Smartphone Purchase Recommendation System Using the K-Nearest Neighbor (KNN) Algorithm

Bayu Rahmat Setiaji*, Dody Qori Utama, Adiwijaya

Informatics, Informatics, Telkom University, Bandung, Indonesia
Email: [1,*]Bayurahmats@student.telkomuniversity.ac.id, [2]dodyqori@telkomuniversity.ac.id,
[3]Adiwijaya@telkomuniversity.ac.id
Correspondence Author Email: Bayurahmats@student.telkomuniversity.ac.id

Abstract−Indonesia is in the fourth position of the countries with the most smartphone users worldwide. Smartphones are needed in today's modern times. Smartphones are also used not only for long-distance communication but also for carrying out daily work. Smartphones are currently used for study and work and also become entertainment to play. Therefore, smartphones are very much sought after for the suitability of users who carry out their daily activities. So this research is very helpful for users to find smartphones that support their daily activities such as studying, working, and playing. This research is based on a website that can make it easier for users to see their smartphone recommendations directly. The analysis uses the K-Nearest Neighbor (KKN) method to see the ratings reviewed by other users who have tried using their smartphones with different phone brands. The calculation method in the current study uses 3 KNN calculations and uses the concept of combining calculations to find the maximum recommendation results. The result of the recommendation system using the K-Nearest Neighbor method is in the form of a review stating whether the user agrees or disagrees. In the current study, there have been 100 reviews from users, and it has a percentage of 78% for success and 22% for failure.

Keywords: Smartphone; Website; Flask; Python; K-Nearest Neighbor (KNN)

1. INTRODUCTION

In the modern era, smartphones have gone through many developments every year. Adults currently use smartphones to elementary school children are already using smartphones. Currently, people are very active in using smartphones which are not just a communication tool but also for their work and education needs. Based on statistical data from the newzoo web, Indonesia is a country that has a fairly large population, which is around 276.5 million Indonesians. 64% of Indonesia's population are active smartphone users and occupy the fourth position in the country, which is active smartphone users, which is around 178.9 million Indonesians [1].

The needs and lifestyles of the people can also determine a person's choice of their smartphone. Because the orientation to consumer satisfaction that the existing phenomenon cannot be separated from a misunderstood understanding, especially related to smartphone demand [2]. Not all smartphone owners meet the status symbol criteria because many are just stylish with smartphones [2]. With that, the function of the smartphone as a smartphone is not utilized optimally by consumers. Then came the idea of mobile phone manufacturers to fulfill all consumers' needs and lifestyles. The existence of needs and lifestyles among the community is also why manufacturers replace new smartphones with similar specifications and prices. So we need a system that helps consumers choose a smartphone according to their needs or lifestyle.

The Recommendation System has several methods and algorithms that can be applied. Performance testing of the application of the algorithm on the Recommendation System can measure the algorithm's success in providing recommendations to users [3]. The recommendation system has been widely used today to assist the community in making decisions. Several studies are related to the recommendation system; for example, The news article recommendation system uses the k nearest neighbor method based on the website [4]. This recommendation system provides good information for users in finding news articles that users want to read or news articles that are common to users. Applying the K-Nearest Neighbor Method for the Car Selection Recommendation System [5] in this study is helpful for user recommendations to choose a car by determining parameters such as price, year, passenger capacity, color, and transmission type.

Furthermore, research on the performance of the K-Nearest Neighbors User Algorithm on the Recommendation System at Tokopedia [3] in this study uses an item rating prediction based on the assessment given by the buyer. As for the research on the implementation of the K-Nearest Neighbor Algorithm on the Laptop Recommendation Website [6], in this study, the user is immensely helped by the existence of a website that makes it easy for users to view specification data or images from laptops recommended on the website. So that the recommendation system using the website is quite easy to provide recommendation information to users. In subsequent research on Recommendation Systems in E-Commerce Using K-Nearest Neighbor [7], this study only compares recommendation systems that use either KNN or Collaborative Filtering. Another study that discussed the recommendation system using the collaborative filtering method and content-based filtering was quite helpful in providing smartphone recommendation results by taking the advantages of each method [8]. However, a Hybrid approach method is used to overcome the weaknesses of the collaborative and content-based filtering methods [9]. From the KNN method, there are shortcomings in dynamically classifying item weights and the best k value rather than manually [10]. In this study, Dynamic KNN and WAKNN are classification search methods that are good enough to determine the results by studying the best k value or using the attribute weighting method [10].

