Implementation of K-Nearest Neighbor Method and Weighted Product Method in Determining High School Majors

Kanda Januar[1], Yunita[*2], Kartika Rahmayani[3]
[1,2,3]Fakultas Ilmu Komputer, Universitas Sriwijaya, Indonesia
E-mail: [1]kandajanuar@gmail.com, [*2]**yunita@ilkom.unsri.ac.id**,
[3]kartikarahmayani99@gmail.com

Abstract

The Ministry of Education and Culture has determined that high school education in Indonesia is divided into 2 majors, namely science and social studies. These majors will usually also affect students when they will continue their education to a higher level of education. Students often find it difficult to determine which major to choose according to their respective abilities, interests and talents. This study aims to assist the school in providing suggestions for the division of student majors and provide more accurate results by using the K-Nearest Neighbor method which will classify students into science or social studies. Furthermore, the Weighted Product Method will be applied so that the student data that has been classified can be sorted based on the best value so that class division can be carried out based on criteria values that have different weight values. The results of this study provide the highest accuracy value for the K-Nearest Neighbor method using the k value configuration of 88% and the accuracy value of 84% using the Weighted Product method.

Keywords *—K-Nearest Neighbor, Majors, High School, Weighted Product*

1. INTRODUCTION

Indonesia is a country that has a learning curriculum that requires high school education to determine majors. The determination of majors in high school itself consists of IPA (Natural Sciences), IPS (Social Sciences) and in some schools there are also language departments. The determination of this major is done so that students can choose their interests and adjust them to their abilities. In determining the majors, it is not determined only based on students' interests, but also the values they get during class X [1]. So that students are divided and are in the appropriate department, majors are classified using the K-Nearest Neighbor method.

The K-Nearst Neighbor algorithm works by looking at the minimum distance from the new data to the k closest neighbors that have been determined. After obtaining the k closest neighbors, the class prediction of the new data is determined based on the majority of the k closest neighbors to the old data [2]. In the research "Application of decision support system using the K-Nearest Neighbor and Weighted Product method for determining the recipients of low-income family scholarship (GAKIN) (case study: Poltekes Kemenkes Semarang). Neighbor up to 99.61% [3]. Based on the results of these studies which have a high level of accuracy, this research was conducted using the K-Nearest Neighbor method to classify the majors of student test data on student training data in accordance with the data they have.

To assist in supporting decision making, the Weighted Product method is used which will sort data on student scores that have been classified by major so that they can be ranked to get the best results in class division in each department. The Weighted Product method performs multiplication to connect the criteria, where the value of each criterion must first be raised to the power of the criteria weight. According to the available data, there are up to 19 criteria with different weights of importance that are used in the decision-making process. The Weighted Product method was chosen because it is able to choose the best alternative from a number of existing alternatives. One of the previous studies using the Weighted Product method is a "Decision Support System to determine the best quality rambutan fruit using the Weighted Product (WP) method" [4]. Another study entitled "The Application of the K–Nearest Neighbor (K-NN) Method and the Weighted Product (WP) Method in Accepting Prospective Teachers and New Administrative Employees with a Technology Insight (Case Study: Muhammadiyah 2 Kediri Vocational High School)." This research uses a combination of the two methods and has an accuracy rate of the Weighted Product method of up to 89% [5].

In this research, the process framework in this research can be described as follows:

1. Conduct a literature study on the K-Nearest Neighbor and Weighted Product methods from related research journals, textbooks, e-books and several websites related to existing methods.
2. Identify problems to be raised in research based on real problems that occur in the scope of research and are carried out in accordance with the stages of research.
3. After formulating the problem and drawing conclusions, will the problem be solved by the method that has been studied and obtain the appropriate literature, then data collection is carried out in the scope that has been determined and compiled in documentation related to software development.
4. Develop the software that has been determined in this study using object-based programming.
5. Perform testing of software that is being developed based on real problems that have been determined previously.
6. Analyze the test results from research on the developed software.
7. Draw conclusions from the research that has been done.The introduction outlines the background of the problem being solved, the issues related to the problem being solved, the research reviews that have been done before by other researchers that are relevant to the research conducted.

The stages of research in this study can be seen as described in Figure 1.

