, and 5-Fold Cross-Validation. Comparative studies have shown that C4.5 consistently outperforms or is competitive with Random Forest, Naive Bayes, and other algorithms in various classification task[13], [14], [15], [16], and produces highly accurate, interpretable models across diverse classification domains [17], [18]. Four comparative algorithms (CART, Random Forest, Naive Bayes, K-NN) were evaluated on identical splits.

2.6 System Architecture

SmartPick uses a three-tier architecture Frontend HTML/CSS/JavaScript with dark/light themes, Backend Flask (Python) REST API with endpoint /api/predict, Model Inference C4.5 algorithm serialized in pickle

(.pkl) format. The 6-step quiz constructs a 14-dimensional feature vector submitted to the backend, which returns predicted class, confidence score, and ranked recommendations filtered by budget.

3. RESULTS AND DISCUSSIONS

3.1 Decision Tree Structure

Table 5 summarizes the resulting decision tree. Price (IDR) was selected as the root node (GR = 0.8460, threshold Rp 2,149,000), effectively separating the Student class from higher-priced segments. The tree has depth 5, 29 nodes, and 15 leaf nodes, remaining compact and interpretable.

Table 5. Decision Tree Structure

Characteristic	Value
Tree Depth	5
Number of Nodes	29
Number of Leaf Nodes	15
Number of Input Features	14
Root Node Feature	Price (IDR)
Root Node Gain Ratio	0.8460

The decision tree visualization is shown in Figure 2. For readability, the visualization is truncated to a maximum depth of 3, while the trained model itself retains its full depth of 5 as reported in Table 5.

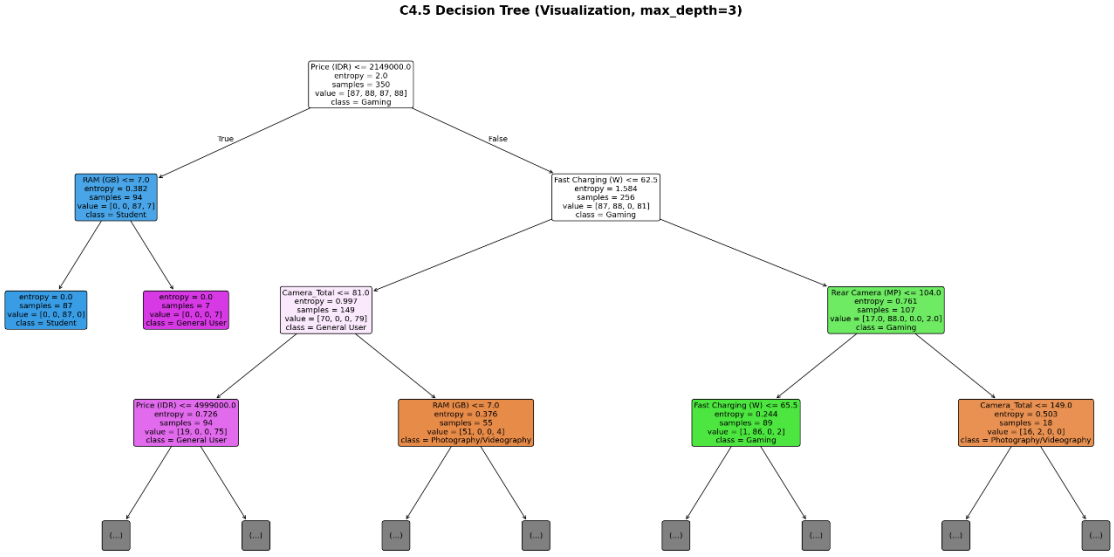

Figure 2. C4.5 Decision Tree Visualization

Figure 2. C4.5 Decision Tree Visualization (truncated to depth 3 for readability; the trained model uses depth 5 as reported in Table 5).

3.2 Algorithm Accuracy

Table 6 presents accuracy of the C4.5 algorithm across all evaluation subsets. The training testing gap of 2.00% and CV standard deviation of ±1.76% confirm robust generalization without significant overfitting.

Table 6. C4.5 Algorithm Accuracy Results

Subset	Data Count	Accuracy
Training	350	98.00%
Validation	75	97.33%
Testing	75	96.00%
5-Fold CV	425	94.82% (±1.76%)

3.3 Confusion Matrix

Table 7 shows the confusion matrix for 75 testing samples. Only 3 misclassifications occurred, all between classes with overlapping feature ranges. Figure 3 shows the heatmap visualization.

