

IMPLEMENTATION OF K-MEANS CLUSTERING METHOD TO LECTURERS BASED ON PUBLICATIONS OF NATIONAL JOURNALS AND ACCREDITED SINTA

*DAHNIAL

State Administration Study Program, Satya Negara Palembang School of Administrative Science (STIA)

Jl. Sukatani I No. 3, 8 Ilir, Kec. Team Ilir. II, Palembang City, South Sumatra 30961

e-mail: dahnial.dp@gmail.com

*Corresponding author

ABSTRACT

Lecturer research or scientific results published both nationally and with Sinta accreditation. This is done to determine the level of publication of lecturer research results that have been published so that they can be easily disseminated so that the knowledge gained influences knowledge in general and has an impact on people's lives. The problem in this research is how to apply the K-Means clustering method with reference to the results of lecturer research publications based on national journals and sinta accredited journals. In this study, the researchers used the K-Means method, which has the goal of minimizing the objective function that has been set in the clustering process and maximizing the variation of data in other clusters. So researchers will look for the best number of clusters in the K-means method. The discussion is carried out based on the data that has been collected in the form of national journal weights and sinta accredited journal weights by utilizing Matlab R2013b software. From the results of system testing, lecturer clustering was obtained related to the results of journal weights, namely lecturers with high national scores and no sinta accreditation weights, lecturers with quite high national weights and high sinta accreditation weights, moderate national weights and low sinta accreditation weights, low national weights and high weights with low Sinta accreditation. This study resulted in the conclusion that the data was processed through several stages, starting with calculating the weight value of each issue of the national journal and sinta accredited journal. Then the results of these weights are recapitulated based on each lecturer so that it can be seen the total value of the lecturers' weights in publishing research results in the form of journals that have been published both in the form of national publications and sinta publications.

Keywords: Application, Clustering, K-Means, Lecturer Assessment, National Journal Publication, Sinta Accreditation

1. INTRODUCTION

Activities regarding the evaluation of the performance or quality of performance of each worker or employee are activities that are commonly carried out by organizations and agencies. This also applies to institutions in higher education, whether in the form of universities, institutes, or high schools. Higher Education is a scientific institution that has the task of organizing education and teaching above the secondary level in tertiary institutions and that provides education and teaching based on Indonesian national culture in a scientific way. According to [1], performance appraisal is a very important aspect in order to improve the quality of human resources (HR). This is one way to find out the condition of performance results, including the performance of a lecturer in publishing the results of his scientific work in the form of national and international journals. Specifically for higher education, it will prioritize discussing the relevance of education to development, which in its implementation steps is known by the terms relatedness and equivalence [2]. Academic and management regulations have working procedures to form a system that must be obeyed with the discipline and dedication of all parties. Academic infrastructure and facilities must be created as a foundation, in addition to the quality foundation of higher education is mainly determined by the role of quality and qualified teaching staff (lecturers) [3].

Lecturers must have the necessary qualifications for conveying their knowledge to students so that everything conveyed in the teaching process can be easily accepted, understood, and developed according to the abilities of the students in the field of science they choose. In relation to this qualification, a lecturer must at least have received an equivalent functional position from the Ministry of Research, Technology, and Higher Education, with the position of Expert Assistant. The higher the functional position of the lecturer shows the level of qualification, both in terms of achievement and prestige [3].

In addition, a lecturer must also have high discipline and a sense of responsibility for the knowledge given to students. Lecturer performance can also be seen from the number of research publications that have been conducted. A lecturer is obliged to conduct research. Law Number 14 of 2005 concerning Teachers and Lecturers emphasizes that lecturers must have academic qualifications, competencies, educational certificates, physically and mentally healthy, meet other qualifications required by the higher education unit where they work, and have the ability to realize national education goals. Lecturers are professional educators and scientists with the main task of transforming, developing, and disseminating science, technology, and art through education, research, and community service. While still believing that basically every lecturer will always have good intentions and behavior as well as high integration with their profession. However, standards, procedures, and procedures for assessing credit scores for proposing lecturers' functional promotion positions must still be pursued so that they can be properly able to provide a promotion. Easily to those who really deserve it, but on the contrary, with the right and ability to easily give sanctions to those who deserve it Directorate General of Higher Education [4].

