Dharmawangsa: International Journals of The Social Science, Education and Humanities.
P - ISSN : 2716-5132, E - ISSN : 2723-0783 Vol. 3 No. 3 November 2022 Page: 25 – 34

ANALYSIS OF K-MEANS ALGORITHM FOR RECOMMENDATIONS STUDENT CAREER DETERMINATION

Cut Fadhilah [1],Nunsina [2], Zakial Viki[3]
Information Systems, Indonesian National Islamic Faculty, Indonesian National Islamic University[1]
Informatics, Indonesian National Islamic Faculty, Indonesian National Islamic University[2, 3]
Corresponding: cutfadhilahzakaria@gmail.com[1], nyak.nunun@gmail.com[2], zakialviki.mkom@gmail.com[3]

ABSTRACT

Career is a person's progress in a job that is obtained through training or work experience during his life. The stages in a career start from knowing the type of job you are interested in based on your expertise, so there is a reference for finding the job you want. After knowing the job you want, the next step is to stay focused and deepen your skills in that field, so you can master the job you're looking for. Based on these stages, a system is needed that can recommend careers that can assist students in determining careers that match their potential based on their academic grades. In this study, the K-Means algorithm was used to analyze the problem. This study designed a k-means algorithm analysis system for career suitability recommendations for web-based students using HTML, PHP, CSS and XAMPP programming languages. The method used in this study is the Unified Modeling Language (UML) method. This research is able to provide career recommendations for students using the k-means clustering algorithm for three types of careers, namely Web Engineer, Programmer and Software Engineering. This study produces an accuracy rate of 96.6% with manual calculations with results in the system This research is able to provide career recommendations for students using the k-means clustering algorithm for three types of careers, namely Web Engineer, Programmer and Software Engineering. This study produces an accuracy rate of 96.6% with manual calculations with results in the system This research is able to provide career recommendations for students using the k-means clustering algorithm for three types of careers, namely Web Engineer, Programmer and Software Engineering. This study produces an accuracy rate of 96.6% with manual calculations with results in the system
Keywords: Career, K-means, Recommendations.

I. INTRODUCTION

Career is a person's progress in a job that is obtained through training or work experience during his life. Everyone's dream is to work or have a career that suits their wishes and interests, because that way a person can channel his full interest in the job. Everyone definitely wants a good and successful career coupled with the rapid advancement of technology that allows someone to be more competitive in the field of work. Apart from relying on digital programs for professional service providers, self-development also needs to be done.

The stages in a career start from knowing the type of job you are interested in based on your expertise, so there is a reference for finding the job you want. After knowing the job you want, the next step is to stay focused and deepen your skills in that field, so you can master the job you want.

A final student who is still confused about choosing a place to work or a career that is in accordance with the expertise he has definitely wants a place to work that is also suitable for that. There are many incidents and observations that occur in the world of work, many scholars and graduates who are not suitable for working in an agency or company that does not match their

potential, talent or do not match the diploma obtained. Based on information quoted from the "Institute for Development of Economics and Finance (INDEF) by Bisnis.com, there are 60.62% of people who work not in accordance with their educational background and skills".

According to Maulida in his book Psychology, an introduction from an Islamic perspective, he explains that interest is a tendency to pay attention to and act on people, activities or situations that are the object of that interest accompanied by feelings or joy.

Therefore, to minimize this happening, career planning is needed in order to reduce anxiety in finding job information and making decisions about the desired career. One of the career planning references that is carried out is to make an analysis to recommend a list of careers that are in accordance with one's potential, educational background, talents and achievements. Careers that are not in accordance with expertise can hinder a person from obtaining the desired career path because it can cause a feeling of discomfort at work and can even lead to an attitude that is not serious at work which eventually resigns or even gets fired by the agency/company (Nurilla, 2017).

Career selection can be determined from several factors, namely in accordance with the expertise possessed, company achievements, office/agency location, experience provided and salary considerations. Based on these factors, a system is needed that can recommend careers that can assist students in determining careers according to their potential. In this study, the K-Means algorithm was used to analyze the problem.

II. LITERATURE REVIEW

A. K-Means Algorithm

K-Means Algorithm K-Means is a clustering algorithm or data recommendation that is Unsupervised Learning, which means that the input from this algorithm accepts data without class labels. The function of this algorithm is to group data into several clusters or classes (Ketutrare, 2018 & Mulyati, 2019). The k-means cluster method attempts to group existing data into several groups, where data in one group have the same characteristics as each other and have different characteristics from data in other groups (Situmorang, 2020).

