 ISSN: 2579-7298

K-Means and K-NN Methods For Determining Student Interest

Guslendra[a1], Sarjon Defit [a,2]*

[a]Putra Indonesia University "YPTK" Padang, Indonesia
[1]guslendra@upiyptk.ac.id
** corresponding author*

ARTICLE INFO

Article history:
Received 22 Ags 2021
Revised 25 Feb 2022
Accepted 28 May 2022

Keywords:
K-Mean,
K-NN,
Specialization,
Student

ABSTRACT

Putra Indonesia University 'YPTK' Padang's Department of Information Systems, Faculty of Computing Science has three specializations, namely Information Technology Management, Business Information Systems and Industrial Information Systems. In the fifth semester, the acquisition of specializations takes place. In the next semester the selection of specialist programs will be determined. The option of the degree is adapted to students' needs and capacities. The acquisition of results generated in the previous semester can be seen. The objective of this survey is to provide students with suggestions for collection of degrees. The study was performed using K-Means and K-Nearest Neighbor methods in order to obtain classification of students and the correlation between recent cases and past cases. This analysis uses 13 characteristics, of which 12 are predictors and 1 is the option. The test results can be used as a way to suggest the student preferences based on preset attributes through the K-Means and K-NN methods

I. Introduction

Each university would like all graduates to know their fields. Graduates' knowledge depends heavily on the suitability of students' preferences and talents. In three fields of specialization, namely Information Technology Management, Business Information Systems and Industrial Information System, the YPTK University Informatics Study Program produces graduates. Starting from the 5th semester, the selection of specializations may be made. In the next semester, the selected specialty will decide the classes that will determine the student's area of expertise. It is dependent on many subjects in order to assess the best option of specialisation. For three specialized fields, 12 topics form the foundation of expertise, where the four subjects are each focused. This does not, of course, entail reducing the expense of other courses to be offered to pupils.

Research in the selection of concentration has also been carried out by [1] on the Yogyakarta Amikom International Class Students using the K-Means method which produces suitable concentrations based on the scores of several subjects. In contrast to the current research, determining the suitable concentration is determined from the results of the cluster produced by the K-Means method.

In another study [2] conducted a study to group the Final Semester Student Concentration Course Classes at Ichsan Gorontalo University using the K-Means and KMeans KNN methods. This study produces clusters to classify Concentration Course Classes for final semester students and each of these clusters has a predictive value for the two clusters.

Grouping and prediction data with the K-Means method and K-NN has also been done by [3] to produce a decision-making tool giving rewards to the employees of the UPI Convention Group, [4] in the Determination Regional Priority Services Birth Certificate Bogor, [5] in his research Implementation method K-Nn To Predict Results Agriculture in Malang, [6] on image Segmentation in wheezing the identification of diseases and diseases, [7] on the Acceptance of New Technology-minded Teacher Candidates and Administrative Employees, and [8] on Sentiment Analysis for Travel Agent Reviews.

In making this specialization selection, students often hesitate in making decisions because of their ignorance and the influence of other friends so that it is often not following their abilities. This will also have an impact on the quality of graduates from the study program.

This study aims to provide recommendations to students in determining their specialization based on the course scores they have obtained in the previous semester. With this research, it is expected that the chosen students will choose according to their abilities so that it will improve the quality of graduates.

II. Methods

Several phases of activity are carried out in this research. The operation starts from the collection of data. The data collected would be pre-processed in order to acquire training data and test data for analysis. Via the grouping process, training data can be performed. The classification is performed using the system K-Means [9][10]. The results of the classification yield clusters of student data that are recommended to identify and not recommended specializations. This is achieved on the basis of the previous values which later form the basis of the course. In order to generate a recommended speciality for students, the recommended data are analyzed using the K-Nearest neighbor system [11][12][13][14][15] and test data. Figure 1 shows the mechanism for this processing of data.

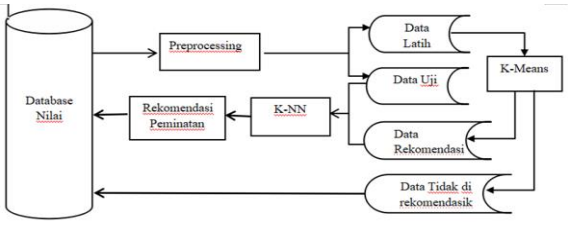


Figure 1 Data Processing Framework

The framework for the above data processing can be explained as follows:

A. Preprocessing

This is the first step taken before processing the data mining process. It is done to eliminate distractions that exist in the data [16][17]. Poor data can be caused by several things, namely:

1. Students have not taken all the subjects that are the basis for determining specialization.
2. Some scores fail either one subject or more than the basic subject of specialization (letter E-grade)
3. Subject data taken are basic courses of specialization in semesters one to four.

B. Training Data

Training data is pre-existing data based on facts that have already occurred. In this study, the training data were taken from the student data of each specialization with the value of the basic subject being the maximum value.

C. Test Data

Test data is data that is already valuable which is used to calculate the accuracy of the formed classification model. In this study, the test data were taken from the results of the student data grouping process that had been carried out using the K-Means method.

D. K-Means

The K-Means algorithm is an iterative grouping of data sets into some predefined clusters [10]. The K-Means method is data grouping by maximizing the similarity of data in one cluster and minimizing the similarity of data between clusters [18]. The measure of similarity used in clusters is a function of distance. So that maximization of data similarity is obtained based on the shortest distance between the data and the centroid point [19]. The K-Means method is a method used to classify student data to obtain the recommended data for selecting specializations and those that have not. The use of the K-Means method is to group data into some clusters by comparing the distance between the data and the centroid [20][21]. Algoritma *K-Means* is not influenced by the order of objects. An important step in using the K-means method is to determine

the centroid, the number of clusters, and the distance of the centroid. By forming several clusters using the K-Means algorithm, you can also find out the distance between the central cluster (centroid) on the data to be analyzedIn this study, the determination of the initial value of the centroid was carried out randomly [1][22]. To determine the distance between the data and the centroid, the following formula is used:

$$d_{ij} = \sqrt{\sum_{k=1}^{p} (X_{ik} - X_{Jk})^2}$$

The process will be carried out until there is no change in the distance between the data and the cluster center. The results of this data grouping process will be used in the next process.