The main task of lecturers in carrying out the Tri Dharma of Higher Education is a unitary Dharma or activity because the three Dharmas can only be distinguished but cannot be separated because they are interrelated and support one another [5]. The Dharma of education and teaching will generate problems and concepts that can drive research to produce scientific publications. On the other hand, research and scientific publications will enrich and renew the repertoire of knowledge to be used in education and teaching. Research results and publications will produce teaching materials that are continuously updated and up-to-date. On the other hand, the results of Dharma research will be applied to the Dharma of community service, and vice versa, the results of Dharma community service will provide inspiration and ideas for research. Thus, it becomes clear that the dharma of research can make a substantial contribution to other dharmas. Therefore, it is no exaggeration if a lecturer's achievements in research and publications are the main benchmarks that describe a lecturer's professionalism as a scientist.

Lecturers are an important parameter in the process of institutional control in higher education, this is due to their very central position, placing lecturers as the main HR as the key holder of operational duties and responsibilities in higher education. With professional skills and close relationships with students and colleagues, lecturers greatly determine the development of the institution, influencing the intellectual and social environment of campus life. Therefore, all efforts made with the aim of improving the performance of lecturers as teaching staff in higher education in a comprehensive manner need to be carried out so that their functions and roles can be carried out optimally in order to achieve organizational goals [6].

According to [7], cluster analysis is the process of partitioning a set of data objects into subsets. Each part is a cluster, so objects in a cluster are similar to one another. However, it is different from the objects in other clusters. The cluster set resulting from cluster analysis can be referred to as clustering. In this context, different clustering methods can result in different clustering of the same data set. Partitioning is not done by humans but by clustering algorithms; therefore, clustering is useful in leading to the discovery of previously unknown groups in the data. Partitioning the dataset into several similar subsets or groups such that the elements of a particular group have a shared set of properties with a high degree of similarity within a group and a low level of similarity between groups. Also called unsupervised learning. If given a number of data points, each of which has a number of attributes, and by using one similarity measure, clusters can be found so that data points in one cluster have greater similarity. Data points in different clusters have little similarity. The similarity measure used is Euclidean distance if the attribute is continuous [8]. The clustering method is a process for finding groups in data. The goal is not to predict the target class variable but to simply derive groupings in the data. For example, customers of a company can be grouped based on consumer behavior. The process of dividing data into meaningful groups is called clustering. In many cases, it is not known which group to look for, and thus the group is difficult to identify. The identified groups are referred to as clusters. Clustering data mining tasks can be used in two different classes: to describe a given data set and as a reprocessing step in other prediction algorithms [9].

Lecturer research is the implementation of the Tri Dharma of Higher Education, including lecturers at STIA Satya Negara Palembang. This research was conducted in order to improve the quality of education as well as the quality of lecturers as educators. Data shows that from 2002 to 2022, there were 273 journal publications for lecturers indexed nationally, in this case at Sinta accreditation. This, of course, still needs to be improved to support the activeness of lecturers in conducting research and publishing research results in international-scale journals.

This research was conducted to determine the level of activeness of lecturers in publishing quality journals; therefore, a calculation is needed to classify the level of activeness of lecturers in publishing research results in the form

of national and international-scale journals using the K-Means Clustering method, which is a method to group research objects based on predetermined criteria. The object to be used in grouping the level of liveliness of lecturer research publications is in the Administrative Sciences Study Program of STIA Satya Negara Palembang.

According to [10], the limitations and problems that are obtained are: how to implement the K-Means Clustering algorithm in grouping lecturers based on national and international journal publications? Likewise, the limitations of the problem in this study are as follows: 1) The criteria used in making decisions are publications in national journals and international journals; 2) The data to be used are primary data taken from the Administrative Sciences Study Program, STA Satya Negara Palembang; 3) Lecturer data to be used are permanent lecturers and extraordinary lecturers (LB) at the STIA Satya Negara Palembang Administrative Science Study Program; 4) Journal publication data to be examined are publications published in the last 3 years until December 2022; 5) Journal data to be examined is data that has been published online; 6) Journal publication data to be examined is data that has been indexed nationally and is accredited by Sinta.

This research was developed from several previous research references that are related to research methods and objects. The use of this reference is intended to provide limitations on methods and systems that will be further developed later. The following are the results of previous studies. Research conducted in [11] entitled "Grouping Students Academic Performance Using One-Way Clustering". The aim of this study is to apply the K-Means clustering algorithm, which serves as an excellent benchmark for monitoring students' progress in learning at school. This study uses the hierarchical method and K-Means to determine student groups. Clustering results were compared, and it was found that K-Means was the most suitable for grouping student achievement [11]. Research entitled "Application of K-Means Clustering Algorithm for Prediction of Students Academic Performance" [12] conducted monitoring student progress in the academic field. Implemented the K-Means clustering algorithm to analyze student data. The clustering algorithm works well in monitoring the development of student performance in the academic field [12]. The advantages of the K-Means clustering method include that even though we can allocate cluster membership absolutely to the data, this can be done at a better granularity level by providing membership percentages [13].