A recommendation system is a system that can help and guide users to find suitable products based on their preferences [11]. There is also a DSS (Decision Support System) application to support the solution of a particular problem or to evaluate an opportunity [12]. This research has a positive impact on customers or subsequent writers because this system makes it easier for consumers to buy smartphones, by using the rating as a parameter. Ratings are very simple, easy to understand, and provide sufficient data to give good decision results [13]. The recommendation system in this study uses the KNN (K-Nearest Neighbor) method which uses rating data as training data, and the implementation of the KNN method on the website uses the python programming language.

2. RESEARCH METHODOLOGY

2.1 Research Steps

Research that has been done in recent years can help to structure the current research. There are many studies on smartphone purchase recommendation systems using collaborative filtering, the implementation of the K-Nearest Neighbor algorithm on laptop recommendation websites, and the classification of snakebite marks using active contour models and k-nearest neighbor. Recommendation systems are generally designed in two stages, 1) Building a user model and 2) Making recommendations [14]. In general, the methods in recommender systems are classified into four main approaches: 1) content-based 2) collaborative filtering 3) knowledge-based 4) hybrid recommendation systems [14] This chapter will focus on recommendation systems, and websites using flask, python, and algorithm systems. K-Nearest Neighbor. The following is presented in Figure 1, which are the stages in the current research to get the results of recommendations to be given to customers.
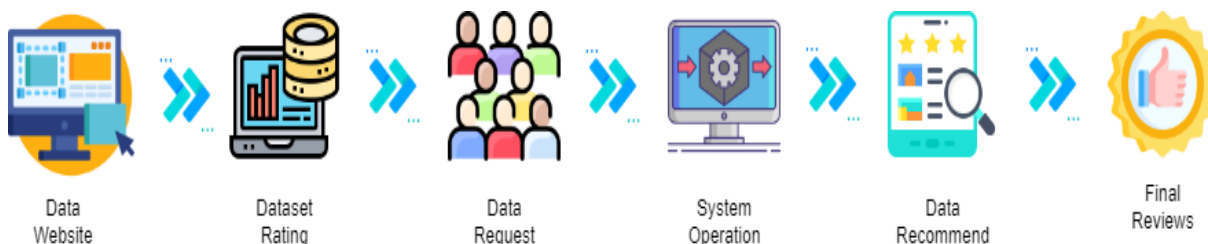

Figure 1. Research Steps

2.2. Python Programming Language

Python is a multipurpose interpretive programming language with a design philosophy focusing on code readability. Python is claimed to be a language that combines capabilities, capabilities with obvious code syntax and is equipped with an extensive and comprehensive standard library functionality. Python is also supported by a large community [15].

2.3. Flask Website

Flask is a web framework written in Python and classified as a type of micro-framework. Flask works as an application framework and the appearance of a web. By using Flask and Python, developers can create a structured web and manage the web behavior more efficiently [16].

Most common functions and components such as form validation, databases, and more are not installed by default in Flask. This is because third parties have provided these functions and components, and Flask can use extensions that make these features and components seem to be implemented by Flask itself. In addition, even though Flask is referred to as a micro-framework, it does not mean that it lacks functionality [17]. The microframework here means that Flask intends to make the core of this application as simple as possible but still easily addable. That way, Flask's flexibility and scalability can be pretty high compared to other frameworks [17]. Flask has advantages and disadvantages compared to other Python-based web frameworks, such as Django, CherryPy, and others [18].