Table 7. Confusion Matrix Testing Data

Actual \ Predicted	Photo/Video	Gaming	Student	Gen. User
Photography/Videography	18	0	0	1
Gaming	1	18	0	0
Student	0	0	19	0
General User	0	0	1	17

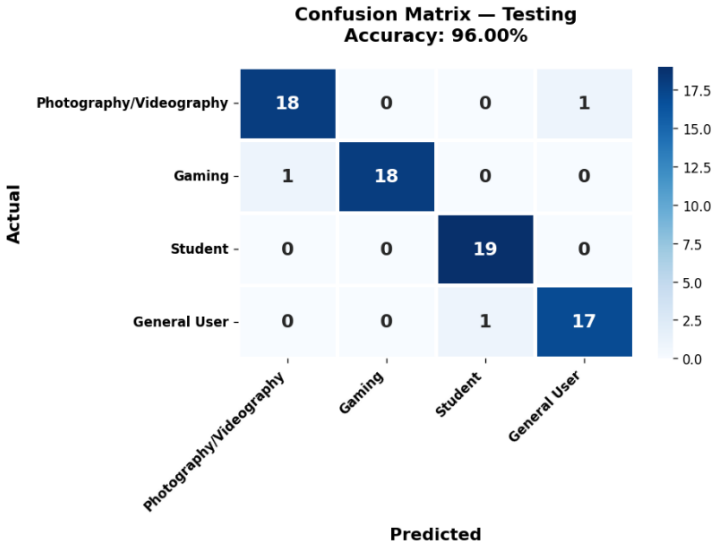

Figure 3. Confusion Matrix Heatmap

Figure 4 shows the correct vs. incorrect prediction counts per class. The Student class achieved zero misclassifications.

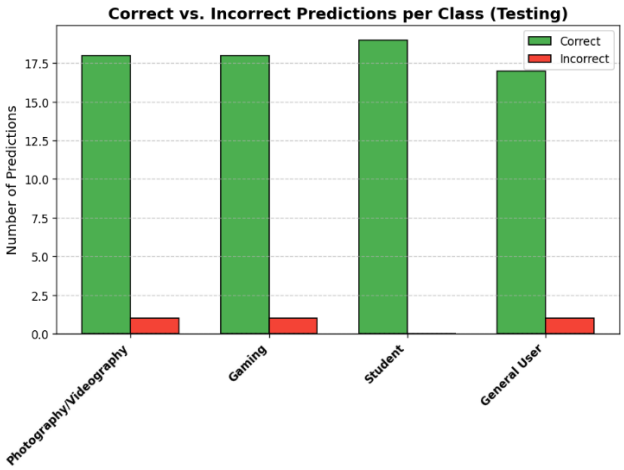

Figure 4. Correct vs. Incorrect Predictions per Class

3.4 Precision, Recall, and F1-Score

Table 8 presents per-class evaluation metrics. Gaming achieved perfect Precision (100.00%); Student achieved perfect Recall (100.00%). General User had the lowest F1-Score (94.00%) due to feature overlap.

Table 8. Per-Class Evaluation Metrics

Class	Precision	Recall	F1-Score	Support
Photography/Videography	95.00%	95.00%	95.00%	19
Gaming	100.00%	95.00%	97.00%	19
Student	95.00%	100.00%	97.00%	19
General User	94.00%	94.00%	94.00%	18
Average	**96.00%**	**96.00%**	**96.00%**	**75**

Table 9 presents ROC-AUC values per class using One-vs-Rest (OvR) approach. Figure 5 shows the ROC Curve visualization.

