Journal of Information Technology and Cyber Security 1(1) January 2023: 9-15 EISSN 2987-386X | PISSN 2987-3878
DOI: 10.30996/jitcs.7559
Research Article

Recommendation System using the K-Nearest Neighbor Approach: A Case Study of Dual Camera Quality as a Smartphone Selection Criterion

Parcelliana Binar Pasha [1,*] and Yusrida Muflihah [2]

[1] Department of Informatics Engineering, Universitas 17 Agustus 1945 Surabaya, Indonesia
[2] Department of Information Systems and Technology, Universitas 17 Agustus 1945 Surabaya, Indonesia
* Corresponding author: parcellianabinar20@gmail.com

Received: 16 November 2022
Revised: 20 December 2022
Accepted: 20 December 2022
Available online: 31 December 2022

To cite this article: Pasha, P. B., & Muflihah, Y. (2023). Recommendation System using the K-Nearest Neighbor Approach: A Case Study of Dual Camera Quality as a Smartphone Selection Criterion. *Journal of Information Technology and Cyber Security, 1*(1), 9–15. https://doi.org/10.30996/jitcs.7559

Abstract

Many smartphones today need to be more precise about choosing one that suits the user's needs. In fact, smartphone sellers sometimes need help recommending smartphones that suit buyers' needs. Generally, buyers search for smartphone specifications with keywords they desire, but the results appear different from what they expected. Users need the main specifications, such as Random Access Memory (RAM) and Read Only Memory (ROM) capacity, battery, and high camera quality. This research aims to implement the K-Nearest Neighbor (KNN) algorithm for recommendation smartphone selection based on the criteria mentioned. The data test results show that the combination of KNN with four criteria has good performance, as indicated by the accuracy, precision, recall, and f-measure values of 95%, 94%, 97%, and 95%, respectively.

Keywords: Euclidean distance, K-Nearest Neighbor, recommendation system, smartphone, specifications.

1. Introduction

Today, smartphones are an essential peripheral. Based on the 2017 Ministry of Communication and Informatics (Menkominfo) survey results, 66.3% of Indonesian people already have smartphones, and 86.60% of smartphone owners come from Java (Kemkominfo, 2017). However, many different types and functions often confuse buyers about choosing a smartphone based on their needs. Buyers often need help selecting their desired items (Bangun, 2017).

Generally, buyers will search for information that fits their needs only with keywords. Searching with these keywords will be understood by other words with the same meaning so that the information that appears not only contains words that are limited to the phrase being searched but also raises information about the equivalent word (Setiawan & Nurkamid, 2012). Similar studies have been conducted on smartphone recommendation systems using the Simple Additive Weighting (SAW) method (Harsiti & Aprianti, 2017; Saputra et al., 2021). Saputra et al. (2021) use rating as a criterion to aid decision-making. Putra (2019) in his research used the K-Nearest Neighbor (KNN) method for smartphone recommendations with the criteria used in the form of Read Only Memory (ROM), Random Access Memory (RAM), smartphone dimensions, rear camera quality, battery capacity, and price. In contrast, Setiaji et al. (2022) used ratings from user reviews in their research.

This study determines RAM capacity, battery, ROM, and camera quality characteristics. Sometimes, the result from the review does not match with user's requirement that a smartphone's capacity fits their needs. In addition, the price of smartphones is highest than the specifications of smartphones and can affect the results of recommendations based on specifications. KNN is an algorithm often used to assist decision-making in recommender systems, such as recommendation systems for choosing cars or movies (Zhang et al., 2017). KNN applies the principle of data classification using the similarity or proximity of the search data to the data in the system (Zhang et al., 2018).

This research aims to study the performance of the KNN algorithm as a smartphone selection recommendation system based on five criteria. These criteria include the capacity of RAM, battery, ROM, and camera quality according to user needs.

2. Literature Review

Sari dan Saputra (2021) dalam penelitiannya membuat sebuah sistem pemilihan *smartphone* berdasarkan spesifikasi untuk mahasiswa dengan metode SAW. Pada penelitian Sari dan Saputra (2021), SAW digunakan sebagai sistem pembobotan untuk menentukan prioritas kriteria pemilihan *smartphone*. Dalam penelitian tersebut kriteria yang digunakan dalam pemilihan *smartphone* adalah jenis *chipset*, kapasitas RAM dan ROM, ukuran layar serta harga *smartphone*. Data yang digunakan merupakan data dari pengisian Google Form oleh mahasiswa program studi Sistem Informasi Universitas Tanjungpura, Indonesia. Hasil survei yang di dapat yaitu Oppo A92 dengan nilai 96; Asus ROG Phone 5 dengan nilai 80; Samsung S21 5G dengan nilai 73,75; Vivo Y17 dengan nilai 81,25; dan Iphone 11 64GB dengan nilai 48,75.