K-Means can be applied to data represented in r -dimensional space. K-Means groups r -dimensional data sets, $X = \{x_i | i=1, \dots, N\}$, where $x_i \in R^d$ denotes the i th data as "data points". K-Means partitions X into K clusters. The K-Means algorithm groups all data points in X so that each point x_i only falls into one of the K partitions. What needs to be considered is which point is in which cluster, which is done by giving each point a cluster ID. Points with the same cluster ID are in the same cluster, while points with different cluster IDs are in different clusters. The parameter that must be entered when using the K-Means algorithm is the value of K . The value of K used is usually based on previously known information about how many data clusters actually appear in X . How many clusters are needed for its implementation, or the type of cluster sought by exploring or experimenting with several K values. How many K values are selected does not require understanding how K-Means partitions the X data set [14].

Research in [16] has also completed a study entitled "An Approach of Improving Student's Academic Performance by using K-Means Clustering Algorithm and Decision Tree". The journal outlines how to reduce significant dropout ratios and improve student performance in academics. The purpose is to partition students based on the same characteristics into groups according to their characteristics and abilities. That research using the data mining process. Clustering K-Means algorithm and decision tree to predict student learning activities [16]. The research entitled "Comparative Analysis & Evaluation of Euclidean Distance Function and Manhattan Distance Function Using K-Means Algorithm" describes the comparison between Euclidean distance and Manhattan distance using the K-Means algorithm when compared according to the number of iterations and the number of double errors. Testing was carried out by utilizing Weka Tools [17]. In [18], there is a difference between the clustering method and the clustering algorithm. A clustering method is a general strategy applied to solve a clustering problem. While the clustering algorithm is only an example of the method. All clustering algorithms can basically be categorized into two main categories. Namely partition and hierarchy. One of the algorithms included in the partition is K-Means [18]. Research entitled "On-Line Clustering of Lecture Performance of the Computer Science Department of Semarang State University Using K-Means Algorithm" describes how to design a system development program in an online form by classifying lecturer performance based on 3 responsibilities using clustering [19].

Based on the description above, a study was conducted on Application of the K-Means Clustering Method to Lecturers Based on National Journal Publications and Accredited by Sinta using K-Means clustering.

2. THEORETICAL BASIS

2.1. Higher Education

The definition of higher education according to Government Regulation Number 60 of 1999, is education at a higher level than secondary education in the school education route. Furthermore, higher education as an educational

unit that organizes higher education can be grouped into two parts, namely the academic and professional paths. Based on the above understanding, students occupy a prestigious position in society and are expected to be of high quality. Likewise, the teaching staff or lecturers are also very influential in determining the quality of higher education [20].

Higher education, as part of the national education system, is expected to have an important and strategic role in achieving educational goals. In Law Number 12 of 2012 concerning Higher Education, Article 1 point 2, what is meant by higher education is the level of education after secondary education which includes diploma programs, undergraduate programs, master programs, doctoral programs, and professional programs, as well as specialist programs, which are organized by universities based on the culture of the Indonesian nation. Higher education has the functions of: (a) developing capabilities and forming dignified national character and civilization in the context of educating the nation's life; (b) developing an innovative, responsive, creative, skilled, competitive, and cooperative academic community through the implementation of the Tri Dharma; and (c) developing science and technology by taking into account and applying humanities values.

2.2. Lecturers

Lecturers are an important component of the education system in higher education. The quality of lecturer performance can be reflected in the productivity and quality of implementation of the three responsibilities, which include activities in the fields of education, research, community service, and other supporting activities. In the context of implementing the Tri Dharma of Higher Education, lecturers carry out three types of activities: education and teaching, research, and community service. As mentioned earlier, the main field of activity for lecturers is carrying out education and teaching. However, research and community service activities must also be carried out by a lecturer. These two activities will greatly support better education and teaching activities.