E. Recommended Data

Recommendation data is data generated from the data grouping process using the K-Means method. The recommended data is data that is complete and has taken all the basic courses of the specialization.

F. Data Not Recommended

Data not recommended is data resulting from the grouping process that has not met the criteria. This could be since there are still some scores that are obtained below the passing standards, especially those for the basic subjects of specialization.

G. K-Nearest Neighbor (K-NN)

K-NN is an algorithm that classifies new objects based on attributes and training samples [23][24]. KNN is a non-parametric lazy learning algorithm because it does not make any assumptions on the distribution of the main data [25][26]. This algorithm works based on the shortest distance from the query instance to the training data [27][28][29]. This algorithm uses a supervised algorithm where the results of the new test samples are grouped based on the majority of the categories on the K-NN [30]. The results of the grouping process carried out in the previous process will be processed with the K Nearest Neighboard method to predict what suitable specializations are taken by a student by looking at the similarity value generated using the following formula [3] :

$$Similarity\ (T.S) = \frac{\sum_{i=1}^{n} f(Ti.Si)\,x\,Wi}{Wi}$$

Where T is the new case, S is the storage case, n is the number of attributes in each case, i is the individual attribute from 1 to n and wi is the attribute feature weight. In this process, the attributes to be used will be determined. In this case, there are 12 attributes used which are taken from the courses which are the basis for each specialization.

H. Specialization Recommendations

The specialization recommendation is the result of a process that has been carried out using the K-NN method. The recommended specialization is that which has the highest similarity value obtained from the closeness of the value between the learning data and the test data

III. Result and Discussion

This study uses data derived from student scores for the Information Systems Study Program at Putra Indonesia University 'YPTK' Padang. The Information Systems Study Program has three specializations that will determine the area of expertise of the graduate. In determining this specialization, students often experience doubts, resulting in students determining their specialization that is not according to their abilities, this can be seen from the scores they obtained in the previous semester. Errors in the selection of specialization will greatly affect the quality of graduates. For this reason, it is necessary to group students who will determine their specialization. The K-Means method will be used to group students into two groups. The first group are students who are recommended to determine specialization and the second group are students who have not been recommended to determine specialization. The process of grouping is carried out by considering several subjects which are the basis for each specialization. There are 12 subjects and

3 specializations that have been assigned to be used in this classification process as shown in Table 1 and Table 2.

Table 1 List Of Subjection Basic Subjection

No.	Course Name	Code
1.	Introduction to Management	pmnj
2.	Introduction to Information Technology	pti
3.	General Organization Theory	tou
4.	Management Information System	sim
5.	Information Systems Concept	ksi
6.	Business Knowledge	pb
7.	Accounting information system	drain
8.	Information System Design Analysis 1	apsi1
9.	Algorithms and Data Structures 1	alg1
10.	Algorithms and Data Structures 2	alg2
11.	Database System	sbd
12.	Database Design	pbd

Table 2 Specialization

No.	Name of Specialization
1.	Management Information System
2.	Business Information Systems
3.	Industrial Information Systems

The sample data of the Information System students at the Faculty of Computer Science, Putra Indonesia University YPTK Padang used can be seen in table 3.

Table 3 One, Two, Three, And Four Semester Student Scores

Student name	Information Technology Management				Business Information Systems				Industrial Information Systems			
	pmnj	pti	tou	sim	ksi	pb	drain	apsi1	alg1	alg2	sbd	pbd
Abidzar Ghifari Zandra	A	A	B	A	A	B	A	A	A	A	A	A
Afdila Zartika	A	B	A	B	A	A	A	A	B	A	A	A
Aisyah Nurminas	B	B	B	B	A	A	A	A	B	B	B	A
Amhar Al unawar	B	B	C	C	B	A	C	D	C	D	C	C
Angga Agustiadi	B	B	B	B	B	B	B	C	A	C	B	C
Ari Suhanda	D	B	C	B	C	C	C	C	A	A	A	A
Ayu Winanda	B	B	B	B	B	B	B	B	B	C	B	B
Deviani	A	B	B	B	A	A	B	B	A	B	B	A
Dino Febrian Doni	B	B	B	C	A	B	B	D	B	B	B	D
Edi Susilo	B	A	B	B	B	A	B	B	B	C	B	A
Elri Suhendra	B	B	B	B	B	B	C	D	B	D	C	B
Fadila Erina	B	B	A	A	A	A	B	B	B	C	B	B
Heru Pramana Firnu	C	C	B	B	C	B	C	B	A	C	B	C
Inspiration of the Son of Gifts	D	C	C	B	C	B	C	C	B	D	D	D
Indri Yani Putri	B	A	A	A	A	A	A	A	A	B	B	A
Kelvin Frendinata	C	C	B	B	C	A	C	C	C	D	B	A
Leonida Cipta Meidayanti	A	B	B	B	A	A	B	B	A	B	B	A
Malfo Dewo	A	A	A	A	A	A	A	A	A	A	A	A
Mohd. Martha. M	B	D	B	B	B	B	C	D	B	B	B	D
Muhammad Fadel	B	B	B	B	B	B	C	C	B	C	B	B
Muhammad Reki Andika	C	B	B	B	B	C	B	D	C	D	B	C
Nada Permata Sari	B	B	B	B	B	A	B	B	B	D	B	B
Nadya Dwi Yasra	B	B	B	B	A	A	A	B	B	A	B	B
Nur Farahana	B	B	B	A	B	A	A	B	B	B	B	B
Nurul Indah Azizah	A	A	A	A	A	A	A	B	A	B	A	A
Panji Patikawa	B	B	B	B	C	B	A	C	A	B	B	B