2.4. K-Nearest Neighbor

The K-Nearest Neighbor (KNN) algorithm is a method used to classify new objects based on attributes and training sample data [6]. These algorithms often produce competitive results and have significant advantages over some other data mining methods [4]. The working principle of KNN is to find the shortest distance between the data used and the K closest neighbors in the training data [6]. The KNN method is used because it has several advantages, including being able to produce more accurate and effective data if it has large enough training data [4]. However, this method also has some drawbacks, such as the computational cost being quite high because it requires calculating the query instance distance in the entire training sample [4]. The following presented in Figure 2 is the form of the KNN method in determining the approach of an object with the KNN calculation.
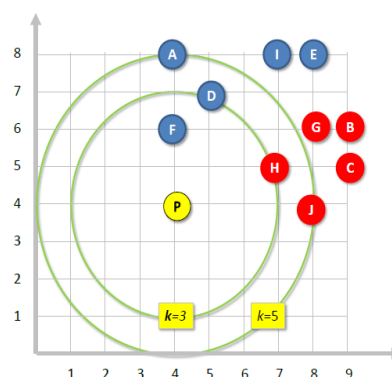
Figure 2. Calculation of the K-Nearest Neighbor Method (Source Google)

Data like in Fig. 2. have new data and a number of training data (training data) by grouping two types of colors, the first is blue and red. Then the new data is presented to determine what color the new data is? There are many ways to measure the proximity between new and old data (training data), including Euclidean distance, Manhattan distance (city block distance), and Minkowski. The most commonly used is Euclidean distance [4]. The equation formula for calculating the distance to be used :

$$d_{st} = (x_s - y_t)(x_s - y_t)^2 \tag{1}$$

$$d_{st} = \sum_{j=1}^{n} |x_{sj} - y_{tj}| \tag{2}$$

$$d(x,y) = (\sum_{i=1}^{m} |x_i - y_i|^r)^{1/r} \tag{3}$$

2.5. Design System

We will discuss the system and data used in the smartphone purchase recommendation system at this stage using the KNN method.

2.5.1. Flowchart Diagrams Recommended System

The flow of the system used in this research uses three calculations to display the system results that are quite good and varied for users or users of this mobile phone purchase recommendation system website. System flow or flowchart diagrams can be shown in Figure 3 below.
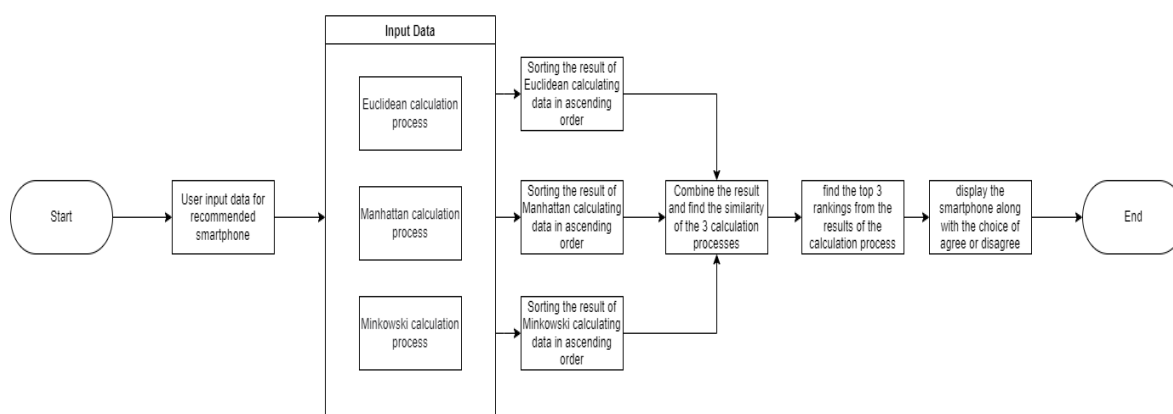

Figure 3. Flowchart Diagrams Recommendation System

The system starts with the user inputting data to get the results of their smartphone recommendations. The data entered will be processed by the Euclidean calculation system, then the Manhattan calculation process, and the Minkowski calculation process. The results of all these calculations will look for the highest data to be combined and look for similarities. After getting the results, the last stage will display all the recommendations generated by the calculation process. The number of recommendations displayed is at least three, such as the flow of this smartphone purchase recommendation system.