In their research, Sari and Saputra (2021) created a smartphone selection system based on specifications for students using the SAW method. Sari and Saputra's research (2021) uses SAW as a weighting system to determine priority smartphone selection criteria. In this study, the criteria for selecting a smartphone were the type of chipset, RAM and ROM capacity, screen size, and smartphone price. The data is from filling out the Google Form by students of the Information Systems study program at the University of Tanjungpura, Indonesia. The survey results obtained were Oppo A92 with a value of 96; Asus ROG Phone 5 with an 80 rating; Samsung S21 5G with a score of 73.75; Vivo Y17 with a score of 81.25; and iPhone 11 64GB with a score of 48.75.

Fauzan et al. (2017) built a system to support web-based smartphone selection decisions using the SAW method. The criteria in this study consist of the processor core, processor clock, RAM, ROM, camera, battery, and price. The study results show that the information display can be used to search for smartphone recommendation systems.

Gafoor et al. (2022) proposed a film recommendation system developed using one of the most powerful, well-known, and widely used KNN machine learning algorithms to improve the prediction of the likelihood of specific digital content to users whose likelihood was previously analyzed.

Rajput & Grover (2022) predict film genres as having interesting problems in designing recommendation systems for audiences, analyzing film box office performance, and understanding film themes. This is a classic multi-label classification problem. The algorithm for detecting film genres in this study is KNN. The basic idea is to identify high-frequency words in a given genre and use them as features to train a classification machine learning model. The best results are obtained using KNN with an average precision for all genres of 77.7% with 200 features. KNN works excellent for Sports and War genres with over 90% precision in some cases.

In their research, Xiong & Yao (2021) proposed that a KNN-based thermal comfort model be developed to form an adaptive thermal comfort environment that is personalized to suit the occupants' preferences. The test results show that the accuracy of the KNN model with 1,000 sets of training data can reach 88.31%.

Rakshit et al. (2023) stated that the way the site knows about products recommended to new users is that the best-selling products of the e-commerce site are products that are in high demand. Therefore, Rakshit et al. (2023) proposed a popularity-based recommendation system using KNN. The result is the top 5 popular products recommended for users who use the KNN algorithm.

Adeniyi et al. (2016) in their research, implemented KNN on an automated web and recommendation system based on current user behavior on a newly developed Simple Syndication (RSS) reader website to provide relevant information to individuals without explicitly asking for it. The test results show that KNN is transparent, consistent, straightforward, easy to understand, has a high tendency to have the desired quality, and is easy to implement.

3. Methods

The method section will explain the stages of data collection and normalization and find recommendations based on the KNN algorithm. Fig. 1 is a flow chart of the research conducted. In this study, the value of k is needed to determine the number of nearest neighbors used to determine the classification. The value of k used in this research is $k = 5$, $k = 10$, and $k = 15$. This value will be used to compare the accuracy, precision, and recall values obtained from the results of this study.

3.1. Dataset

The dataset used in this study consists of 100 smartphone specification data found at the Birawa Cell counter located in Sidoarjo, East Java, Indonesia. The comparison between training data and test data is 80%:20%. The parameters used in the dataset are RAM capacity, battery, ROM, and quality of the main camera and front camera.

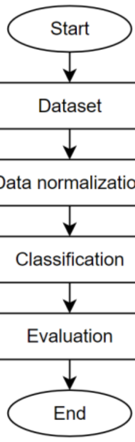

Fig. 1. Research flow chart.

Table 1
The original data.

No	Item	RAM	ROM	Battery	Main Camera	Front Camera
1	OPPO A76	6	128	5000	8	13
2	Redmi 9A	2	32	5000	5	13
3	VIVO Y01	2	32	5000	5	8
4	OPPO RENO7 5G	8	256	4500	32	64
5	Oppo Reno6	8	128	4310	44	64
6	VIVO V21 5G	8	256	4000	44	64
7	VIVO V23 5G	8	128	4200	50	64
8	OPPO A74 5G	6	128	5000	16	48
9	REALME C25Y	4	64	5000	8	50
10	REALME C31	3	32	5000	5	13

3.2. Data normalization

The process carried out to prepare data before classification is the process of normalization. Normalization is done to equalize the shape of the data that is not uniform and make the data have a range from 0 to 1 (Zhu et al., 2023). Normalization is also interpreted as changing data into Gaussian data or equivalent value equations (Peterson, 2021). This study uses min-max normalization. In this process, several steps need to be taken. Min-max normalization can be done by Eq. (1),

$$norm = \frac{v - min_k}{max_k - min_k} \quad (1)$$

where $norm$ is normalization, v is the original value, min_k is the minimum value in column k, and max_k is the maximum value in column k. The normalization process is used for each attribute used in the research. Table 1 is an example of data before it is normalized. Table 2 is an example of normalized sample data.