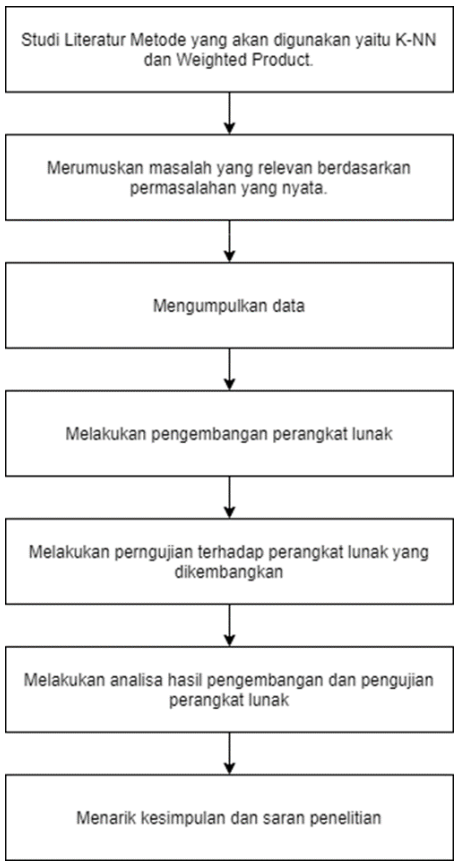

Figure 1. Research Stages

2. RESEARCH METHOD

This research methodology uses K-Nearsest Neighbor and Weighted Product. K-Nearset Neighbor is used to classify student data which will be classified into science and social studies majors. The Weighted Product method is used to rank student data in each department.

2.1. K-Nearset Neighbor

K-Nearest Neighbor is a group of instead-based learning which is also a lazy learning technique. The purpose of this algorithm is to classify new objects based on attributes and training samples. Given a query point, it will find a number of k objects or (training points) closest to the query point. Classification uses a value that is close to the value of k objects. The K-Nearest Neighbor algorithm uses neighboring classification as the predictive value of the new query instance. Because this research defines the distance between two objects, the formula used is the distance calculation formula using Euclidian Distance.

The plot description or notation on the K-Nearest Neighbor algorithm method using Euclidian Distance can be seen as follows [2]:

$$E\ (x, y) = \sqrt{\sum_{i=0}^{n}(xi - yi)^2} \quad (1)$$

The K-Nearest Neighbor algorithm works based on the minimum distance from the new data to the k nearest neighbors that have been set. After determining the k nearest neighbors, the class prediction of the new data classification will be determined based on the majority of the k nearest neighbors. K-Nearest Neighbor can be divided into 2 types based on neighbors that are used as a reference for calculations, namely:

1. Nearest Neighbor, namely the classification is carried out on 1 neighbor data which has the closest label.
2. K-Nearest Neighbor, namely the classification is carried out on k neighbor data which has the closest label with k must be greater than 1 and odd.

Steps to use the K-Nearest Neighbor algorithm:

1. Determine the parameter k = number of nearest neighbors.
2. Calculate the distance between testing data and all training data using Euclidean Distance.
3. Sort the distance results and set the number of nearest neighbors accordingly with a specified value of k.
4. Take the index value of the number of k as the closest data.
5. Change the distance index to a label

2.2. Weighted Product

Weighted Product is one of the methods used to solve the problem of Multi Attribute Decision Making. This method is able to choose the best alternative from a number of alternatives. The Weighted Product method uses multiplication to connect the criteria values, where the value of each criterion must first be raised to the power of the criteria concerned [4].

$$S_i = \prod_{i=1}^{n} X_{ij}^{wj} \quad (2)$$

With i = 1 , 2, ..., m and j = 1, 2,…, n
Description:
= product
Si = score / value of each alternative
Xij = i-th alternative value of the j-th attribute
wj = weight of each attribute

Where = 1 is a positive rank for the benefit / benefit attribute and a negative value for the cost / cost attribute.

For ranking / looking for the best alternative is done by the following formula:

$$V_i = \frac{\prod_{j=1}^{n} X_{ij}^{wj}}{\prod_{j=i}^{n} \left(X_{ij}^{*}\right)^{wj}} \quad (3)$$

Determination of the weight value can be done by giving a priority scale to each of the existing criteria. The criteria that will be the benchmark in this study are the values of compulsory subjects in each of their respective majors. Determination of criteria and weights in this study can be seen in table 1.