Qaidah Pratari Rahmatullah	A	A	A	A	B	A	B	B	B	B	B	A
Randi Sulaeman	B	B	B	B	A	A	B	A	B	A	C	A
Renza Nazirah Faiqah	B	B	A	B	A	A	B	A	B	B	B	B
Rika Ridla Juita	B	B	B	B	B	B	B	B	B	B	B	B
Ririn Uswatun H	B	A	B	B	A	A	B	B	B	B	B	A
Rizki Permana Putra	D	C	D	B	C	A	D	D	D	C	B	C
Rudi Greetings	B	C	C	B	B	B	D	D	B	C	C	D
Shinta Amelia Ananda	B	B	B	A	A	A	B	D	C	B	B	A
Sukri Azhari	C	B	B	B	C	B	B	D	C	C	C	C
Syamsuri Sani	B	B	C	B	B	A	D	D	B	C	C	C
Teguh Arisandi	C	B	B	B	C	B	B	B	B	D	C	C
Thia Pratama Tanjung	B	B	B	D	B	A	D	D	D	C	C	D
Tri Wanda Yasman	B	B	B	B	B	A	C	D	B	D	B	C
Viki Pratama	C	B	B	B	C	C	C	D	B	C	B	B
Wahyu Ferdiant	B	B	B	B	B	C	B	B	A	C	D	B
Windika Pilyadi	B	A	B	B	A	A	C	D	A	A	C	D
Yogi Fajr Bahari	B	A	B	A	A	A	A	A	B	B	A	B

The process of grouping values with K-Means is a process of numerical values. So that the data above needs a value transformation process so that all natural values are in the form of numbers. The transformation is made with the values A = 4, B = 3, C = 2 and D = 1. The results of the value transformation to be processed can be seen in table 4

Table 4 .Transformation Of Value For One, Two, Three, And Four Semester Students

Student name	**Information Technology Management**				**Business Information Systems**				**Industrial Information Systems**			
	pmnj	**pti**	**tou**	**sim**	**ksi**	**pb**	**drain**	**apsi1**	**alg1**	**alg2**	**sbd**	**Pbd**
Abidzar Ghifari Zandra	4	4	3	4	4	3	4	4	4	4	4	4
Afdila Zartika	4	3	4	3	4	4	4	4	3	4	4	4
Aisyah Nurminas	3	3	3	3	4	4	4	4	3	3	3	4
Amhar Al Munawar	3	3	2	2	3	4	2	1	2	1	2	2
Angga Agustiadi	3	3	3	3	3	3	3	2	4	2	3	2
Ari Suhanda	1	3	2	3	2	2	2	2	4	4	4	4
Ayu Winanda	3	3	3	3	3	3	3	3	3	2	3	3
Deviani	4	3	3	3	4	4	3	3	4	3	3	4
Dino Febrian Doni	3	3	3	2	4	3	3	1	3	3	3	1
Edi Susilo	3	4	3	3	3	4	3	3	3	2	3	4
Elri Suhendra	3	3	3	3	3	3	2	1	3	1	2	3
Fadila Erina	3	3	4	4	4	4	3	3	3	2	3	3
Heru Pramana Firnu	2	2	3	3	2	3	2	3	4	2	3	2
Inspiration of the Son of Gifts	1	2	2	3	2	3	2	2	3	1	1	1
Indri Yani Putri	3	4	4	4	4	4	4	4	4	3	3	4
Kelvin Frendinata	2	2	3	3	2	4	2	2	2	1	3	4
Leonida Cipta Meidayanti	4	3	3	3	4	4	3	3	4	3	3	4
Malfo Dewo	4	4	4	4	4	4	4	4	4	4	4	4
Mohd. Martha. M	3	1	3	3	3	3	2	1	3	3	3	1
Muhammad Fadel	3	3	3	3	3	3	2	2	3	2	3	3
Muhammad Reki Andika	2	3	3	3	3	2	3	1	2	1	3	2
Nada Permata Sari	3	3	3	3	3	4	3	3	3	1	3	3
Nadya Dwi Yasra	3	3	3	3	4	4	4	3	3	4	3	3
Nur Farahana	3	3	3	4	3	4	4	3	3	3	3	3
Nurul Indah Azizah	4	4	4	4	4	4	4	3	4	3	4	4

Panji Patikawa	3	3	3	3	2	3	4	2	4	3	3	3
Qaidah Pratari Rahmatullah	4	4	4	4	3	4	3	3	3	3	3	4
Randi Sulaeman	3	3	3	3	4	4	3	4	3	4	2	4
Renza Nazirah Faiqah	3	3	4	3	4	4	3	4	3	3	3	3
Rika Ridla Juita	3	3	3	3	3	3	3	3	3	3	3	3
Ririn Uswatun Hasanah	3	4	3	3	4	4	3	3	3	3	3	4
Rizki Permana Putra	1	2	1	3	2	4	1	1	1	2	3	2
Rudi Greetings	3	2	2	3	3	3	1	1	3	2	2	1
Shinta Amelia Ananda	3	3	3	4	4	4	3	1	2	3	3	4
Sukri Azhari	2	3	3	3	2	3	3	1	2	2	2	2
Syamsuri Sani	3	3	2	3	3	4	1	1	3	2	2	2
Teguh Arisandi	2	3	3	3	2	3	3	3	3	1	2	2
Thia Pratama Tanjung	3	3	3	1	3	4	1	1	1	2	2	1
Tri Wanda Yasman	3	3	3	3	3	4	2	1	3	1	3	2
Viki Pratama	2	3	3	3	2	2	2	1	3	2	3	3
Wahyu Ferdiant	3	3	3	3	3	2	3	3	4	2	1	3
Windika Pilyadi	3	4	3	3	4	4	2	1	4	4	2	1
Yogi Fajr Bahari	3	4	3	4	4	4	4	4	3	3	4	3

The first step is to determine the data centroid. Centroid is determined randomly. In this process, 2 centroids are selected as shown in Table 5 below:

Table 5 Centroid Table

Centroid	Student name	pmnj	pti	tou	sim	ksi	Pb	drain	apsi1	alg1	alg2	sbd	pbd
C1	NADYA DWI YASRA	B	B	B	B	A	A	A	B	B	A	B	B
C2	RIZKI PERMANA PUTRA	D	C	D	B	C	A	D	D	D	C	B	C

The value of the centroid is transformed according to the predefined conditions. The results of the transformation can be seen in table 6. below.

