

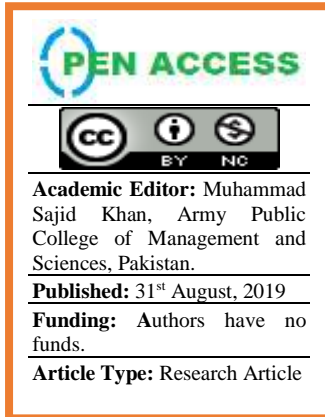
K-NEAREST NEIGHBOR ALGORITHM FOR DETERMINATION OF HIGH SCHOOL PROGRAMS

AGUNG TRIAYUDI, DENY HIDAYATULLAH, VIDILA ROSALINA

Universitas Nasional, Indonesia
Universitas Serang Raya, Indonesia

Email : agungtriayudi@civitas.unas.ac.id, rafadeari@gmail.com, vidila.suhendarsah@gmail.com

ABSTRACT



This study aims to assist teachers in determining the direction to be more effective and efficient. This research was conducted using K-Nearest Neighbor method. This study took 50 data placements of students from major data of 228 students. The 50 data values are obtained and placement of students is divided into two parts, namely 40 and 10 training data and test data. The study was conducted in SMA Negeri 1 Waringinkurung starting from May until August. Data are obtained directly from the students that are then analysed by manual calculation and then applied to the determination of the application until a match is found between the area of manual calculations with the application. This application uses the Java programming language with NetBeans as the IDE and MySQL supporting data storage. The application prompts the user to enter student data and aspects are used to determine the majors. The end result of this application form calculations using K-Nearest Neighbor method and browse the student. Based on the results of this study concluded that: the time to determine

the majors is faster and directly determines the majors.

Keywords: K-Nearest Neighbor; prediction; data analysis; determination; study programs;

1. INTRODUCTION

Majors basically has the ideal concept of putting students according to their interests, talents and potential intelligence. Moreover, it can prepare students to pursue higher education. Determination of the majors will affect the next academic level and will affect the field of science or studies for students who will continue to pursue higher education will thus improper majors could very detrimental to students and their future [1].

Majors in SMAN 1 Waringinkurung has been done since the beginning of class X. Majors' knowledge is available in a Natural Science (IPA) and Social Sciences (IPS). In determining the majors, students are required to follow the Test Psikotes and Homogeneous seed Academic Potential Test (PUHA) which is held on the third day of Orientation School (MOS), which on Wednesday. Psikotes test aims to determine the students' talent and intelligence and PUHA test consisting of questions Biology, Physics, Chemistry, Mathematics, and English which aims to find out how deep knowledge of the IPA

However, because there are too many students who would be led temporarily held by teachers who belong to the team deciding majors only three days from the day after the execution of the test, make teachers difficult to determine the majors because the majors are still done manually by grouping existing values by the appropriate department, and still consider the interest of the students so that the method is less effective and efficient because it requires quite a long time [2, 3].

Therefore, it takes a system that can help the process of majors that more effective and efficient. The system created a support system decision due later this system will assist in decision-making departments where appropriate for each student [4]. There are various methods of decision support, however that will be used is the K-Nearest Neighbor method. One technique that is created in the Data Mining is how to search for the data that is to build a model, and then use that model in order to identify the pattern of other data that are not in the database stored, KNN (K-Nearest Neighbor) is a classification algorithm that uses adjacency as the predicted value of the new query instance [5][6]. Algorithm KNN method is simple, operates on the shortest distance from the query instance to the training sample to determine its KNN [7, 8]. Based on the description of the background issues that have been raised, the formulation of the problem is:

1. How to implement the K-Nearest Neighbor algorithm in an application that aims to help the majors in SMAN 1 Waringinkurung?

2. Analyse whether the K-Nearest Neighbor method is suitable when applied to determining the direction in SMAN 1 Waringinkurung?