The characteristics of this algorithm are:

1. Has n pieces of data
2. Input in the form of the amount of data and the number of clusters (groups)
3. In each cluster / group has a centroid that represents the cluster.

In simple terms the K-Means algorithm starts from the following stages:

1. Choose K centroid points.
2. Calculating data distance with centroid.
3. Update the centroid point value.
4. Repeat steps 2 and 3 until the value of the centroid point no longer changes.

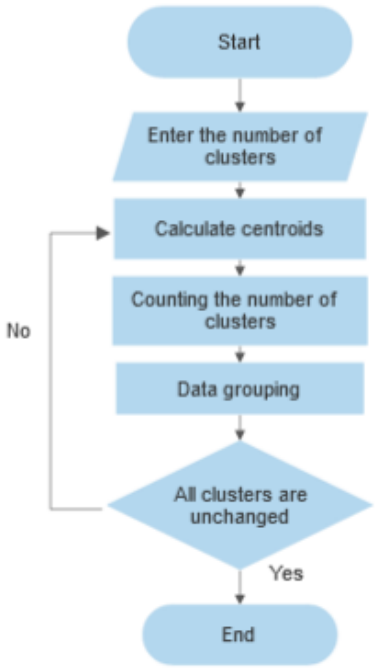

Figure 1. K-Means Workflow

1. Determine k (the value is independent) as the number of clusters you want to form.
2. Generate random values for the initial cluster center (centroid) as much as k.
3. Calculating the distance of each input data to each centroid using the Euclidean Distance formula until the closest distance of each data to the centroid is found. The following is the Euclidian Distance equation:

$$\boldsymbol{d(xi, \mu j) = \sqrt{\sum(xi - \mu j)^2}} \qquad (1)$$

4. Classify each data based on its proximity to the centroid (smallest distance).
5. Updating the value The new centroid value is obtained from the average cluster in question using the formula:

$$\boldsymbol{\mu j\ (t+1) = \frac{1}{Nsj}\sum_{j \in sj} xj} \qquad \textbf{(2)}$$

Repeat from step 3 to 5, until nothing changes in the members of each cluster.

B. Data Classification

Classification is processing to find a model or function that describes and characterizes a concept or class of data, for a particular purpose. Group analysis as a method for classifying data into several groups using the association size measurement method, so that the same data is in one group and data with large differences is placed in another data group. The input for the group analysis system is a data set and the similarity in size between the two data (Fadhilah, 2020). While the results of group analysis are a number of groups that form a partition or partition structure of the data set and a general description of each group, which is very important for a

deeper analysis of the characteristics contained in the data. Data grouping must use an approach to look for similarities in data so as to be able to place data into the right groups. Data grouping will divide the data set into several groups where the similarities in one group are greater when compared to other groups (Utomo, 2020).

III. RESEARCH METHODOLOGY

. Thinking Framework According to (Sugiyono, 2019) a conceptual model of how theory relates to various factors that have been identified as important problems.

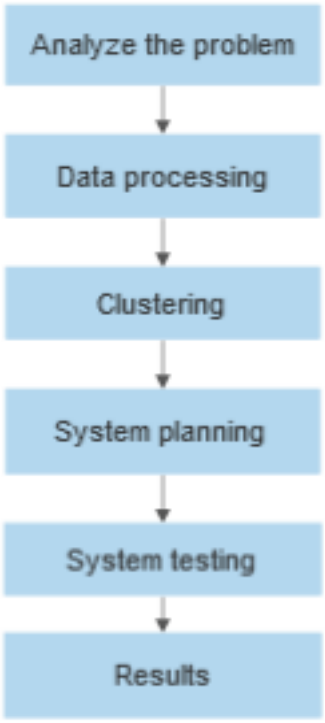

Image 2. Framework of thinking

A. Analysis of Requirement

At this stage the researcher made an understanding of the necessary data related to the research objectives. The data needed is data on student grade transcripts from the fourth semester to the seventh semester from 23 students in the class of 2018, Faculty of Computers and Multimedia, Indonesian National Islamic University. From the value transcript, the researcher can determine the attributes to be used, namely Semester Achievement Index (IPS) 4-7, the average Web Engineer supporting score, the average Programming supporting score and the average Software Engineering supporting score (RPL). These attributes include

1. Semester Achievement Index (IPS)
   Is the average value obtained from the sum of semester 4 to semester 7 scores.

2. The average value of supporting Web Engineer
   Is the average value obtained from the sum of the course scores related to the Web Engineer career. In this attribute, 6 courses were taken, namely: Information Systems, Information Systems Design, Internet & Multimedia, Web Design I, Information Systems Practicum and Web Design II.
3. Programmer support score average
   Is the average value obtained from the sum of the course scores related to the Programmer's career. In this attribute, 6 courses were taken, namely: Programming Techniques and Methods, Object-Oriented Programming, Object-Oriented Programming Practicum, Compilation Techniques, Human-Computer Interaction and Design Patterns.