Table 6 Centroid Transformation

Centroid	Student name	pmnj	pti	tou	sim	ksi	Pb	drain	apsi1	alg1	alg2	sbd	pbd
C1	Nadya Dwi Yasra	3	3	3	3	4	4	4	3	3	4	3	3
C2	Rizki Permana Putra	1	2	1	3	2	4	1	1	1	2	3	2

The second step is to determine the distance between the data and its centroid values. Table 7. The following is the result of the process according to the formula mentioned in the previous chapter.

Table 7 Data Distance With Centroid Value

Student name	Mnj Information Technology				Business Information Systems				Industrial Information Systems				Distance To		Score
	pmnj	pti	tou	sim	ksi	pb	drain	apsi1	alg1	alg2	sbd	pbd	C1	C2	Cluster
Abidzar Ghifari Zandra	4	4	3	4	4	3	4	4	4	4	4	4	2.83	7.68	C1
Afdila Zartika	4	3	4	3	4	4	4	4	3	4	4	4	2.24	7.35	C1
Aisyah Nurminas	3	3	3	3	4	4	4	4	3	3	3	4	1.73	6.32	C1
Amhar Al Munawar	3	3	2	2	3	4	2	1	2	1	2	2	4.80	3.46	C2
Angga Agustiadi	3	3	3	3	3	3	3	2	4	2	3	2	3.16	5.00	C1
Ari Suhanda	1	3	2	3	2	2	2	2	4	4	4	4	4.58	5.10	C1
Ayu Winanda	3	3	3	3	3	3	3	3	3	2	3	3	2.65	4.90	C1
Deviani	4	3	3	3	4	4	3	3	4	3	3	4	2.24	6.32	C1

Dino Febrian Doni	3	3	3	2	4	3	3	1	3	3	3	1	3.46	5.00	C1
Edi Susilo	3	4	3	3	3	4	3	3	3	2	3	4	2.83	5.39	C1
Elri Suhendra	3	3	3	3	3	3	2	1	3	1	2	3	4.47	4.36	C2
Fadila Erina	3	3	4	4	4	4	3	3	3	2	3	3	2.65	5.66	C1
Heru Pramana Firnu	2	2	3	3	2	3	2	3	4	2	3	2	4.12	4.47	C1
Inspiration of the Son of Gifts	1	2	2	3	2	3	2	2	3	1	1	1	5.74	3.74	C2
Indri Yani Putri	3	4	4	4	4	4	4	4	4	3	3	4	2.65	7.35	C1
Kelvin Frendinata	2	2	3	3	2	4	2	2	2	1	3	4	4.69	3.61	C2
Leonida Cipta Meidayanti	4	3	3	3	4	4	3	3	4	3	3	4	2.24	6.32	C1
Malfo Dewo	4	4	4	4	4	4	4	4	4	4	4	4	2.83	7.94	C1
Mohd. Martha. M	3	1	3	3	3	3	2	1	3	3	3	1	4.36	4.24	C2
Muhammad Fadel	3	3	3	3	3	3	2	2	3	2	3	3	3.32	4.24	C1
Muhammad Reki Andika	2	3	3	3	3	2	3	1	2	1	3	2	4.69	4.12	C2
Nada Permata Sari	3	3	3	3	3	4	3	3	3	1	3	3	3.32	4.90	C1
Nadya Dwi Yasra	3	3	3	3	4	4	4	3	3	4	3	3	0.00	5.92	C1
Nur Farahana	3	3	3	4	3	4	4	3	3	3	3	3	1.73	5.48	C1
Nurul Indah Azizah	4	4	4	4	4	4	4	3	4	3	4	4	2.83	7.42	C1
Panji Patikawa	3	3	3	3	2	3	4	2	4	3	3	3	2.83	5.57	C1
Qaidah Pratari Rahmatullah	4	4	4	4	3	4	3	3	3	3	3	4	2.83	6.40	C1
Randi Sulaeman	3	3	3	3	4	4	3	4	3	4	2	4	2.00	6.24	C1
Renza Nazirah Faiqah	3	3	4	3	4	4	3	4	3	3	3	3	2.00	6.08	C1
Rika Ridla Juita	3	3	3	3	3	3	3	3	3	3	3	3	2.00	5.00	C1
Ririn Uswatun Hasanah	3	4	3	3	4	4	3	3	3	3	3	4	2.00	5.74	C1
Rizki Permana Putra	1	2	1	3	2	4	1	1	1	2	3	2	5.92	0.00	C2
Rudi Greetings	3	2	2	3	3	3	1	1	3	2	2	1	5.10	3.61	C2
Shinta Amelia Ananda	3	3	3	4	4	4	3	1	2	3	3	4	3.00	4.90	C1
Sukri Azhari	2	3	3	3	2	3	3	1	2	2	2	2	4.24	3.61	C2
Syamsuri Sani	3	3	2	3	3	4	1	1	3	2	2	2	4.58	3.46	C2
Teguh Arisandi	2	3	3	3	2	3	3	3	3	1	2	2	4.24	4.58	C1
Thia Pratama Tanjung	3	3	3	1	3	4	1	1	1	2	2	1	5.57	4.00	C2
Tri Wanda Yasman	3	3	3	3	3	4	2	1	3	1	3	2	4.36	4.00	C2
Viki Pratama	2	3	3	3	2	2	2	1	3	2	3	3	4.58	4.00	C2
Wahyu Ferdiant	3	3	3	3	3	2	3	3	4	2	1	3	3.87	6.00	C1
Windika Pilyadi	3	4	3	3	4	4	2	1	4	4	2	1	3.87	5.66	C1
Yogi Fajr Bahari	3	4	3	4	4	4	4	4	3	3	4	3	2.24	6.48	C1

The third step repeats the process by determining the new centroid value by calculating the centroid of each cluster for each attribute until the position of the cluster does not change.