Table 9. ROC-AUC Values per Class

Class	AUC Value
Photography/Videography	0.972
Gaming	1.000
Student	0.991
General User	0.961
Macro-Average AUC	**0.981**

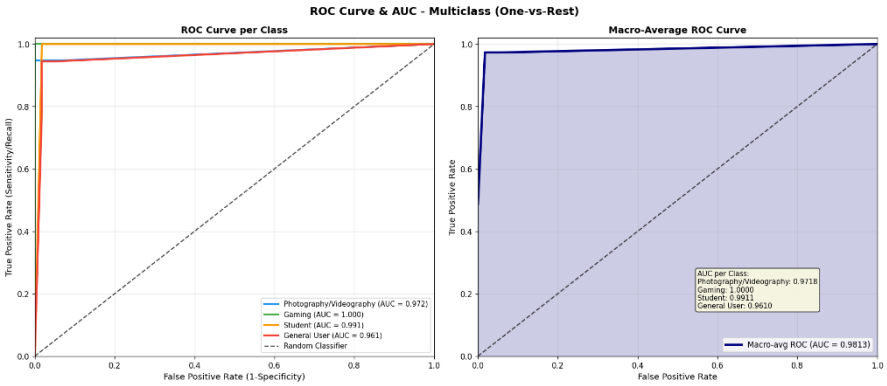

Figure 5. ROC Curve and AUC Multiclass C4.5 Model

3.5 5-Fold Cross-Validation

The learning curve (Figure 6) shows training and CV scores converging as data volume increases. C4.5 achieved a mean CV score of 94.82% (±1.76%) across 5 folds, confirming model stability. Table 10 presents detailed per-fold scores for all compared models.

Table 10. 5-Fold Cross-Validation Scores

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean ± Std
C4.5	**94.12%**	**95.29%**	**91.76%**	**96.47%**	**96.47%**	**94.82% (±1.76%)**
CART	97.65%	94.12%	96.47%	96.47%	96.47%	96.24% (±1.15%)
Random Forest	94.12%	95.29%	95.29%	96.47%	97.65%	95.76% (±1.20%)
Naive Bayes	62.35%	78.82%	71.76%	68.24%	72.94%	70.82% (±5.44%)
K-NN (k=19)	62.35%	71.76%	67.06%	74.12%	70.59%	69.18% (±4.10%)

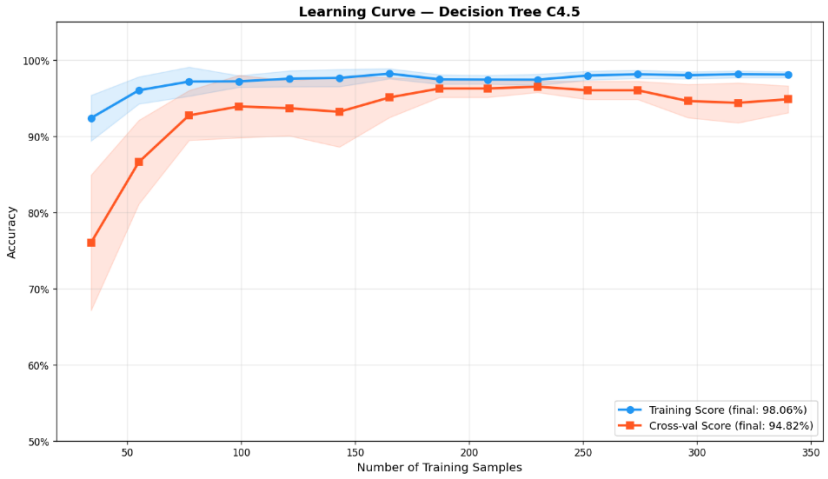

Figure 6. Learning Curve Decision Tree C4.5

3.6 Algorithm Comparison

Table 11 compares C4.5 against five algorithms on identical data splits. C4.5 achieved the highest testing accuracy (96.00%). Figure 7 visualizes the comparison.

Table 11. Algorithm Accuracy Comparison

Algorithm	Train	Val	Test	CV Mean	Prec	Recall	F1
C4.5	**98.00%**	**97.33%**	**96.00%**	**94.82%**	**96.07%**	**96.00%**	**96.00%**
CART	96.86%	97.33%	94.67%	96.24%	94.65%	94.67%	94.63%
Random Forest	100.00%	97.33%	94.67%	95.76%	94.65%	94.67%	94.63%
Naive Bayes	73.14%	62.67%	66.67%	70.82%	73.24%	66.67%	66.51%
K-NN (k=19)	74.29%	61.33%	69.33%	69.18%	70.50%	69.33%	68.60%