2.5.2. Flowchart Diagrams Euclidean Calculation System

The flow of the Euclidean, Manhattan, and Minkowski calculation systems can be seen in the flowchart of the calculation system.
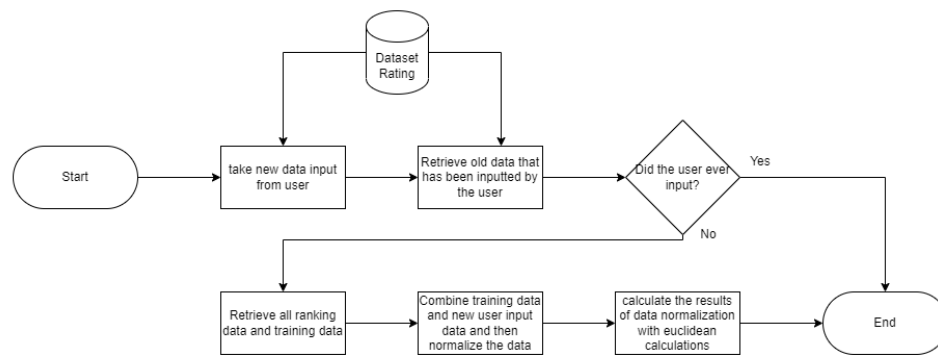
Figure 4. Flowchart Diagrams Euclidean Calculation System

The Euclidean, Manhattan, and Minkowski systems of calculation are not very different, and it's just that the formulas used in the calculations determine different values. Starting from retrieving new data from the user and retrieving old data on the dataset, both those that have been previously inputted will display the results of the previous user's data. If the user never has input data, then the data will be processed and combined with the training data and normalized. The final stage will be calculated using the Euclidean calculation formula to get the Euclidean value for all training data.

3. RESULTS AND DISCUSSION

In this section, we will discuss the results and data processing to become recommendation data for smartphone buyers. System implementation will also be discussed in this section.

3.1 Dataset Review (origin of the dataset)

The review dataset in this study was taken from the DXOMARK website for reviews made by those who have done the research, and this site focuses on providing reviews of cameras and displays. Retrieval of review data using manual methods, not automatic data retrieval methods from other websites or a web crawling system [19]. The review data taken were 223 smartphone review data and five different review content. Five review content includes camera, selfie, audio, display, and battery reviews.

3.2 Cleaning Data

Cleaning data is needed to make a calculation, or data processing can be done properly. Data cleaning can be done in many ways, such as normalizing data, searching for outliers, or changing character data to nominal for easy calculations. Table 1 is the initial data that has not been cleaned of data.

Table 1. Initial Dataset Table

No	Name	Lunch	Price	Camera	Selfie	Audio	Dispaly	Battery
1	Huawei P50 Pro	Jul 2021	$907	144	106		93	
2	Xiaomi Mi 11 Ultra	Mar 2021	$1200	143	94	71	87	69
3	Apple Iphone 13 Pro Max	Sep 2021	$1099	137	99	75	99	89

After the data has been cleaned, the data is ready to be used for calculation. The data that has gone through the cleaning process can be seen in table 2, namely the data in the form of normalization, and the data is no longer a character data type except for the name of the smartphone.

Table 2. Dataset Table After Cleaning

No	Name	Norm_Price	Norm_Camera	Norm_Selfie	Norm_Audio	Norm_Dispaly	Norm_Battery
1	Huawei P50 Pro	0.45	1.0	1.0	0.0	0.93	0.0
2	Xiaomi Mi 11 Ultra	0.60	0.99	94	0.71	0.87	0.69
3	Apple Iphone 13 Pro Max	0.68	0.96	99	0.75	0.99	0.89
4	Max	1999	144	106	100	100	100

3.3 Operation System

The Operation System applied in this study uses the K-Nearest Neighbor (KNN) algorithm for a website-based smartphone purchase recommendation system. This algorithm is chosen because it is in the classification stage for attributes based on learning data closest to the smartphone attributes. The recommendation system on this website is divided into two ways to provide recommendations. The first stage is checking the input data. Has it been filled with other users? If true, the recommendation results will be displayed like other user data. In the second stage, if it is wrong to check another user's data, then the input from the user will go through the KNN calculation process. The data used in calculating this recommendation include price, camera, selfie, audio, display, and battery in determining the recommendation. Examples of training and testing data can be seen in table 1.