Table 1. Criteria and Weights

No	Criteria (Subject)	Weight	Description	Attribute
1	Pkn	2	Need	
2	Seni Budaya	2	Need	
3	Penjaskes	2	Need	
4	Prakarya dan Kewirausahaan	2	Need	
5	Agama	2	Need	
6	UN B. Indonesia	2	Need	
7	UN B. Inggris	2	Need	
8	UN Matematika	3	Quite Influential	
9	B. Indonesia	4	Take Effect	
10	B. Inggris	4	Take Effect	
11	Sejarah Indonesia	4	Take Effect	
12	Sosiologi	5	Very Influential	**COST**
13	Ekonomi	5	Very Influential	**COST**
14	Geografi	5	Very Influential	**COST**
15	Matematika	5	Very Influential	**BENEFIT**
16	Biologi	5	Very Influential	**BENEFIT**
17	Fisika	5	Very Influential	**BENEFIT**
18	Kimia	5	Very Influential	**BENEFIT**
19	Matematika Peminatan	5	Very Influential	**BENEFIT**

3. RESEARCH RESULTS AND DISCUSSION

In this study, the data used were 100 student data consisting of names, majors and the student's grades in each subject. The 100 data will be divided into two parts, namely 25 student data which is used as testing data and 75 student data which is used as training data. Research testing will be carried out using the configuration of the number of k parameters, namely 3, 7, 11, 13, 25, 31 and 73.

The other k values produce output with the same Tp, Tn, Fp and Fn values so that they get the same confusion matrix results. Then the value of k was chosen because it can represent the possible values of k which have different Tp, Tn, Fp and Fn values for each k value so that they get different confusion matrix results. The results of the classification of majors from the

calculation process of the K-Nearset Neighbor method using the k value parameter 3,7,11,13,25,31 and 73 and the ranking of student data using the Weighted Product method will be shown in the Table 2 to Table 8.

Table 2. Calculation Results k=3

Rank	Name	Major (K-NN)
1	Mhd Azkha Aldine Munthe	IPA
2	Ahmad Askari Akhyar	IPA
3	Khairil Abdy Ritonga	IPA
4	Aldha Mufidha	IPA
5	Ayu Putri Aisyah	IPA
6	Satria Parbarita	IPA
7	Ahmad Haidar Ramadhan Syanie	IPA
8	Anggi Nai Fitri Rambe	IPA
9	Ririn Wardhani	IPA
10	Raihan Munthe	IPA
11	Puan Betsyeda	IPS
12	Hefla Syahputri	IPA
13	Putri Handayani	IPA
14	Evitaria Sihotang	IPS
15	Putri Rahadani Tambunan	IPA
16	Siti Nuriah	IPS
17	April Nurul Giyanti	IPA
18	Rukiyah Br Rambe	IPA
19	Cindy Claudia	IPA
20	Muhammad Syaifan	IPA
21	Sri Rezeki	IPS
22	Mutiah	IPS
23	Meli Saputri Saragih	IPA
24	Anggi Tri Yuliasih	IPS
25	Rispiza Syahputri	IPS

Table 3. Calculation Results k=7

Rank	Name	Major (K-NN)
1	Mhd Azkha Aldine Munthe	IPA
2	Ahmad Askari Akhyar	IPA
3	Khairil Abdy Ritonga	IPA
4	Aldha Mufidha	IPA
5	Ayu Putri Aisyah	IPA
6	Satria Parbarita	IPA
7	Ahmad Haidar Ramadhan Syanie	IPA
8	Anggi Nai Fitri Rambe	IPA
9	Ririn Wardhani	IPA
10	Raihan Munthe	IPA
11	Puan Betsyeda	IPA
12	Hefla Syahputri	IPA
13	Putri Handayani	IPA
14	Evitaria Sihotang	IPA
15	Putri Rahadani Tambunan	IPA
16	Siti Nuriah	IPS
17	April Nurul Giyanti	IPA
18	Rukiyah Br Rambe	IPA