Is the average value obtained from the sum of course scores related to software engineering careers. In this attribute, 4 courses are taken, namely: Introduction to Robotics, Software Architecture, Software Project Management and Model & Simulation.

1. Determining the center point (centroid) on the cluster.

Table 1. First centroid iteration

Centroid name	ips	Web	prg	Rpl
K1	3,3	3,3	3,1	3,2
K2	3,3	3,4	3,3	3,2
K3	3,4	3,1	3,5	3,5

The first centroid data table is from the student value data table, the data taken are to 4, 11 and 17 (picked at random) data.

2. Calculating data distance to the centroid using the enclidean distance formula

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

It is:

D (x,y) the range of x data to y data

X_1 is a test data to-i

Y_1 is a training data to- i

Source : Suntoro (2019)

a. First Iteration

The first data centroid with 1st data as follows:

$=\sqrt{(3,4 - 3,3)2 + (3,1 - 3,3)2 + (3,1 - 3,1)2 + (3,5 - 3,2)^2}$
$= \sqrt{(0,1)2 + (-0,2)2 + (0,3)2}$
$=\sqrt{0,01 + 0,04 + 0,09}$
$=\sqrt{0,14}$
$=0,374$

Second centroid data with 1st as follows:

$=\sqrt{(3,4-3,3)2+(3,1-3,4)2+(3,1-3,3)2+(3,5-3,2)^2}$
$=\sqrt{(0,1)2+(-0,3)2+(-0,2)^2+(0,3)2}$
$=\sqrt{0,01+0,09+0,04+0,09}$
$=\sqrt{0,23}$
$=0,479$

Third data centroid with 1st as follows:

$=\sqrt{(3,4-3,4)2+(3,1-3,1)2+(3,1-3,5)2+(3,5-3,5)^2}$
$=\sqrt{(-0,4)2}$
$=\sqrt{0,16}$
$=0,4$

From the above calculations it has been obtained that the centroid data to 1 with the object's data to 1 is 0.374, the centroid to 2 with 1 object's data is 0.479 and the -3 data of the centroid with the object's data to 1 is 0.4. In the next step do repetition using the same formula.

IV. RESULT AND DISCUSSION

This study is analyzed the appropriate career-fit recommendations for students based on the academic scores each student has obtained. The study included 23 student files that were used to support the study. The kind of careers that we recommend for this research are web engineers, programmers, and software engineers. Here's a description of the research that's been done.

Tabel 2. Result of Data *Clustering*

Name	Ips	Web	Prg	Rpl	K1	K2	K3	Cluster
Monica Sari	3,4	3,1	3,1	3,5	0,250289089	0,440315286	0,36986484	1
Humaira Fitri	3,5	3,3	3,5	3,5	0,559146339	0,397697454	0,150996689	3
Rita Riskila	3,3	3,3	3,1	3,2	0,206505758	0,271804251	0,498798557	1
Youlanur	3,3	3,3	3,1	3,3	0,180678245	0,299318955	0,427551167	1
Nurul Ulya	3,5	3,1	3,1	3,5	0,318960087	0,464010908	0,367151195	1
Nurul Izzati	3,4	3,1	3,3	3,2	0,289056602	0,154523626	0,45033321	2
Oya Monica	3,3	3,1	3,5	3,2	0,442008937	0,198976975	0,482493523	2
Hayatun Nufus	3,3	3,1	3	3,5	0,219315238	0,503862631	0,474130784	1
M Fauzan	3,3	3,1	3,3	3,2	0,253895274	0,149829835	0,474130784	2
Kulbahri	3,4	3,3	3,3	3,2	0,335625726	0,111574995	0,393446311	2
Rahul Mahfud	3,3	3,4	3,3	3,2	0,371616972	0,187899235	0,427551167	2
Rahmah	3,4	3,3	3,3	3	0,438913007	0,202534955	0,588897275	2
Epa Yanti	3,2	3	3,1	3	0,331537864	0,410077146	0,764722172	1
Etc	.....	...	.....	.....	.....	.....	.....	.....