Table 2
Data after normalization.

No	Item	RAM	ROM	Battery	Main Camera	Front Camera
1	OPPO A76	0.6667	0.4286	1	0.0667	0.0893
2	Redmi 9A	0	0	1	0	0.0893
3	VIVO Y01	0	0	1	0	0
4	OPPO RENO7 5G	1	1	0.5	0.6	1
5	Oppo Reno6	1	0.4286	0.31	0.8667	1
6	VIVO V21 5G	1	1	0	0.8667	1
7	VIVO V23 5G	1	0.4286	0.2	1	1
8	OPPO A74 5G	0.6667	0.4286	1	0.2444	0.7143
9	REALME C25Y	0.3333	0.1429	1	0.0667	0.75
10	REALME C31	0.1667	0	1	0	0.0893

This study uses five criteria, as presented in Table 2, namely:

1) RAM represents RAM capacity,
2) ROM represents ROM capacity,

3) Battery represents battery capacity,
4) Main Camera represents the quality of the main camera, and
5) Front Camera represents the quality of the front camera.

3.3. Classification

The most important step in classification is determining the best classifier (Kowsari et al., 2019). This study uses one of the simplest classification methods, namely KNN. KNN is sensitive to the distance function used to select the nearest neighbors (Suyanto et al., 2022). Eq. (2) is the Euclidean distance equation used in KNN,

$$d_{euclid}(x,y) = \sqrt{\sum_{p=0}^{q}\left(x_p - y_p\right)^2} \tag{2}$$

where $d_{euclid}(x,y)$ as Euclidean distance, x_p as a value test iteration-p, y_p as a value data training iteration-p, p as variable data, and q as parameters quantity.

Table 1 must be normalized first, and then the normalization results will be used to find Euclidean values. Search for Euclidean values according to Eq. (2). After the normalization process is complete, it will find the Euclidean value. Table 3 is the result of calculating the Euclidean value based on training data with RAM specifications of 6, ROM of 64, battery capacity of 5000, front camera with 44 MP, and a main camera with 64 MP.

Table 3
The result of euclidean value.

No	Item	*Euclidean Value*	*Rank*	Classifier
1	OPPO A76	1.2454	6	Medium
2	Redmi 9A	1.4302	9	Low
3	VIVO Y01	1.4886	10	Low
4	OPPO RENO7 5G	1.0802	5	High
5	Oppo Reno6	0.8178	2	Medium
6	VIVO V21 5G	1.3586	7	High
7	VIVO V23 5G	0.9222	4	Medium
8	OPPO A74 5G	0.7419	1	Medium
9	REALME C25Y	0.9020	3	Low
10	REALME C31	1.3605	8	Low

Based on the results from Table 3, the data in Table 4 will be classified based on the nearest neighbor. In this sample, the value of $k = 5$ is used. Then the results of the nearest neighbor search can be seen in Table 6.

Based on Table 4, the classification results are as follows:
1) The closest neighbor with the top classification is 1;
2) Closest neighbors with low classification are 1;
3) Meanwhile, the nearest neighbors with medium classification have the highest number, namely 3.

So the results of the classification of smartphones in the test data of Table 3 are middle-class smartphones.

Table 4
Nearest neighbor.

Item	Euclid Value	Rank	Classifier
OPPO A74 5G	0.7419	1	Medium
Oppo Reno6	0.8178	2	Medium
REALME C25Y	0.9020	3	Low
VIVO V23 5G	0.9222	4	Medium
OPPO RENO7 5G	1.0802	5	High

3.4. Evaluation

Evaluation in this study aims to see the performance of KNN. The evaluation used in this study is the confusion matrix. The confusion matrix is a way to evaluate and review the accuracy and identify errors in calculations made (Thammasiri et al., 2014).

Table 5
Confusion matrix.

		Recent Values	
		Positive	Negative
Prediction Value	Positive	False Positive (FP)	False Positive (FP)
	Negative	True Negative (TN)	True Negative (TN)

The confusion matrix is searched by comparing the output predicted by the system with the elaboration output done manually. Eq. (3), (4), (5), and (6) are the equations used in the confusion matrix in this study (Lee et al., 2022; Thammasiri et al., 2014),

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

$$F - Measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{6}$$

with TP as a True Positive value, TN as a True Negative value, FP as a False Positive value, and FN as a False Negative value, accuracy is the result of the accuracy, recall is recall, precision is precision and F-Measure.