The new centroid values can be seen in table 8 below:

Table 8 Iteration Centroid Value 2.

Cluster 1	3.10	3.27	3.20	3.27	3.43	3.57	3.20	2.93	3.37	2.87	3.03	3.23
Cluster 2	2.38	2.54	2.54	2.77	2.62	3.31	1.85	1.15	2.38	1.62	2.38	2.00

Repetition will be carried out until there is no more change in the Cluster value. In this study, there were 4 repetitions of the process in determining the classification of student data.From the results of the process, there were two groups of students, Group C1 was advised to take specialization while group C2 was not recommended. The results of this grouping can be seen in the following tables 9 and 10.

Table 9 Group C1

No.	Student name	Information Technology Management				Business Information Systems				Industrial Information Systems				Distance To		Cluster
		pmnj	pti	tou	sim	ksi	pb	drain	apsi1	alg1	alg2	sbd	pbd	C1	C2	
1	Abidzar Ghifari Zandra	4	4	3	4	4	3	4	4	4	4	4	4	2.38	5.53	C1
2	Afdila Zartika	4	3	4	3	4	4	4	4	3	4	4	4	2.21	5.30	C1
3	Aisyah Nurminas	3	3	3	3	4	4	4	4	3	3	3	4	1.44	4.36	C1
4	Ari Suhanda	1	3	2	3	2	2	2	2	4	4	4	4	4.18	4.25	C1
5	Ayu Winanda	3	3	3	3	3	3	3	3	3	2	3	3	1.65	2.35	C1

6	Deviani	4	3	3	3	4	4	3	3	4	3	3	4	1.44	3.94	C1
7	Edi Susilo	3	4	3	3	3	4	3	3	3	2	3	4	1.60	3.28	C1
8	Fadila Erina	3	3	4	4	4	4	3	3	3	2	3	3	1.75	3.25	C1
9	Indri Yani Putri	3	4	4	4	4	4	4	4	4	3	3	4	1.87	5.02	C1
10	Leonida Cipta Meidayanti	4	3	3	3	4	4	3	3	4	3	3	4	1.44	3.94	C1
11	Malfo Dewo	4	4	4	4	4	4	4	4	4	4	4	4	2.41	5.71	C1
12	Nada Permata Sari	3	3	3	3	3	4	3	3	3	1	3	3	2.32	2.59	C1
13	Nadya Dwi Yasra	3	3	3	3	4	4	4	3	3	4	3	3	1.60	3.88	C1
14	Nur Farahana	3	3	3	4	3	4	4	3	3	3	3	3	1.38	3.41	C1
15	Nurul Indah Azizah	4	4	4	4	4	4	4	3	4	3	4	4	2.04	5.02	C1
16	Panji Patikawa	3	3	3	3	2	3	4	2	4	3	3	3	2.39	3.10	C1
17	Qaidah Pratari Rahmatullah	4	4	4	4	3	4	3	3	3	3	3	4	1.70	4.14	C1
18	Randi Sulaeman	3	3	3	3	4	4	3	4	3	4	2	4	2.02	4.40	C1
19	Renza Nazirah Faiqah	3	3	4	3	4	4	3	4	3	3	3	3	1.58	3.83	C1
20	Rika Ridla Juita	3	3	3	3	3	3	3	3	3	3	3	3	1.32	2.62	C1
21	Ririn Uswatun Hasanah	3	4	3	3	4	4	3	3	3	3	3	4	1.22	3.69	C1
22	Shinta Amelia Ananda	3	3	3	4	4	4	3	1	2	3	3	4	2.78	3.40	C1
23	Wahyu Ferdiant	3	3	3	3	3	2	3	3	4	2	1	3	3.15	3.24	C1
24	Yogi Fajr Bahari	3	4	3	4	4	4	4	4	3	3	4	3	1.89	4.58	C1

Table 10 Group C2

No.	Student name	Information Technology Management				Business Information Systems				Industrial Information Systems				Distance To		Cluster
		pmnj	pti	tou	sim	ksi	pb	Drain	apsi1	alg1	alg2	sbd	pbd	C1	C2	
1	Amhar Al Munawar	3	3	2	2	3	4	2	1	2	1	2	2	4.49	1.93	C2
2	Angga Agustiadi	3	3	3	3	3	3	3	2	4	2	3	2	2.56	1.95	C2
3	Dino Febrian Doni	3	3	3	2	4	3	3	1	3	3	3	1	3.79	2.53	C2
4	Elri Suhendra	3	3	3	3	3	3	2	1	3	1	2	3	3.66	1.72	C2
5	Heru Pramana Firnu	2	2	3	3	2	3	2	3	4	2	3	2	3.49	2.42	C2
6	Inspiration of the Son of Gifts	1	2	2	3	2	3	2	2	3	1	1	1	5.45	2.82	C2
7	Kelvin Frendinata	2	2	3	3	2	4	2	2	2	1	3	4	3.89	2.83	C2
8	Mohd. Martha. M	3	1	3	3	3	3	2	1	3	3	3	1	4.45	2.48	C2
9	Muhammad Fadel	3	3	3	3	3	3	2	2	3	2	3	3	2.41	1.57	C2
10	Muhammad Reki Andika	2	3	3	3	3	2	3	1	2	1	3	2	4.24	2.18	C2
11	Rizki Permana Putra	1	2	1	3	2	4	1	1	1	2	3	2	5.80	3.35	C2
12	Rudi Greetings	3	2	2	3	3	3	1	1	3	2	2	1	4.86	1.97	C2
13	Sukri Azhari	2	3	3	3	2	3	3	1	2	2	2	2	4.00	1.72	C2
14	Syamsuri Sani	3	3	2	3	3	4	1	1	3	2	2	2	4.18	1.76	C2
15	Teguh Arisandi	2	3	3	3	2	3	3	3	3	1	2	2	3.56	2.34	C2
16	Thia Pratama Tanjung	3	3	3	1	3	4	1	1	1	2	2	1	5.57	3.12	C2
17	Tri Wanda Yasman	3	3	3	3	3	4	2	1	3	1	3	2	3.74	1.54	C2
18	Viki Pratama	2	3	3	3	2	2	2	1	3	2	3	3	3.87	2.07	C2
19	Windika Pilyadi	3	4	3	3	4	4	2	1	4	4	2	1	4.12	3.44	C2

The data in the C1 group will be further processed to determine the appropriate specialization for these students. The process will be carried out using the K Nearest Neighbor method to produce a similarity value with the test data used.