Table 4. Calculation Results k=11

Rank	Name	Major (K-NN)
1	Mhd Azkha Aldine Munthe	IPA
2	Ahmad Askari Akhyar	IPA
3	Khairil Abdy Ritonga	IPA
4	Aldha Mufidha	IPA
5	Ayu Putri Aisyah	IPA
6	Satria Parbarita	IPA
7	Ahmad Haidar Ramadhan Syanie	IPA
8	Anggi Nai Fitri Rambe	IPA
9	Ririn Wardhani	IPA
10	Raihan Munthe	IPA
11	Puan Betsyeda	IPA
12	Hefla Syahputri	IPA
13	Putri Handayani	IPA
14	Evitaria Sihotang	IPA
15	Putri Rahadani Tambunan	IPA
16	Siti Nuriah	IPA
17	April Nurul Giyanti	IPA
18	Rukiyah Br Rambe	IPA
19	Cindy Claudia	IPA
20	Muhammad Syaifan	IPA
21	Sri Rezeki	IPA
22	Mutiah	IPS
23	Meli Saputri Saragih	IPA
24	Anggi Tri Yuliasih	IPS
25	Rispiza Syahputri	IPS

Table 5. Calculation Result k-13

Rank	Name	Major (K-NN)
1	Mhd Azkha Aldine Munthe	IPA
2	Ahmad Askari Akhyar	IPA
3	Khairil Abdy Ritonga	IPA
4	Aldha Mufidha	IPA
5	Ayu Putri Aisyah	IPA
6	Satria Parbarita	IPA
7	Ahmad Haidar Ramadhan Syanie	IPA
8	Anggi Nai Fitri Rambe	IPA
9	Ririn Wardhani	IPA
10	Raihan Munthe	IPA
11	Puan Betsyeda	IPA
12	Hefla Syahputri	IPA
13	Putri Handayani	IPA
14	Evitaria Sihotang	IPS
15	Putri Rahadani Tambunan	IPA
16	Siti Nuriah	IPA
17	April Nurul Giyanti	IPA
18	Rukiyah Br Rambe	IPA
19	Cindy Claudia	IPA
20	Muhammad Syaifan	IPA
21	Sri Rezeki	IPA
22	Mutiah	IPS
23	Meli Saputri Saragih	IPA
24	Anggi Tri Yuliasih	IPS
25	Rispiza Syahputri	IPS

Table 6. Calculation Result k=25

Rank	Name	Major (K-NN)
1	Mhd Azkha Aldine Munthe	IPA
2	Ahmad Askari Akhyar	IPA
3	Khairil Abdy Ritonga	IPA
4	Aldha Mufidha	IPA
5	Ayu Putri Aisyah	IPA
6	Satria Parbarita	IPA
7	Ahmad Haidar Ramadhan Syanie	IPA
8	Anggi Nai Fitri Rambe	IPA
9	Ririn Wardhani	IPA
10	Raihan Munthe	IPA
11	Puan Betsyeda	IPA
12	Hefla Syahputri	IPA
13	Putri Handayani	IPA
14	Evitaria Sihotang	IPA
15	Putri Rahadani Tambunan	IPA
16	Siti Nuriah	IPA
17	April Nurul Giyanti	IPA
18	Rukiyah Br Rambe	IPA
19	Cindy Claudia	IPA
20	Muhammad Syaifan	IPA
21	Sri Rezeki	IPA
22	Mutiah	IPA
23	Meli Saputri Saragih	IPA
24	Anggi Tri Yuliasih	IPS
25	Rispiza Syahputri	IPS

Table 7. Calculation Result k-31

Rank	Name	Major (K-NN)
1	Mhd Azkha Aldine Munthe	IPA
2	Ahmad Askari Akhyar	IPA
3	Khairil Abdy Ritonga	IPA
4	Aldha Mufidha	IPA
5	Ayu Putri Aisyah	IPA
6	Satria Parbarita	IPA
7	Ahmad Haidar Ramadhan Syanie	IPA
8	Anggi Nai Fitri Rambe	IPA
9	Ririn Wardhani	IPA
10	Raihan Munthe	IPA
11	Puan Betsyeda	IPA
12	Hefla Syahputri	IPA
13	Putri Handayani	IPA
14	Evitaria Sihotang	IPA
15	Putri Rahadani Tambunan	IPA
16	Siti Nuriah	IPA
17	April Nurul Giyanti	IPA
18	Rukiyah Br Rambe	IPA
19	Cindy Claudia	IPA
20	Muhammad Syaifan	IPA
21	Sri Rezeki	IPA
22	Mutiah	IPA
23	Meli Saputri Saragih	IPA
24	Anggi Tri Yuliasih	IPA
25	Rispiza Syahputri	IPS