The author of this study has a purpose, namely to assist teachers in determining the direction for the students of SMAN 1 Waringinkurung. The previous study that became the reference of the research is a study entitled "Implementation of the K-Nearest Neighbor algorithm As Classification Decision Support PPA and BBM Scholarship Recipients. The main idea of the study discusses the k-nearest neighbor algorithm is applied to the data of students who will be selected for admission scholarship PPA and BBM. Comparison of these studies by this research is in common use k-nearest neighbor algorithms for solving problems, but in a previous study k-nearest neighbor algorithm used to determine the scholarship recipients PPA and BBM based on data variables that have been determined. While in this study k-nearest neighbor algorithm is used for determining the direction the students of SMAN 1 Waringinkurung based on data from the previous generation of students' majors [9].

On "Classification Feasibility Awarding Venture Capital Farmers Group Salt Using K-Nearest Neighbors in the Context of Economic Empowerment in the district of Bireuen" explained that the salt business people in North Aceh district became part of the economic development broadly that will encourage the development of national in scope. But the number of venture capital request to the Department of maritime and fisheries become an obstacle in determining the feasibility of providing venture capital to prospective farmers salts according to their respective capabilities. Therefore we need a system that can help the Department of Marine and Fisheries in determining the prospective farmers eligible or not to receive venture capital. The preferred method of making this system is a K-Nearest Neighbor as it can provide accurate results [10].

On "Algorithms K-Nearest Neighbor Classification For System Prediction Predicate Student Achievement" explained that the achievement of students is a form of achievement during the academic activities at a higher education institution in which the predicate is obtained from a prediction were calculated using a K-Nearest Neighbor. K-NN algorithm which is based on the proximity of the old data history (training) with new data (testing). Attribute determination is obtained based on the results of previous studies that have similarities in the case of student predictions are further validated by academic faculty [11].

On "Classification White Blood Cells Based on Feature Color and Form by Method K-Nearest Neighbor (K-NN)" explains that in order to obtain information on the type of blood cells to classify cells according to the procedure using a microscope in a laboratory hematology with a series of laboratory tests can take a long time. As an attempt to overcome this and for the purpose of early diagnosis can use based on morphological image processing techniques to classify blood cells and white blood cells. The image processing algorithms used are hough circle, thresholding, further extraction characteristics for the classification process used method of K-Nearest Neighbor (K-NN). The test results demonstrate the accuracy of segmentation by 78% and 64% classification testing of 100 images were tested [12].

On "Four Wheel Vehicle Classification Based on K-NN" explained that it is very important to choose a safe vehicle for safety. The car is one vehicle that has an adequate level of security. However, the diversity of the car with various brands make consumers should be more careful in choosing. To facilitate consumers in making choices needed a system that could help to decision-making. A classification system using the K-NN is considered to be the best choice because it is able to deduce the car are included in the classification is desirable and has a high accuracy. The more data the higher the level of accuracy [13].

2. RESULTS AND DISCUSSION

In this chapter will be discussed regarding the data to be used in research, the data will be calculated using K-Nearest Neighbor algorithm.

2.1 Data used

The data used in this study is as much as 50 data results of placement of students with attribute values of Potential Commodity Homogeneous Academic (PUHA), the value Psychotest Recommendation Psikotes, Interests Students and Results Majors Of the 30 data majors existing students will be divided into two parts, namely 30 20 training data and test data.

Table. 1 Data sample

No.	Value PUHA	Value Psikotes	Recommendations Psikotes	Student Interest	Result
1.	23	121	IPS	IPA	IPA
2.	20	110	IPS	IPA	IPA
3.	19	103	IPA	IPS	IPS
4.	19	111	IPA	IPA	IPA
5.	17	119	IPA	IPA	IPA
6.	16	117	IPA	IPA	IPA
7.	16	118	IPA	IPA	IPA
8.	16	129	IPA	IPA	IPA
9.	16	123	IPA	IPA	IPA
10.	15	109	IPA	IPA	IPA
11.	15	109	IPA	IPS	IPA
12.	15	102	IPS	IPA	IPA
13.	15	114	IPA	IPA	IPA
14.	14	113	IPA	IPA	IPA
15.	14	109	IPS	IPA	IPA
16.	14	87	IPS	IPS	IPA
17.	13	101	IPS	IPA	IPA
18.	13	109	IPA	IPA	IPA
19.	12	107	IPA	IPA	IPA
20.	12	90	IPA	IPA	IPA
21.	11	89	IPA	IPS	IPS
22.	11	104	IPS	IPA	IPA
23.	11	89	IPS	IPA	IPA
24.	11	112	IPA	IPA	IPA
25.	11	116	IPA	IPA	IPA
26.	10	90	IPS	IPA	IPA
27.	10	82	IPA	IPS	IPS
28.	10	90	IPA	IPA	IPS
29.	10	110	IPS	IPA	IPA
30.	10	116	IPA	IPA	IPA