3.4 Implementation

3.4.1 Main Menu Page

Figure 5 shows the first page when the user opens the website. This page only provides information about the menu provided to get smartphone recommendations using data that the user will input later. This website has a home menu, input menu, Smartphone data menu, and results in the menu.

Figure 5. Main Menu Page

3.4.2 Smartphone Data Menu Page

Figure 6 shows the smartphone menu data page. When the user presses, it will display smartphone information and several smartphone brands, which will become recommendation data to the user. There are three menu tabs, and the first is Rating Data which displays smartphone rating information from various brands. The second tab displays information on smartphone specifications from various brands. The third tab of data results displays information on input results from users who have provided reviews of the results of recommendations given to users. Examples of reviews given by users such agree or disagree.

Figure 6. Data Smartphone Page

3.4.3 User Data Input Page

Figure 7 shows a user input page consisting of 6 contents that the user must fill out. The first content of the user must determine the smartphone's price, with the lowest price of 1 million to the highest at 30 million. The content of both users must determine the camera rating, ranging from a rating value of 0 to 150. The third content user must determine the selfie rating value, starting from a rating value of 0 to 106 rating values. The fourth, fifth, and sixth content users must specify a rating value ranging from 0 to 100. After pressing the submit button, the system will run two stages to make smartphone recommendations.

Figure 7. User Data Input Page

3.4.4 Recommended Results Page

Figure 8 shows the recommendation results page by displaying smartphone specification data in the dataset. This page will exit after the user enters data in the input menu. This recommendation results page has an agree or disagree button, which aims to make users provide reviews of the recommendations that this website-based recommendation system has given. Users can only choose one of the recommendations on this page so that user reviews can be determined from the most relevant ones that users want to buy. If the user does not agree with the results of this recommendation, they do not have to choose a smartphone first; just click the disagree button.

Figure 8. Recommended Result Page

3.4.5 Data Pages All Recommended Results

Figure 9 shows the results of all user reviews who have looked for smartphone recommendations on this website. One hundred users have commented on the recommendation system in the current study. This page only displays data that has been entered by the user and the smartphone they have chosen to be the most preferred smartphone by the user from the results of the recommendations given. The following is presented in Figure 10 which is the result of a review from users who have used the recommendation system and gave the results of agreeing or disagreeing.

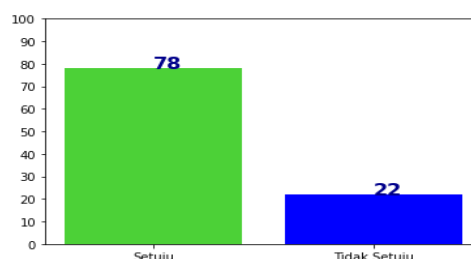**Figure 9.** Data All Recommended Result

Figure 10. Reviews User

4. CONCLUSION

The conclusion of this research and discussion is to provide recommendations to users to determine the smartphone they want to buy. Of the six content that the user has entered, the system will perform the calculation process and check the data that the user has entered. The checking process on the dataset is done manually, so there is no need for a particular method to check the data. The calculation process on the input data by the user uses the KNN method, which has three different calculations, namely Euclidean, Manhattan, and Minkowski. From this study, we will not compare the 3 KNN calculations but combine the results of the 3 KNN calculations to become a recommendation for smartphones. The results of this study provide a review of the recommendation system using the KNN calculation method, which has received 100 reviews. 78 reviews out of 100 agree with the recommended smartphone, and 22 out of 100 disagree. So there is a percentage for this recommendation system to get a score of 78% for success and 22% for failure. This study also gives value to the KNN method. With the results obtained, the KNN calculation method can be a good recommendation system for providing recommendations.

REFERENCES

[1] newzoo.com, "Top Countries by Smartphone Users," newzoo.com, 2020. [Online]. Available: https://newzoo.com/insights/rankings/top-countries-by-smartphone-penetration-and-users. [Diakses June 2022].