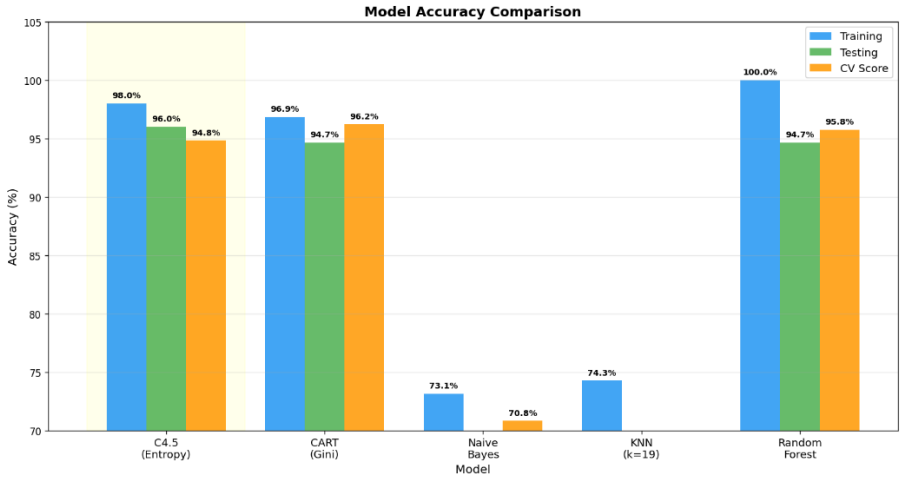

Figure 7. Algorithm Accuracy Comparison Chart

3.7 System Implementation

3.7.1 Homepage

The C4.5 model was deployed as a three-tier web prototype (Section 2.6) to validate real-world feasibility: a six-step quiz collects user inputs, a Flask backend runs inference, and results are returned with a predicted class, confidence score, and ranked recommendations. The prototype is available at smartpicksahabatponsel.my.id Figure 8 shows the homepage.

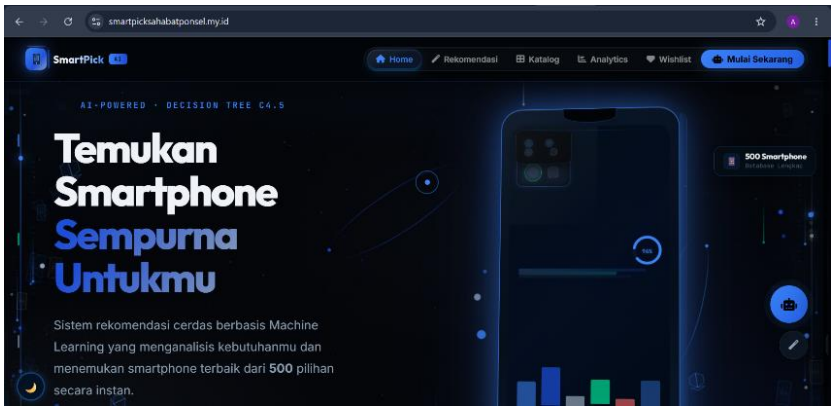

Figure 8. SmartPick Homepage Dark Theme

.

3.8 Functional Testing

Black-box testing covering the AI quiz, C4.5 classification, catalog browsing and filtering, comparison, wishlist, rating and review, the analytics dashboard, and contact information across 10 scenarios returned a 100% pass rate, confirming the prototype's core functionality operates as intended alongside the classification results reported above.

3.9 Discussion

SmartPick addresses gaps in prior work. Amperawansyah & Putri [5] achieved 95% accuracy but covered only low-budget segments. Ramadhan et al. [6] reported 81% accuracy without feature engineering. Zuhdiansyah & Luthfiarta [2] applied K-Means + SVD collaborative filtering, which is vulnerable to cold-start problems when new devices are added, a critical limitation for a store that regularly updates its catalog. SmartPick integrates multi-variable classification (14 engineered features) into a single interpretable C4.5 Decision Tree, achieving 96.00% testing accuracy on a balanced 500-device dataset, and Section 2.4 outlines why C4.5 was preferred over ensemble and distance-based alternatives for this task. Compared with CART and Random Forest, C4.5 trades a small amount of raw testing accuracy (96.00% vs. 94.67% for both) for a single, fully traceable decision path rather than an ensemble of trees; Random Forest in particular reached 100.00% training accuracy against 98.00% for C4.5, indicating a larger train-test gap and a higher risk of overfitting to this dataset's specific feature distribution despite similar testing performance. Naive Bayes and K-NN trail by roughly 25-30 percentage points, reflecting that their independence and distance assumptions do not fit the mixed, correlated feature set (e.g., price, RAM, and camera specifications are jointly informative rather than independent). In terms of model complexity, C4.5's depth-5, 15-leaf tree is substantially simpler than a Random Forest ensemble while remaining close in accuracy, which matters directly for retail deployment: staff can explain a recommendation as a short sequence of threshold checks rather than pointing to an opaque ensemble vote.