4. Results and Discussion

In this study, two trial scenarios were carried out. The first test scenario is based on the number of criteria used, as presented in Table 6. The number of criteria 1 means that the test is carried out with only one type of criterion, such as a search using only RAM. The number of criteria 2 means that the test is carried out using RAM, ROM, etc. This test is continued until all five criteria are used. The aim is to determine whether the number of criteria used affects the performance results of the KNN. While the second trial scenario is based on the number of parameter k values, as presented in Table 7.

Table 6
Test result criteria's.

Criteria	Accuracy	Precision	Recall	F-Measure
1	0.49	0.61	0.63	0.62
2	0.74	0.79	0.74	0.76
3	0.84	0.79	0.84	0.81
4	0.84	0.85	0.85	0.85
5	**0.95**	**0.94**	**0.97**	**0.95**

Table 7
Test result based on k value.

k	Accuracy	Precision	Recall	F-Measure
5	0.80	0.87	0.87	0.87
10	0.85	0.90	0.89	0.89
15	0.95	0.94	0.97	0.95

Based on Table 7, $k = 15$ gets the highest performance value. Each performance value is accuracy with a value of 0.95; precision with a value of 0.94; recall with a value of 0.97; and the F-Measure value is 0.95. Therefore, the value of k used in this study is 15. Based on Table 6, it can be concluded that the number of criteria affects the accuracy of the KNN performance. The more criteria used, the higher the resulting accuracy.

Based on the testing results, this study was declared successful because the proposal had an accuracy value of 95%, which was considered good. However, according to Dio et al. (2021), fairly good accuracy is 80% to 110%. This means that the method is able to provide smartphone recommendations that are close to user needs. Factors that influence the success of the research are based on the amount of data used and the number of criteria used or the absence of missing values. As found in testing with a number of criteria. The higher the number of criteria used, the higher the accuracy.

5. Conclusions

This research aims to look at the performance of the KNN algorithm as a smartphone selection recommendation system based on five criteria. Based on the test results, it can be concluded that this study was successful with good performance when using a value of $k = 15$ and all the proposed criteria. This research still has limitations, so it is recommended that future research be continued by adding other criteria, other types of data that have the potential to be needed by prospective buyers, better algorithms than KNN, and cross-validation.

6. CRediT Authorship Contribution Statement

Parcelliana Binar Pasha: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, and Project administration. **Yusrida Muflihah:** Validation and Visualization.

7. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

8. Acknowledgments

The authors would like to thank the anonymous referees for their helpful comments and suggestions.

9. Data Availability

Data will be made available on request.

10. Funding

No funding was received for this study.

11. Ethical Approval

Ethical approval No patient-identifying parts in this paper were used or known to the authors. Therefore, no ethical approval was requested.

12. References

Adeniyi, D. A., Wei, Z., & Yongquan, Y. (2016). Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method. *Applied Computing and Informatics*, *12*(1), 90–108. https://doi.org/10.1016/J.ACI.2014.10.001

Bangun, A. A. A. (2017). *Pengaruh Brand Ambassador Terhadap Brand Image Serta Dampaknya Terhadap Keputusan Pembelian (Studi pada Pengguna Smartphone Oppo di Kota Medan)*. Universitas Sumatera Utara. https://repositori.usu.ac.id/handle/123456789/17167

Dio, R. G. R., Bahri, S., Kiswandono, A. A., & Supriyanto, R. (2021). Validasi Metode Fotodegradasi Congo Red Terkatalis ZNO/Zeolit Y Secara Spektrofotometri UV-VIS. *Analit: Analytical and Environmental Chemistry*, *6*(2), 134–144. https://doi.org/10.23960/aec.v6.i2.2021.p134-144

Fauzan, R., Saberan, S., & Ridwan, M. (2017). Sistem Pendukung Keputusan Pemilihan Smartphone Menggunakan Metode Simple Additive Weighting (SAW). *Prosiding SNRT (Seminar Nasional Riset Terapan)*.