Table 8. Calculation Result k-73

Rank	Name	Major (K-NN)
1	Mhd Azkha Aldine Munthe	IPA
2	Ahmad Askari Akhyar	IPA
3	Khairil Abdy Ritonga	IPA
4	Aldha Mufidha	IPA
5	Ayu Putri Aisyah	IPA
6	Satria Parbarita	IPA
7	Ahmad Haidar Ramadhan Syanie	IPA
8	Anggi Nai Fitri Rambe	IPA
9	Ririn Wardhani	IPA
10	Raihan Munthe	IPA
11	Puan Betsyeda	IPA
12	Hefla Syahputri	IPA
13	Putri Handayani	IPA
14	Evitaria Sihotang	IPA
15	Putri Rahadani Tambunan	IPA
16	Siti Nuriah	IPA
17	April Nurul Giyanti	IPA
18	Rukiyah Br Rambe	IPA
19	Cindy Claudia	IPA
20	Muhammad Syaifan	IPA
21	Sri Rezeki	IPA
22	Mutiah	IPA
23	Meli Saputri Saragih	IPA
24	Anggi Tri Yuliasih	IPA
25	Rispiza Syahputri	IPA

The Weighted Product method performs the data calculation process to get the results of ranking student data which will later be divided into each class for each major. The table 9 and 10 will show the results of the student data calculation process based on the data used by using the Weighted Product method which is divided based on the majors that students have obtained.

Table 9. IPA Student Rankings Results

Rank	Name	Major (K-NN)
1	Mhd Azkha Aldine Munthe	IPA
2	Ahmad Askari Akhyar	IPA
3	Khairil Abdy Ritonga	IPA
4	Aldha Mufidha	IPA
5	Ayu Putri Aisyah	IPA
6	Satria Parbarita	IPA
7	Ahmad Haidar Ramadhan Syanie	IPA
8	Anggi Nai Fitri Rambe	IPA
9	Ririn Wardhani	IPA
10	Raihan Munthe	IPA
11	Puan Betsyeda	IPA
12	Hefla Syahputri	IPA
13	Putri Handayani	IPA
14	Evitaria Sihotang	IPA
15	Putri Rahadani Tambunan	IPA
16	April Nurul Giyanti	IPA
17	Rukiyah Br Rambe	IPA
18	Cindy Claudia	IPA
19	Muhammad Syaifan	IPA

Rank	Name	Major (K-NN)
20	Meli Saputri Saragih	IPA
21	Rispiza Syahputri	IPA

Table 10. IPS Student Rankings Results

Rank	Name	Major (K-NN)
1	Siti Nuriah	IPS
2	Sri Rezeki	IPS
3	Mutiah	IPS
4	Anggi Tri Yuliasih	IPS

4. CONCLUSION

Based on the results of the test calculations for the K-Nearest Neighbor method using the confusion matrix to determine the values of accuracy, recall, precision and f-1 score on the system that has been built, the calculation results will be displayed using the k value comparison configuration, namely 3, 7, 11, 13 , 25, 31 and 73. It can be seen that the equations of the values of accuracy, recall, precision and f-1 score with different k values can be seen. Based on the table 11, it can be seen that the test experiment using the value of k=11 has the highest accuracy, recall and f-1 score and the highest precision value is found in the test experiment using the value of k=7.