In this process, the attributes that will be used will be determined where there are 12 attributes taken from the course which are the basis for each specialization. Each attribute is given a symbol and weighted according to its level of need. The attribute names, symbols, and weighting can be seen in table 11. below:

Table 11 Name Attributes And Symbols

No.	(Symbol) Attribute Name	Weight
1	(A) Introduction to Management	0.9
2	(B) Introduction to Information Technology	0.9
3	(C) General Organization Theory	0.9
4	(D) Management Information Systems	0.9
5	(E) Information Systems Concepts	0.9
6	(F) Business Knowledge	0.9
7	(G) Accounting Information Systems	0.9
8	(H) Analysis of Information System Design	0.9
9	(I) Algorithms and Data Structures 1	0.9
10	(J) Algorithms and Data Structures 2	0.9
11	(K) Database Systems	0.9
12	(L) Database Design	0.9

Each attribute has an attribute value which can be seen in table 12. below:

Table 12 Attributes Value

No.	Attribute	Attribute Value	No.	Attribute	Attribute Value
1	(A) Introduction to Management	(A1) Grade A.	7	(G) Accounting Information Systems	(G1) Value A.
		(A2) Value of B.			(G2) Value of B
		(A3) Value of C			(G3) Value of C
		(A4) Value of D			(G4) Value of D
2	(B) Introduction to Information Technology	(B1) Grade A.	8	(H) Analysis of Information System Design	(H1) Value A
		(B2) Value of B			(H2) Value B
		(B3) Value of C			(H3) Value of C
		(B4) Value of D			(H4) Value of D
3	(C) General Organization Theory	(C1) Grade A.	9	(I) Algorithms and Data Structures 1	(I1) Grade A.
		(C2) Value of B			(I2) Value of B
		(C3) Value of C			(I3) Value of C
		(C4) Value of D			(I4) Value of D
4	(D) Management Information Systems	(D1) Grade A.	10	(J) Algorithms and Data Structures 2	(J1) Grade A.
		(D2) Value of B			(J2) Value of B
		(D3) Value of C			(J3) Value of C
		(D4) Value of D			(J4) Value of D
5	(E) Information Systems Concepts	(E1) Grade A.	11	(K) Database Systems	(K1) Grade A.
		(E2) Value of B			(K2) Value of B
		(E3) Value of C			(K3) Value of C
		(E4) Value of D			(K4) Value of D
	(F) Business Knowledge	(F1) Value A	12	(L) Database Design	(L1) Grade A.
		(F2) Value of B			(L2) Value of B
		(F3) Value of C			(L3) Value of C
		(F4) Value of D			(L4) Value of D

The next process is to calculate the proximity of the attribute values for all the attributes used. Tables 13. and 14. The following are examples of the calculation process in finding the value of the proximity of each Attribute Value.

Table 13 Classification (A) Introduction To Management

No.	Attribute Value	Score	Weight
1	(A1) Grade A.	4	0.9
2	(A2) Value of B.	3	
3	(A3) Value of C	2	
4	(A4) Value of D	1	

Table 14 Approach To Attributes Value (A) Introduction To Management

	A1	A2	A3	A4
A1	1	0.75	0.5	0.25
A2	0.75	1	0.67	0.33
A3	0.5	0.67	1	0.5
A4	0.25	0.33	0.5	1

The next step is to select sample data for training which can be seen in table 15. below:

Table 15 Training Data Sample

No.	Name	A	B	C	D	E	F	G.	H	I	J	K	L	Decision
1	Afdila Zartika	A1	B2	C1	D2	E1	F1	G1	H1	I2	J1	K1	L1	Business Information Systems
2	Nurul Indah Azizah	A1	B1	C1	D1	E1	F1	G1	H2	I1	J2	K1	L1	Information Technology Management
3	Abidzar Ghifari Zandra	A1	B1	C2	D1	E1	F2	G1	H1	I1	J1	K1	L1	Industrial Information Systems
4	Aisyah Nurminas	A2	B2	C2	D2	E1	F1	G1	H1	I2	J2	K2	L1	Business Information Systems
5	Qaidah Pratari Rahmatullah	A1	B1	C1	D1	E2	F1	G2	H2	I2	J2	K2	L1	Information Technology Management
6	Ari Suhanda	A4	B2	C3	D2	E3	F3	G3	H3	I1	J1	K1	L1	Industrial Information Systems

The testing data used can be seen in Table 16 below:

Table 16 Testing Data

No.	Name	A	B	C	D	E	F	G.	H	I	J	K	L	Decision.
7	Leonida Cipta Meidayanti	A1	B2	C2	D2	E1	F1	G2	H2	I1	J2	K2	L1	

The next step is to calculate the closeness of the new case in Table 16. and the old case Table 15. The closeness of the new case to case number 1 in the old case is shown in Table 17 below:

Table 17 New Case Approach With Old Case.