Gafoor, A., Srujana, A. L., Nagasri, A., Durgaprasad, G. S. S., & Dasari, L. S. K. (2022). KNN based Entertainment Enhancing System. *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, 1056–1061. https://doi.org/10.1109/ICOEI53556.2022.9777225

Harsiti, H., & Aprianti, H. (2017). Sistem Pendukung Keputusan Pemilihan Smartphone dengan Menerapkan Metode Simple Additive Weighting (SAW). *JSiI (Jurnal Sistem Informasi)*, *4*, 19–24. https://doi.org/10.30656/JSII.V4I0.372

Kemkominfo, B. (2017). *Survey Penggunaan TIK 2017*. https://literasidigital.id/buku/797-2

Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text Classification Algorithms: A Survey. *Information*, *10*(4). https://doi.org/10.3390/info10040150

Lee, G. T., Nam, H., Kim, S.-H., Choi, S.-M., Kim, Y., & Park, Y.-H. (2022). Deep learning based cough detection camera using enhanced features. *Expert Systems with Applications*, *206*, 117811. https://doi.org/10.1016/j.eswa.2022.117811

Peterson, R. A. (2021). Finding Optimal Normalizing Transformations via bestNormalize. *The R Journal*, *13*(1), 310–329.

Putra, A. E. (2019). *Sistem Rekomendasi Pemilihan Handphone Menggunakan K-Nearest Neighbor (KNN)*.

Rajput, N. K., & Grover, B. A. (2022). A multi-label movie genre classification scheme based on the movie's subtitles. *Multimedia Tools and Applications*, *81*(22), 32469–32490. https://doi.org/10.1007/s11042-022-12961-6

Rakshit, P., Saha, S., Chatterjee, A., Mistri, S., Das, S., & Dhar, G. (2023). A Popularity-Based Recommendation System Using Machine Learning. *Machine Learning in Information and Communication Technology*, *498*, 143–150. https://doi.org/10.1007/978-981-19-5090-2_14

Saputra, B. D., Subagja, M. H., Aldiansyah, M., Setiawan, W., & Rosyani, P. (2021). Sistem Pendukung Keputusan Pemilihan Smartphone dengan Metode Simple Additive Weighting (SAW) | Scientia Sacra: Jurnal Sains, Teknologi dan Masyarakat. *Scientia Sacra*, *1*(3). http://www.pijarpemikiran.com/index.php/Scientia/article/view/67

Sari, R. P., & Saputra, B. (2021). Sistem Pemilihan Smartphone Berdasarkan Spesifikasinya Pada Mahasiswa Sistem Informasi Universitas Tanjungpura Menggunakan Metode Simple Additive Weighting (SAW). *Jurnal Sistem Komputer Dan Informatika (JSON)*, *2*(3), 329–338. https://doi.org/10.30865/json.v2i3.3038

Setiaji, B. R., Utama, D. Q., & Adiwijaya, A. (2022). Smartphone Purchase Recommendation System Using the K-Nearest Neighbor (KNN) Algorithm. *Jurnal Media Informatika Budidarma*, *6*(4), 2180–2186. https://doi.org/10.30865/MIB.V6I4.4753

Setiawan, R. R., & Nurkamid, M. (2012). Teknologi Web Semantik Untuk Bibliografi Perpustakaan. *Semantik*, *2*(1). http://publikasi.dinus.ac.id/index.php/semantik/article/view/60

Suyanto, S., Yunanto, P. E., Wahyuningrum, T., & Khomsah, S. (2022). A multi-voter multi-commission nearest neighbor classifier. *Journal of King Saud University - Computer and Information Sciences*, *34*(8), 6292–6302. https://doi.org/10.1016/J.JKSUCI.2022.01.018

Thammasiri, D., Delen, D., Meesad, P., & Kasap, N. (2014). A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. *Expert Systems with Applications*, *41*(2), 321–330. https://doi.org/10.1016/j.eswa.2013.07.046

Xiong, L., & Yao, Y. (2021). Study on an adaptive thermal comfort model with K-nearest-neighbors (KNN) algorithm. *Building and Environment*, *202*, 108026. https://doi.org/10.1016/J.BUILDENV.2021.108026

Zhang, S., Li, X., Zong, M., Zhu, X., & Cheng, D. (2017). Learning k for kNN Classification. *ACM Transactions on Intelligent Systems and Technology*, *8*(3). https://doi.org/10.1145/2990508

Zhang, S., Li, X., Zong, M., Zhu, X., & Wang, R. (2018). Efficient kNN classification with different numbers of nearest neighbors. *IEEE Transactions on Neural Networks and Learning Systems*, *29*(5), 1774–1785. https://doi.org/10.1109/TNNLS.2017.2673241

Zhu, X., Ma, B., Su, G., Hu, Y., & Liu, D. (2023). Blank design method for C-section profile ring rolling based on FEM and RSM. *Alexandria Engineering Journal*, *65*, 649–660. https://doi.org/10.1016/J.AEJ.2022.10.007