2.2 Calculation of K-Nearest neighbor

The initial phase should be done in the calculation of majors using K-NN method is to determine the size of the distance between the two data or dissimilarity are sometimes considered a recommendation Psikotes and Interest Students.

Table. 2 Size distance between two data

	IPA	IPS
IPA	0	1
IPS	1	0

As an example of the calculations will be done with the data in the following table:

Table. 3 Data test

No.	value PUHA	value Psikotes	Recommendations Psikotes	Student interests	result
1.	10	97	IPS	IPA	?

To calculate the distance between the test data with the data sample used Euclidean Distance formula.

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \quad (1)$$

Thus, obtained the following results:

Table. 4 Calculation results distance has sorted

No.	value PUHA	Value Psi-kotes	Recommendations Psi-kotes	Student interests	result	discrepancy miripan	The distance (d)
17	13	101	IPS	IPA	IPA	1	5
26	10	90	IPS	IPA	IPA	1	7
12	15	102	IPS	IPA	IPA	1	7.071067812
22	11	104	IPS	IPA	IPA	1	7.071067812
28	10	90	IPA	IPA	IPS	0	7.071067812
20	12	90	IPA	IPA	IPA	0	7.348469228
21	11	89	IPA	IPS	IPS	1	8.062257748
23	11	89	IPS	IPA	IPA	1	8.062257748
19	12	107	IPA	IPA	IPA	0	10.24695077
3	19	103	IPA	IPS	IPS	1	10.81665383
16	14	87	IPS	IPS	IPA	0	10.81665383
18	13	109	IPA	IPA	IPA	0	12.40967365
15	14	109	IPS	IPA	IPA	1	12.64911064
11	15	109	IPA	IPS	IPA	1	13
29	10	110	IPS	IPA	IPA	1	13
10	15	109	IPA	IPA	IPA	0	13.03840481
27	10	82	IPA	IPS	IPS	1	15
24	11	112	IPA	IPA	IPA	0	15.06651917
2	20	110	IPS	IPA	IPA	1	16.40121947
14	14	113	IPA	IPA	IPA	0	16.52271164
4	19	111	IPA	IPA	IPA	0	16.673332
13	15	114	IPA	IPA	IPA	0	17.74823935
30	10	116	IPA	IPA	IPA	0	19.02629759
25	11	116	IPA	IPA	IPA	0	19.05255888
6	16	117	IPA	IPA	IPA	0	20.90454496
7	16	118	IPA	IPA	IPA	0	21.86321111
5	17	119	IPA	IPA	IPA	0	23.10844002
9	16	123	IPA	IPA	IPA	0	26.70205985
1	23	121	IPS	IPA	IPA	1	27.29468813
8	16	129	IPA	IPA	IPA	0	32.57299495

The next stage is to determine the shortest distance to the order of the value of K predetermined. Unknown if the value K is 3 then grade science and social studies class = 2 = 1. This science class to get the most amount so obtained results for "IPA". So, the conclusion is that students have a value corresponding to the test data have been calculated predicted to get into the majors "IPA".