No.	Attribute	Case 1	New Case	Proximity (s)	Weight (w)	S * W
1	A	A1	A1	1	0.9	0.9
2	B	B2	B2	1	0.9	0.9
3	C	C1	C2	0.75	0.9	0.675
4	D	D2	D2	1	0.9	0.9
5	E	E1	E1	1	0.9	0.9
6	F	F1	F1	1	0.9	0.9
7	G.	G1	G2	0.75	0.9	0.675
8	H	H1	H2	0.75	0.9	0.675
9	I	I2	I1	0.75	0.9	0.675
10	J	J1	J2	0.75	0.9	0.675
11	K	K1	K2	0.75	0.9	0.675
12	L	L1	L1	1	0.9	0.9
					10.8	9.45

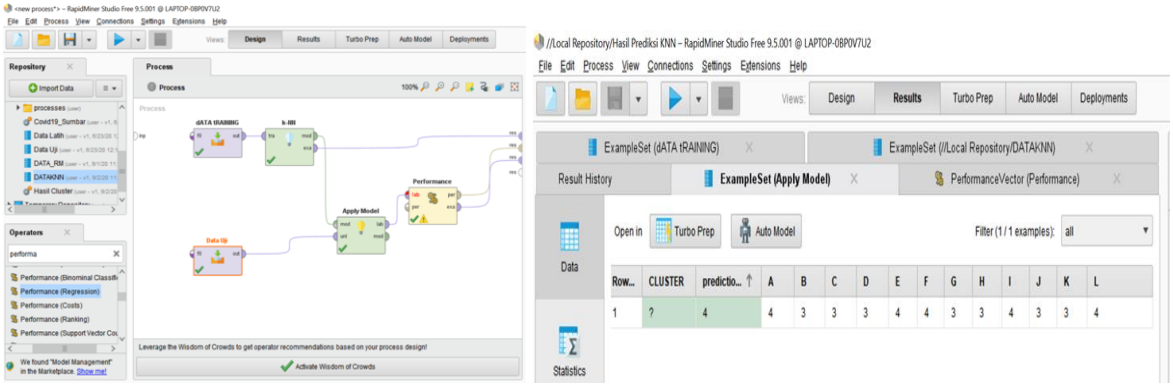
From Table 17, the closeness of the new case to case number 1 can be calculated as follows:

$$Similarity = \frac{0.9 + 0.9 + 0.675 + 0.9 + 0.9 + 0.9 + 0.675 + 0.675 + 0.675 + 0.675 + 0.675 + 0.9}{0.9 + 0.9 + 0.9 + 0.9 + 0.9 + 0.9 + 0.9 + 0.9 + 0.9 + 0.9 + 0.9 + 0.9}$$

$$= 0.875$$

The same process is carried out on all training data so that the highest similarity value is obtained for new cases. The results of the calculation process in this study were tested using data mining applications. The test results prove that the results of the calculation process have the same value as the results of the application.

Prediction With K-Nearest Neighbors



IV. Conclusion

This research was conducted by taking training data and test data from the student data of the Putra Indonesia University Information System 'YPTK' Padang. This study uses 12 attributes taken from existing courses in semesters up to 4. This research has been able to provide recommendations for students who will choose specializations under the basic courses of specialization they have mastered, this can be seen from the results of the course scores obtained. From the results of tests carried out using Rapid Miner, the calculation results of the methods used have similarities. This proves that the K-Means and K-NN methods can be used in determining student interest. This research can also be developed using other methods to get maximum results. Research can also be carried out to determine the direction of student research based on the grades they got in previous semesters

References

[1] J. Aranda and W. A. G. Natasya, "Penerapan Metode K-Means Cluster Analysis Pada Sistem Pendukung Keputusan Pemilihan Konsentrasi Untuk Mahasiswa International Class Stmik Amikom Yogyakarta," *Semnasteknomedia Online*, vol. 4, no. 1, pp. 4-2–1, 2016, [Online]. Available: https://ojs.amikom.ac.id/index.php/semnasteknomedia/article/view/1293.

[2] S. Rustam and H. Annur, "Akademik Data Mining (Adm) K-Means Dan K-Means K-Nn Untuk Mengelompokan Kelas Mata Kuliah Kosentrasi Mahasiswa Semester Akhir," *Ilk. J. Ilm.*, vol. 11, no. 3, pp. 260–268, 2019, doi: 10.33096/ilkom.v11i3.487.260-268.

[3] E. Praja, W. Mandala, M. Ridwan, and E. Putri, "DATA MINING PEMBERIAN REWARD PADA KARYAWAN UPI," pp. 37–44, 2019.

[4] A. M. M. Anwar, P. Harsani, and A. Maesya, "Penentuan Daerah Prioritas Pelayanan Akta Kelahiran Dengan Metode K-Nn Dan K-Means," *Komputasi J. Ilm. Ilmu Komput. dan Mat.*, vol. 17, no. 1, pp. 319–328, 2020, doi: 10.33751/komputasi.v17i1.1884.

[5] A. Pungky, "Penerapan metode k-nn untuk memprediksi hasil pertanian di kabupaten malang," vol. 3, no. 1, pp. 235–242, 2019.

[6] F. G. Febrinanto, C. Dewi, and A. T. Wiratno, "Implementasi Algoritme K-Means Sebagai Metode Segmentasi Citra Dalam Identifikasi Penyakit Daun Jeruk," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 11, pp. 5375–5383, 2018.

[7] N. N. Dzikrulloh 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 (Studi Kasus : Sekolah Menengah Kejuruan Muhammadiyah 2 Kediri)," *Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 5. pp. 378–385, 2017.

[8] S. Ernawati and R. Wati, "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel," *J. Khatulistiwa Inform.*, vol. VI, no. 1, pp. 64–69, 2018, [Online]. Available: https://ejournal.bsi.ac.id/ejurnal/index.php/khatulistiwa/article/view/3802/2626.

[9] I. Kamila, U. Khairunnisa, and M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, p. 119, 2019, doi: 10.24014/rmsi.v5i1.7381.

[10] F. Nur, M. Zarlis, and B. B. Nasution, "Penerapan Algoritma K-Means Pada Siswa Baru Sekolahmenengah Kejuruan Untuk Clustering Jurusan," *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 1, no. 2, pp. 100–105, 2017, doi: 10.30743/infotekjar.v1i2.70.

[11] C. Paramita, E. Hari Rachmawanto, C. Atika Sari, and D. R. Ignatius Moses Setiadi, "Klasifikasi Jeruk Nipis Terhadap Tingkat Kematangan Buah Berdasarkan Fitur Warna Menggunakan K-Nearest Neighbor," *J. Inform. J. Pengemb. IT*, vol. 4, no. 1, pp. 1–6, 2019, doi: 10.30591/jpit.v4i1.1267.