Stability across the 5-Fold Cross-Validation runs (94.82% ±1.76%) is comparable to CART (96.24% ±1.15%) and Random Forest (95.76% ±1.20%), suggesting the performance gap between tree-based methods is small and within the range of ordinary sampling variation on a dataset of this size, rather than evidence that one tree-based method is categorically superior. This reinforces that C4.5's advantage over CART and Random Forest here lies less in raw accuracy and more in interpretability and reduced model complexity, while its advantage over Naive Bayes and K-NN is substantial on both accuracy and stability grounds. The approach is best suited to retail contexts where a small, curated catalog and explainable rules are valued over marginal accuracy gains from ensemble methods for example, single-store or small-chain electronics retailers, where staff turnover makes an auditable rule set more sustainable than a black-box model. It is less suited to marketplaces with thousands of frequently changing SKUs, where the manual labeling effort behind the four usage categories would not scale, and where the modest dataset size (discussed in Section 2.2) would need to be addressed before deployment at that scale.

4. CONCLUSION

This study successfully developed a C4.5 Decision Tree model for classifying smartphone user needs into four categories Gaming, Photography/Videography, Student, and General User achieving 98.00% training, 97.33% validation, and 96.00% testing accuracy, with 94.82% (±1.76%) on 5-Fold Cross-Validation, indicating a well-generalized model without significant overfitting. The resulting decision tree (depth 5, 29 nodes, 15 leaf nodes) uses Price (IDR) as the primary splitting attribute with the highest Gain Ratio of 0.8460, producing transparent and interpretable classification rules. C4.5 outperformed all four comparative algorithms in testing accuracy (96.00% vs. 94.67% for CART and Random Forest, 69.33% for K-NN, and 66.67% for Naive Bayes), while maintaining single-tree interpretability superior to ensemble methods. The model was implemented as a web-based prototype, SmartPick, at Sahabat Ponsel Tembilahan retail store to validate its real-world feasibility. These results should be interpreted with the dataset's scale in mind: the 500 sample dataset, while balanced and cross-validated, remains modest relative to the full diversity of commercially available smartphones. Future work should therefore prioritize evaluating the model on progressively larger datasets (e.g., 1,000 and 2,000 samples) to characterize scalability and overfitting risk as sample size grows, alongside periodic model retraining using user rating data, hybrid C4.5 and collaborative filtering for personalized recommendations, user authentication with history tracking, and replication at other retail stores to validate generalizability of the proposed approach.

REFERENCES

[1] A. Ernawati and S. Wahyuni, "Analisis Data Mining Pola Penggunaan Seluler dan Klasifikasi Perilaku Pengguna Menggunakan Metode C4.5," Bulletin of Information Technology, vol. 5, no. 4, pp. 262–268, 2024, doi: 10.47065/bit.v5i2.1689.

[2] I. Zuhdiansyah and A. Luthfiarta, "Sistem Rekomendasi Pembelian Smartphone berbasis Algoritma K-Means dan SVD," Jurnal Nasional Teknologi dan Sistem Informasi, vol. 10, no. 1, pp. 45–53, 2024, doi: 10.25077/teknosi.v10i1.2024.45-53.

[3] A. Fitriansyah, N. Sucahyo, and A. E. Verawati, "Sistem Rekomendasi Pemilihan Smartphone Dengan Case Based Reasoning," Jurnal Teknologi Informatika dan Komputer, vol. 8, no. 2, pp. 1–16, 2022, doi: 10.37012/jtik.v8i2.1134.

[4] D. A. Mukhsinin et al., "Implementasi Algoritma Decision Tree untuk Rekomendasi Film pada Netflix," MALCOM, vol. 4, no. 2, pp. 570–579, 2024, doi: 10.57152/malcom.v4i2.1255.

[5] R. Amperawansyah and D. P. Putri, "Pemilihan Jenis Ponsel Pintar Sesuai dengan Kebutuhan Menggunakan Metode Forward Chaining dan Decision Tree," Jurnal Inovtek Polbeng - Seri Informatika, vol. 7, no. 1, pp. 1–8, 2022.