Table 11. Testing Scenarios of k Value

Score K	Tp	Fp	Accuracy (%)	Recall (%)	Precision (%)	F-1 Score (%)
3	18	0	0.88	0.86	1	0.92
7	19	1	0.88	0.91	0.95	0.93
11	20	2	0.88	0.95	0.91	0.93
13	19	2	0.84	0.91	0.91	0.91
25	20	3	0.84	0.95	0.87	0.91
31	20	4	0.80	0.95	0.83	0.89
73	21	4	0.80	0.85	0.83	0.89

While the calculation of the confusion matrix in the Weighted Product method is carried out by adjusting the original ranking data with the results of calculations using the system, the accuracy value is 84%, precision is 91%, recall is 91% and the f-1 score is 91%. With values Tp= 19, Tn=2, Fp=2 and Fn=2. An overview of the comparison of the ranking of the original data with the system can be seen in the Table 12.

Table 12. Weighted Product Test Scenarios

Rank	Name	System	Original
1	Mhd Azkha Aldine Munthe	IPA	IPA
2	Ahmad Askari Akhyar	IPA	IPA
3	Khairil Abdy Ritonga	IPA	IPA
4	Aldha Mufidha	IPA	IPA
5	Ayu Putri Aisyah	IPA	IPA
6	Satria Parbarita	IPA	IPA
7	Ahmad Haidar Ramadhan Syanie	IPA	IPA
8	Anggi Nai Fitri Rambe	IPA	IPA
9	Ririn Wardhani	IPA	IPA

Rank	Name	System	Original
10	Raihan Munthe	IPA	IPA
11	Puan Betsyeda	IPA	IPA
12	Hefla Syahputri	IPA	IPA
13	Putri Handayani	IPA	IPA
14	Evitaria Sihotang	IPA	IPA
15	Putri Rahadani Tambunan	IPA	IPS
16	Siti Nuriah	IPA	IPA
17	April Nurul Giyanti	IPA	IPA
18	Rukiyah Br Rambe	IPA	IPA
19	Cindy Claudia	IPA	IPA
20	Muhammad Syaifan	IPA	IPS
21	Sri Rezeki	IPA	IPS
22	Mutiah	IPS	IPS
23	Meli Saputri Saragih	IPA	IPA
24	Anggi Tri Yuliasih	IPS	IPS
25	Rispiza Syahputri	IPS	IPA

The use of the K-Nearest Neighbor method in classifying majors on 25 student testing data at SMAN 3 using different k values can produce classifications with different labels, depending on the number of closest data labels. And in this study the use of the K-Nearest Neighbor method produces the highest Accuracy value of up to 80%. And the use of the Weighted Product method in ranking student data which depends on the weight value on each different criterion produces a sequence of student testing data in each major. The Weighted Product method has a fairly high accuracy value, namely 84% in performing calculations to get the results of the student testing data sequence.

5. REFERENCES

[1] D. Kurniawan and A. Saputra, "Penerapan K-Nearest Neighbour dalam Penerimaan Peserta Didik dengan Sistem Zonasi," *J. Sist. Inf. BISNIS*, vol. 9, no. 2, 2019, doi: 10.21456/vol9iss2pp212-219.

[2] R. D. Y. Prakoso, B. S. Wiriaatmadja, and F. W. Wibowo, "Sistem Klasifikasi Pada Penyakit Parkinson Dengan Menggunakan Metode K-Nearest Neighbor," in *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, 2020, pp. 63–68.

[3] L. A. Nasher and N. Bahtiar, "Application of decision support system using the K-Nearest Neighbor and Weighted Product method for determining the recipients of low-income family scholarship (GAKIN) (case study: Poltekkes Kemenkes Semarang)," in *Journal of Physics: Conference Series*, 2019, vol. 1217, no. 1, doi: 10.1088/1742-6596/1217/1/012117.

[4] Y. Zai, Mesran, and E. Buulolo, "Sistem Pendukung Keputusan untuk Menentukan Buah Rambutan Dengan Kualitas Terbaik Menggunakan Metode Weighted Product (WP)," *Media Inform. Budidarma*, vol. 1, no. 1, 2017.

[5] N. N. Dzikrulloh, Indriati, and B. D. Setiawan, "Penerapan Metode K–Nearest Neighbor(KNN) dan Metode Weighted Product (WP) Dalam Penerimaan Calon Guru Dan Karyawan Tata Usaha Baru Berwawasan Teknologi," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 5, 2017.