[2] M. F. N. D. Chuzaimah, "Smartphone: Antara Kebutuhan dan E-Lifestyle," Seminar Nasional Informatika 2010, pp. E-312, 2010.

[3] R. D. Syah, "Performa Algoritma User K-Nearest Neighbors pada Sistem Rekomendasi di Tokopedia," Jurnal Informatika Universitas Pamulang, vol. 5, no. 3, pp. 302-306, 2020.

[4] S. Abraham dan Y. D. Rahayu, "Sistem Rekomendasi Artikel Berita Menggunakan Metode K-Nearest Neighbor Berbasis Website," Prosiding SENSEI 2017, pp. 179-187, 2017.

[5] N. L. G. P. Suwirmayanti, "Penerapan Metode K-Nearest Neighbor Untuk Sistem Rekomendasi Pemilihan Mobil," Techno.COM, vol. 2, no. 16, pp. 120-131, 2017.

[6] C. A. Rahardja, T. Juardi dan H. Agung, "Implementasi Algoritma K-Nearest Neighbor Pada Website Rekomendasi Laptop," Jurnal Buana Informatika, vol. 10, no. 1, pp. 75-84, 2019.

[7] C. S. D. Prasetya, "SISTEM REKOMENDASI PADA E-COMMERCE MENGGUNAKAN K-NEAREST NEIGHBOR," Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), vol. 4, no. 3, pp. 194-200, 2017.

[8] H. H. Arfisko dan A. T. Wibowo, "Sistem Rekomendasi Film Menggunakan Metode Hybrid Collaborative Filtering Dan Content-Based Filtering," e-Proceeding of Engineering, vol. 9, no. 3, pp. 2149-2159, 2022.

[9] K. A. DS, Z. Baizal dan Adiwijaya, "Recommender System Based on User Functional Requirements Using Euclidean Fuzzy," International Conference on Information and Communication Technology (ICoICT), pp. 455-460, 2015.

[10] J. Wu, Z.-h. Cai* dan S. Ao, "Hybrid dynamic k-nearest-neighbour and distance and attribute weighted method for classification," International Journal of Computer Applications in Technology, 2012.

[11] Y. REDITYAMURTI, ADIWIJAYA dan Z. A. BAIZAL, "Compound Critiquing for Conversational Recommender System Based on Functional Requirement," Advanced Science Letters, vol. 22, no. 8, pp. 1892-1896, 2017.

[12] E. Turban, R. Sharda dan D. Delen, Decision Support And Business Intelligence Systems (9th Edition), New Jersey: Pearson Education India, 2013.

[13] K. MCGEE, "The Advantages of Rating Scales," Classroom.synonym, 3 August 2018. [Online]. Available: https://classroom.synonym.com/advantages-rating-scales-6151387.html. [Diakses 2 June 2022].

[14] L. Rokach, B. Shapira dan F. Ricci, Recommender Systems Handbook, Boston, MA: Springer US, 2010.

[15] T. K. Akbar Nur Syahrudin, "INPUT DAN OUTPUT PADA BAHASA PEMROGRAMAN PYTHON," Jurnal Dasar Pemograman Python STMIK, 2018.

[16] Flask, "User's Guide," Flask.com, [Online]. Available: https://flask.palletsprojects.com/en/2.1.x/. [Diakses Jun 2022].

[17] R. Maulid, "Mengenal Flask, Library Machine Learning Python Idaman Developer," dqlab, 1 September 2021. [Online]. Available: https://dqlab.id/mengenal-flask-library-machine-learning-python-idaman-developer. [Diakses Jun 2022].

[18] R. Irsyad, "Penggunaan Python Web Framework Flask Untuk Pemula," Institut Teknologi Bandung, 2018.

[19] R. Hanifah dan I. S. Nurhasanah, "IMPLEMENTASI WEB CRAWLING UNTUK MENGUMPULKAN INFORMASI WISATA KULINER DI BANDAR LAMPUNG," Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), vol. 5, no. 5, pp. 531-536, 2014.