[6] A. T. Ramadhan et al., "Penerapan Algoritma Decision Tree Dalam Klasifikasi Harga Handphone," Jurnal Sistem Informasi dan Ilmu Komputer, vol. 1, no. 4, pp. 195–206, 2023, doi: 10.59581/jusiik-widyakarya.v1i4.1861.

[7] D. F. S. Septiana, W. A. Triyanto, and S. Muzid, "Penerapan Metode Decision Tree Algoritma C4.5 Dalam Sistem Rekomendasi Jurusan Bagi Calon Mahasiswa Baru," Jurnal Unitek, vol. 18, no. 1, pp. 167–178, 2025, doi: 10.52072/unitek.v18i1.1322.

[8] A. Pariddudin and F. S. Warsa, "Penerapan Algoritma C4.5 Untuk Rekomendasi Mentor Santri Baru," Teknois, vol. 13, no. 1, pp. 44–49, 2023, doi: 10.36350/jbs.v13i1.

[9] R. Musfikar, H. Apriadinata, and B. Yusuf, "Aplikasi Prediksi Prestasi pada Siswa Menggunakan Algoritma C4.5," JAMIKA, vol. 13, no. 2, pp. 148–162, 2023, doi: 10.34010/jamika.v13i2.10649.

[10] D. Nasution et al., "Pengujian Algoritma C4.5 Untuk Mengevaluasi Kinerja Pegawai," SURPLUS, vol. 2, no. 2, pp. 412–424, 2024.

[11] A. D. Kuswanto et al., "Penerapan Algoritma C4.5 dalam Klasifikasi Prestasi Atlet," Jurnal Informatika dan Sains Teknologi (MODEM), vol. 1, no. 3, pp. 52–61, 2024, doi: 10.62951/modem.v1i3.110.

[12] L. N. Ningsih, R. Septiani, A. Pramadjaya, and S. Nuralisah, "Implementing the C4.5 Algorithm for Customer Satisfaction Classification," MALCOM, vol. 5, no. 4, pp. 1470–1480, 2025, doi: 10.57152/malcom.v5i4.2211.

[13] Amrin, O. Pahlevi, and H. Rianto, "Analisa Komparasi Model Data Mining Algoritma C4.5, CHAID, dan Random Forest Untuk Penilaian Kelayakan Kredit," Computer Science (CO-SCIENCE), vol. 5, no. 1, pp. 49–57, 2025, doi: 10.31294/coscience.v5i1.6208.

[14] P. A. Firnanda, L. Shofwatillah, F. Rahma, and F. Fauzi, "Analisis Perbandingan Decision Tree dan Random Forest dalam Klasifikasi Penjualan Produk pada Supermarket," Emerging Statistics and Data Science Journal, vol. 3, no. 1, pp. 445–461, 2025, doi: 10.20885/esds.vol3.iss.1.art2.

[15] N. L. Desmita, S. Lonang, and D. T. Kumoro, "Comparative Analysis of Decision Tree and Random Forest for Predicting Diabetes Mellitus," Journal of Information Systems and Informatics, vol. 7, no. 1, pp. 1–12, 2025.

[16] K. Handayani, L. Lisnawanty, A. Latif, M. R. Firdaus, and F. N. Hasan, "Komparasi Algoritma C4.5 dan Naïve Bayes dalam Penentuan Status Kelayakan Donor Darah," Sistemasi: Jurnal Sistem Informasi, vol. 10, no. 3, pp. 676–687, 2021, doi: 10.32520/stmsi.v10i3.1440.

[17] H. Kadar and A. Budiyantara, "Penentuan Kelayakan Karyawan Baru Menggunakan Data Mining C4.5," Merkurius: Jurnal Riset Sistem Informasi dan Teknik Informatika, vol. 2, no. 6, pp. 29–41, 2024, doi: 10.61132/merkurius.v2i6.389.

[18] N. D. Saputri, K. Khalid, and D. Rolliawati, "Komparasi Penerapan Metode Bagging dan Adaboost pada Algoritma C4.5 untuk Prediksi Penyakit Stroke," Sistemasi: Jurnal Sistem Informasi, vol. 11, no. 3, pp. 567–577, 2022, doi: 10.32520/stmsi.v11i3.1684.

.