[12] R. Enggar Pawening, W. Ja, and far Shudiq, "KLASIFIKASI KUALITAS JERUK LOKAL BERDASARKAN TEKSTUR DAN BENTUK MENGGUNAKAN METODE k-NEAREST NEIGHBOR (k-NN)," *Ejournal.Unuja.Ac.Id*, vol. 1, no. 1, pp. 10–17, 2020, [Online]. Available: http://ejournal.unuja.ac.id/index.php/core.

[13] A. Budiyantara, I. Irwansyah, E. Prengki, P. A. Pratama, and N. Wiliani, "Komparasi Algoritma Decision Tree, Naive Bayes Dan K-Nearest Neighbor Untuk Memprediksi Mahasiswa Lulus Tepat Waktu," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 5, no. 2, pp. 265–270, 2020, doi: 10.33480/jitk.v5i2.1214.

[14] M. A. Rahman, N. Hidayat, and A. Afif Supianto, "Komparasi Metode Data Mining K-Nearest Neighbor Dengan Naïve Bayes Untuk Klasifikasi Kualitas Air Bersih (Studi Kasus PDAM Tirta Kencana Kabupaten Jombang)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (JPTIIK)*, vol. 2, no. 12. pp. 925–928, 2018.

[15] M. A. Maricar and Dian Pramana, "Perbandingan Akurasi Naïve Bayes dan K-Nearest Neighbor pada Klasifikasi untuk Meramalkan Status Pekerjaan Alumni ITB STIKOM Bali," *J. Sist. dan Inform.*, vol. 14, no. 1, pp. 16–22, 2019, doi: 10.30864/jsi.v14i1.233.

[16] R. P. Fitrianti, "Analisis Sentimen terhadap Review Restoran dengan Teks Bahasa Indonesia Menggunakan Algoritma K-Nearest Neighbor," pp. 27–32, 2018.

[17] Yusra, D. Olivita, and Y. Vitriani, "Perbandingan Klasifikasi Tugas Akhir Mahasiswa Jurusan Teknik Informatika Menggunakan Metode Naïve Bayes Classifier dan K-Nearest Neighbor," *J. Sains, Teknol. dan Ind.*, vol. 14, no. 1, pp. 79–85, 2016.

[18] K. U. Syaliman, M. Zulfahmi, and A. A. Nababan, "Perbandingan Rapid Centroid Estimation (RCE) — K Nearest Neighbor (K-NN) Dengan K Means — K Nearest Neighbor (K-NN)," *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 2, no. 1, pp. 79–89, 2017, doi: 10.30743/infotekjar.v2i1.166.

[19] R. A. Asroni, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *Ilm. Semesta Tek.*, vol. 18, no. 1, pp. 76–82, 2015, doi: 10.1038/hdy.2009.180.

[20] W. Dhuhita, "Clustering Menggunakan Metode K-Mean Untuk Menentukan Status Gizi Balita," *J. Inform. Darmajaya*, vol. 15, no. 2, pp. 160–174, 2015.

[21] L. Rusdiana and T. Informatika, "Perbandingan Metode K-Nearest Neighbor Dan Fuzzy C-Means Dalam Menentukan Predikat Kelulusan Mahasiswa," vol. 1, no. 1, pp. 21–26, 2017.

[22] K. Kunci, "KOMPARASI METODE KLASIFIKASI DATA MINING UNTUK PREDIKSI Abstrak," vol. 5, no. 1, pp. 23–29, 2019.

[23] C. A. Rahardja, T. Juardi, and H. Agung, "Implementasi Algoritma K-Nearest Neighbor Pada Website Rekomendasi Laptop," *J. Buana Inform.*, vol. 10, no. 1, p. 75, 2019, doi: 10.24002/jbi.v10i1.1847.

[24] A. Sulistiyo, "Penentuan Jurusan Sekolah Menengah Atas Menggunakan Metode K-Nearest Neighbor Classifier Pada SMAN 16 Semarang," *Fasilkom Udinus*, vol. 1, no. 1, pp. 1–5, 2014.

[25] W. T. Panjaitan, "Penerapan Algoritma Knn Pada Prediksi Produksi," *J. Univ. AMIKOM Yogyakarta*, pp. 61–66, 2018.

[26] G. A. Pradnyana and A. A. J. Permana, "Sistem Pembagian Kelas Kuliah Mahasiswa Dengan Metode K-Means Dan K-Nearest Neighbors Untuk Meningkatkan Kualitas Pembelajaran," *JUTI J. Ilm. Teknol. Inf.*, vol. 16, no. 1, p. 59, 2018, doi: 10.12962/j24068535.v16i1.a696.

[27] D. Z. Abidin, S. Nurmaini, and R. F. Malik, "Penerapan Metode K-Nearest Neighbor dalam Memprediksi Masa Studi Mahasiswa (Studi Kasus : Mahasiswa STIKOM Dinamika Bangsa)," *Pros. Annu. Res. Semin.*, vol. 3, no. 1, pp. 133–138, 2017.

[28] H. Risman, D. Nugroho, and Y. Retno, "Penerapan Metode K-Nearest Neighbor Pada Aplikasi Penentu Penerima Beasiswa Mahasiswa Di Stmik Sinar Nusantara Surakarta," *TIKomSiN*, pp. 19–25, 2012.

[29] Luh Gede Pivin Suwirmayanti, "Penerapan Metode K-Nearest Neighbor Untuk Sistem Rekomendasi Pemilihan Mobil Implementation of K-Nearest Neighbor Method for Car Selection Recommendation System," *Techno.COM*, vol. 16, no. 2, pp. 120–131, 2017.

[30] Y. Yahya and W. Puspita Hidayanti, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada 'Lombok Vape On,'" *Infotek J. Inform. dan Teknol.*, vol. 3, no. 2, pp. 104–114, 2020, doi: 10.29408/jit.v3i2.2279.