2.3 System implementation

Applications built is based desktop where making use of the Java programming language with its Netbeans as the IDE. Here is a view of the application:

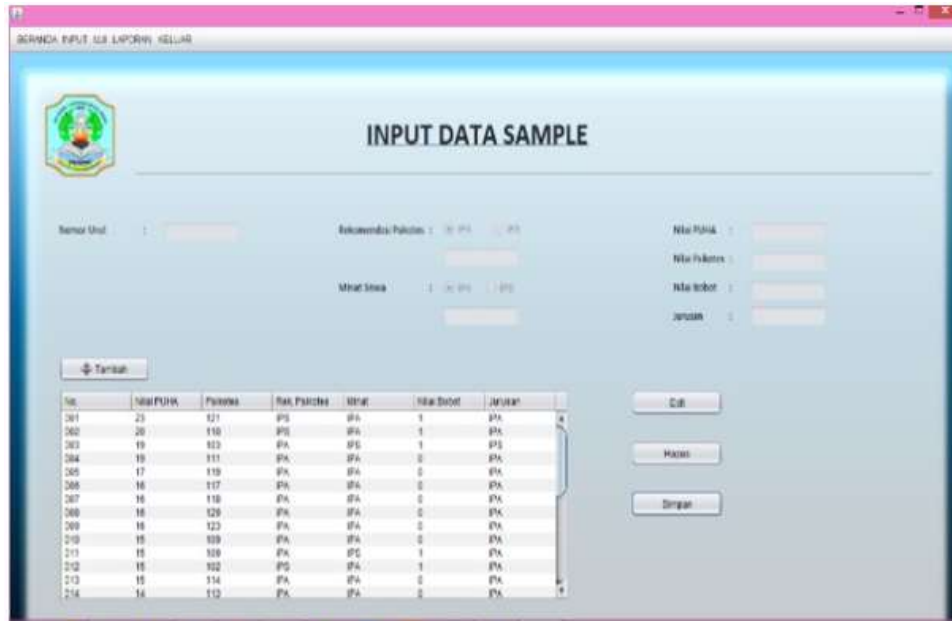


Figure. 1 Data view sample

Figure 1 is a form that serves to add, delete or modify the sample data and to display the sample data that has been entered.

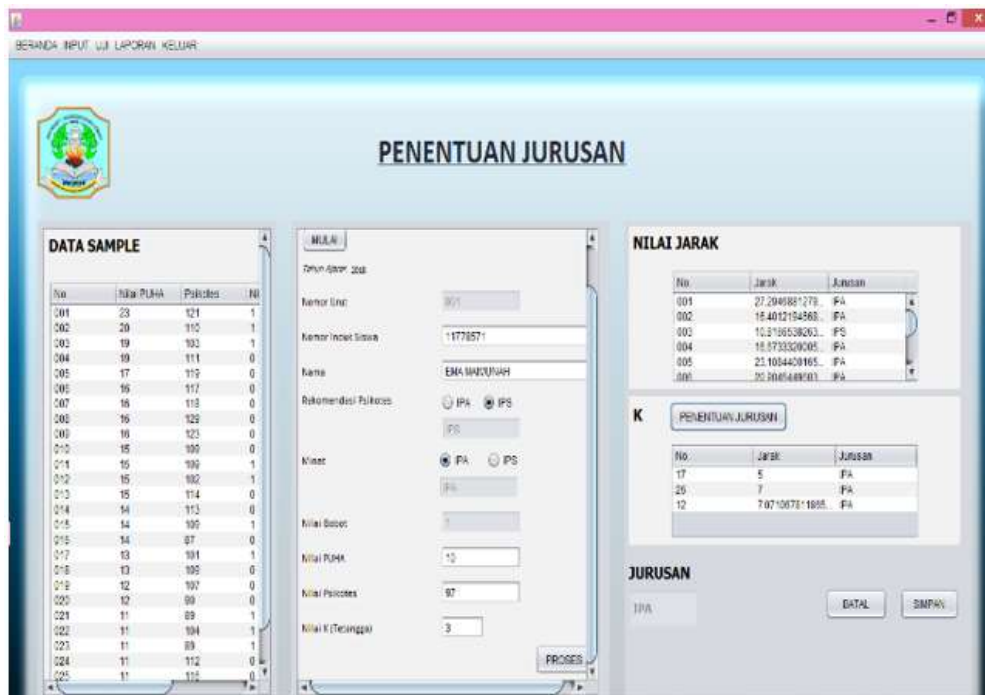


Figure. 2 Display calculation method of determining department K-NN

Figure 2 is a form of testing which in this form of determining the majors with the calculation using the K-NN done. To process the determination of the majors, there are several attributes that must be entered. There are nine attributes that must be entered, two of which are attributes that are filled automatically, the serial number is automatically filled in accordance with the order of students tested, and the weights are filled automatically in accordance with the value of the dissimilarity between attributes on a psychological test and interests of students. Then two attributes that contain the identity of the student is the student identification number and the name of the student, and the five attributes used for the calculation process using the K-NN.