According to the chart, the five times maximum realignment of the student's data can include the following:

From table 2above comes the following:

1. Cluster 1 contains 11 data which is for a career of the web engineer

2. Cluster 2 contains seven sets of data for a programmer's career
3. Cluster 3 contains 5 data which is for the engineering career of the software

Table 3. Data Cluster

Name	Cluster
Monica Sari	1
Humaira Fitri	3
Rita Riskila	1
Youlanur	1
Nurul Ulya	1
Nurul Izzati	2
Oya Monica	2
Hayatun Nufus	1
M Fauzan	2
Kulbahri	2
Rahul Mahfud	2
Rahmah	2
Epa Yanti	1
Cut Putri Safina S	1
Fajrul Syakban	1
Uswatul Hasanah	1
Ulfa Rahmi	3
Ine santianur	3
Taibatul Aini	2
Syahrel	1
Aulia Faqrurrazi	1
Taufikur Rahman	3
Agus Maulida	3

From the table, the graphic results are obtained as follows:

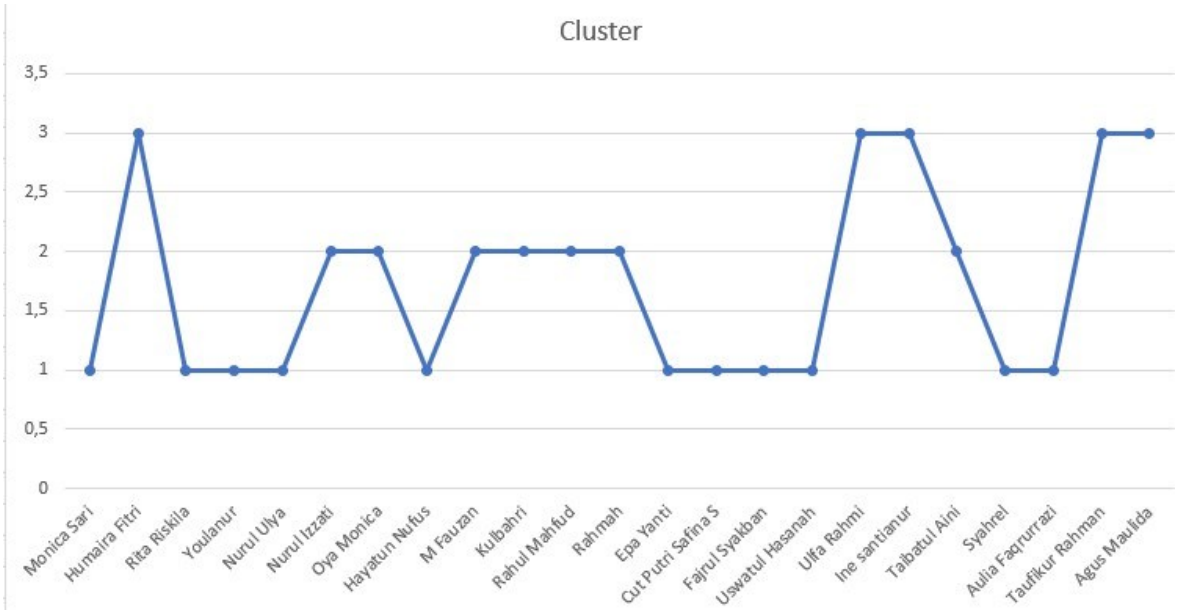

Figure 3. Graph of Clustering Results

From the chart it can be seen that cluster 1 is present in the first and last data gets cluster 3. Clustering in the system is a double file that produces the following data:

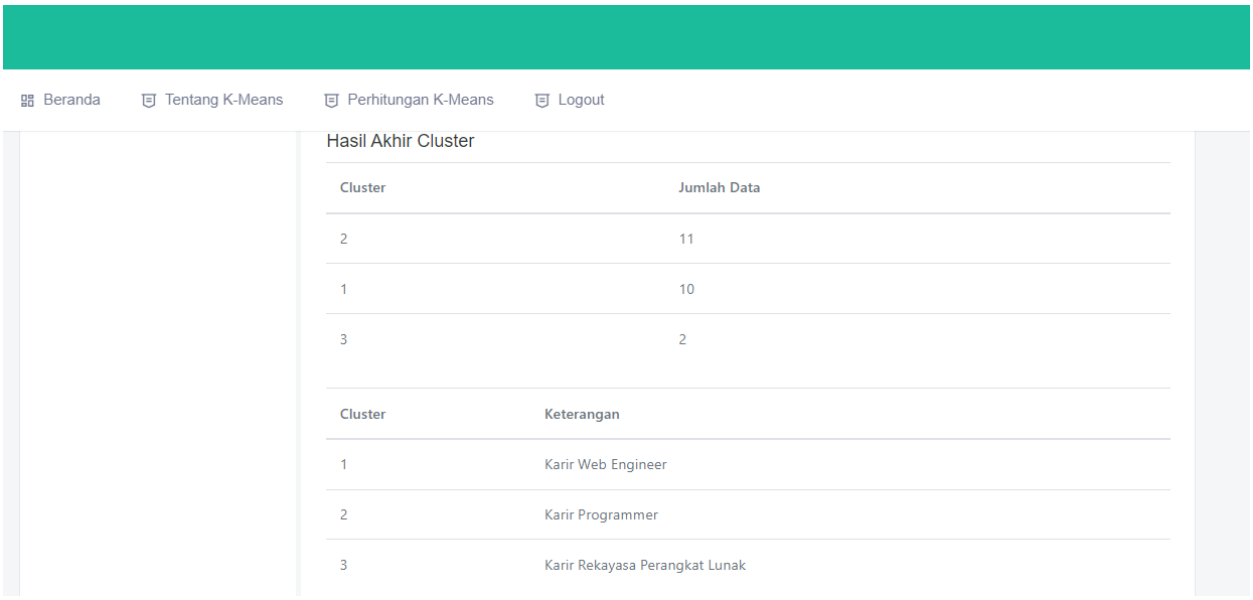

Figure 4. System Clustering Result Data

From the picture are the following results:

1. Cluster 1 contains 10 data
2. Cluster 2 contains 11 data
3. Cluster 3 has 2 data

From the clustering results obtained, the K-Means method is able to group random data to form a new arrangement of criteria according to the iterations needed.

V. CONCLUSION

Based on the results of the discussion of analysis, implementation, and testing of the system, the following conclusions can be drawn: From 23 student data obtained from the academic center of the Faculty of Computers and Multimedia at the Indonesian National Islamic University, the data is grouped into 3 clusters of career compatibility recommendations . The recommended career types are Web Engineer, Programmer and Software Engineering. Based on these data, Cluster 1 has 11 data for Web Engineer careers, Cluster 2 has 8 data for Programmer careers and Cluster 3 has 4 data for Software Engineering careers. In designing this system, it is proven that the k-means algorithm is able to provide career recommendations for students based on the academic value of each student.

REFERENCES

Fadhilah, C., Efendi, S., Zarlis, M., & Azhari, T. (2020). Comparative Performance of Random Forest Algorithms and Discriminant Analysis in Classifications. In 2020 3rd International Conference on Mechanical, Electronics, Computer, and Industrial Technology (MECnIT) (pp. 311-316). IEEE.

Ketutrare (2018), K-Means Clustering Algorithm.

Maulida, D., & Shaleh, AR (2018). The effect of psychological capital and work totality on subjective well-being. Psychohumanities: Journal of Psychological Research, 2(2), 107-124.

Nurrillah, SL (2017). Career Guidance Program to Improve Student Career Maturity. Journal of Innovative Counseling: Theory, Practice, and Research, 1(01).

Mulyati, M., Sadiah, HT, & Febry, AJ (2019, November). 863 Application of the K-Means Clustering Algorithm in Determining Web-Based High School Majors (Case Study of Sma 1 Cisarua Bogor). In National Seminar & Scientific Conference on Information Systems, Informatics & Communication (pp. 863-868).

Priyatman, H., Sajid, F., & Haldivany, D. (2019). Clustering Using the K-Means Clustering Algorithm to Predict Student Graduation Time. Journal of Informatics Education and Research (JEPIN), 5(1), 62.

Situmorang, Z. (2020, June). Analysis optimization k-nearest neighbor algorithm with certainty factor in determining student career. In 2020 3rd International Conference on Mechanical, Electronics, Computer, and Industrial Technology (MECnIT) (pp. 306-310). IEEE.

Sugiyono, S., Sutarman, S., & Rochmadi, T. (2019). Development of a school-level computer based test (CBT) system. Indonesian Journal of Business Intelligence (IJUBI), 2(1), 1-8.

Suntoro, J. (2019). DATA MINING: Algorithm and Implementation with php Programming. Elex Media Komputindo.

Utomo, DP, & Mesran, M. (2020). Comparative Analysis of Data Mining Classification Methods and Attribute Reduction in Heart Disease Data Sets. Budidarma Informatics Media Journal, 4(2), 437-444.

About the Author(s)

The writers had different backgrounds but still within one clump of sciences, as the first authors had expertise in data analysis and system development. The authors, in turn, focus on robotics-based projects and information systems.