From Figure 2 has shown the results of tests on students based on the data indicated. Acyl displayed is the calculation of the distance between the test data to each of the sample data that was ordered by the order of the

data sample, then the panel K perform distance calculations which have been sorted in ascending and limited based on the value of K that has been entered previously and the last is the result of the determination department.

3. CONCLUSION

From the results of research on decision support system for the determination of the majors SMA 1 Waringinkurung using K-Nearest Neighbor method can be concluded as follows:

1. After testing with test data, the results obtained in applications by using K-Nearest Neighbor method in accordance with the manual calculation of these methods so that the algorithm used in the application is appropriate and can be used.
2. K-Nearest Neighbor method is a method in data mining using sample data in which way this works is to calculate the distance between the test data with the data sample. The shortest distance is the result of the test so in other words this method works based on the similarity of data making it suitable for use in SMA Negeri 1 Waringinkurung where there is no change in behavior / rules in determining the majors.

ACKNOWLEDGEMENT

This research is an outcome of the internal competitive grant scheme of the National University of Indonesia.

REFERENCES

1. Jain AK, Murty MN, Flynn PJ. *Data clustering: a review*. ACM computing surveys (CSUR). 1999 Sep 1;31(3):264-323.
2. Xu R, Wunsch DC. *Survey of clustering algorithms*. 2005.
3. Triayudi A, Fitri I. *A new agglomerative hierarchical clustering to model student activity in online learning*. Telkonnika. 2019 Jun 1;17(3):1226-35.
4. Tan PN. *Introduction to data mining*. Pearson Education India; 2018.
5. Gan G, Ma C, Wu J. *Data clustering: theory, algorithms, and applications*. Siam; 2007 Jul 12.
6. Triayudi A, Fitri I. *Comparison of Parameter-Free Agglomerative Hierarchical Clustering Methods*. ICIC Express Letters. 2018;12(10):973-80.
7. Lopez MI, Luna JM, Romero C, Ventura S. *Classification via clustering for predicting final marks based on student participation in forums*. International Educational Data Mining Society. 2012 Jun.
8. Jain AK, Dubes RC. *Algorithms for clustering data*. Englewood Cliffs, NJ: Prentice hall; 1988 Jan 1.
9. Triayudi A, Fitri I. *Alg Clustering To Analyze The Behavioural Patterns Of Online Learning Students*. Journal of Theoretical & Applied Information Technology. 2018;96(16):5327-37.
10. Elfiana E, Mukhlis M. *Klasifikasi Kelayakan Pemberian Modal Usaha Kelompok Petani Garam Menggunakan K Nearest Neighbors dalam Rangka Pembedayaan Ekonomi Masyarakat di Kab Bireun*. Jurnal Sains Pertanian. 2017;1(1).
11. Mustakim M, Oktaviani G. *Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa*. Jurnal Sains dan Teknologi Industri. 2016 Jun 4;13(2):195-202.
12. Khasanah MN, Harjoko A, Candradewi I. *Klasifikasi Sel Darah Putih Berdasarkan Ciri Warna dan Bentuk dengan Metode K-Nearest Neighbor (K-NN)*. IJEIS (Indonesian Journal of Electronics and Instrumentation Systems). 2016 Oct;6(2):151-62.
13. Nouvel A. *Klasifikasi Kendaraan Roda Empat Berbasis K-NN*. Jurnal Blanglala Informatika, Vols. 2015 Sep;3(2):66-9.
14. Sumarlin S. *Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM*. JSINBIS (Jurnal Sistem Informasi Bisnis). 2015;5(1):52-62.

AUTHORS PROFILE