Universities that have qualified lecturers will be in great demand by the public. Therefore, a program to improve the quality of lecturers is an obligation that is not negotiable at this time or in the future. Universities that do not want to keep up with current and future changes will be abandoned by society and sooner or later will experience setbacks that will eventually collapse.

a. Lecturer Journal Publication

The functional position of a lecturer is basically an acknowledgment, appreciation, and trust in competence, performance, integrity, and responsibility in carrying out tasks, as well as lecturer ethics in carrying out the Tri Dharma of Higher Education. While still believing that basically every lecturer will always have good intentions and behavior, as well as high integration with their profession. However, standards, procedures, and procedures for assessing credit scores for proposing lecturers' functional promotion positions must still be pursued so that they can be properly able to provide a promotion. Easily to those who really deserve it, but on the contrary, with the right and ability to easily give sanctions to those who deserve them, the Directorate General of Higher Education.

b. Assessment of Lecturer Research Credit Scores

Performance appraisal is a very important aspect in order to improve the quality of human resources (HR). This is one way to find out the condition of performance results, including the performance of a lecturer in publishing the results of his scientific work in the form of national and international journals.

The type of scientific work that is the main requirement for occupying a certain level of academic position may differ from one another. In addition, for certain scientific works that are used in the promotion of academic positions, a recognized highest limit is applied. The determination of the highest limit that is recognized is adjusted to the criteria for academic positions. Types of activities, criteria, and highest credit score submissions in research and dissemination of science, technology, and the arts (IPTEKS) are presented in Table 1.

A scientific journal, periodical, or scientific magazine, hereinafter referred to as a journal, is a form of publication that functions to register scholarly activities, certify the results of activities that meet minimum scientific requirements, disseminate them widely to the general public, and archive all findings resulting from the scientific activities of scientists and pundits that they publish.

c. Lecturer in Administrative Science Study Program STIA Satya Negara Palembang

The Institute for Research and Community Service (LP2M) at STIA Satya Negara Palembang noted that since the last 3 years, from 2019 to December 2022, as many as 54 lecturers have conducted research, and as many as 154 journals have been indexed in Sinta accreditation from 2002 to December 2022, 104 of which are the result of research conducted by lecturers at the STIA Satya Negara Palembang Administrative Study Program. Of course, this is still relatively low when compared to other universities, which are very active in publishing the research results of educators on a national scale or on a Sinta-accredited scale.

Table 1 Types of Activities and Credit Scores

No.	Type of activity	Code	Evidence of	Cr	Activity
1.	1) Jurnal internasional bereputasi (terindeks pada <i>database</i> internasional bereputasi dan berfaktor dampak)	II.A.1.b.1.1	Pindai halaman sampul, dewan redaksi/redaksi pelaksana, daftar isi dan bukti kinerja Pindai halaman sampul,	40	
	2) Jurnal internasional terindeks pada <i>database</i> internasional bereputasi	II.A.1.b.1.2	Dewan redaksi/redaksi pelaksana, daftar isi dan bukti kinerja	30	
	3) Jurnal internasional terindeks pada <i>database</i> internasional di luar kategori 2	II.A.1.b.1.3	Pindai halaman sampul, redaksi pelaksana, daftar isi dan bukti kinerja	20 15	
	5) a. Jurnal nasional berbahasa Indonesia terindeks pada DOAJ b. Jurnal nasional berbahasa Inggris atau bahasa resmi (PBB) terindeks pada DOAJ	II.A.1.b.2.1	Pindai halaman sampul, Redaksi pelaksana, daftar isi dan bukti kinerja	20	
	6) Jurnal nasional	II.A.1.b.3	Pindai halaman sampul, Dewan redaksi/redaksi pelaksana, daftar isi dan bukti kinerja Tinggi	10	Paling tinggi 25% dari AK unsur penelitian yang diperlukan untuk pengusulan ke Lektor Kepala dan Profesor yang diterbitkan di jurnal nasional

The academic community of STIA Satya Negara Palembang continues to launch programs to increase the activeness of teaching staff in conducting research, including the selection of research funding for teaching staff. This is a strategy to increase activity and improve the quality of the research results of educators conducting research in the form of scientific journals so that the results of these journals can be accredited both as national journals and as sinta accreditation.

The higher the level of active teaching staff conducting research, of course, can have a positive impact on the academic community of STIA Satya Negara Palembang, and the more teaching staff in a department conducting research can also impact the accreditation value of a department. This is of course highly expected so that a department, especially at STIA Satya Negara Palembang, will get the best accreditation from the higher education. This is because, in assessing the accreditation of a department, there is a form from the research results of a lecturer in the department that adds value to the accreditation of the department at STIA Satya Negara Palembang.

3. RESEARCH METHOD

The stages of the research are described in the figure 1.

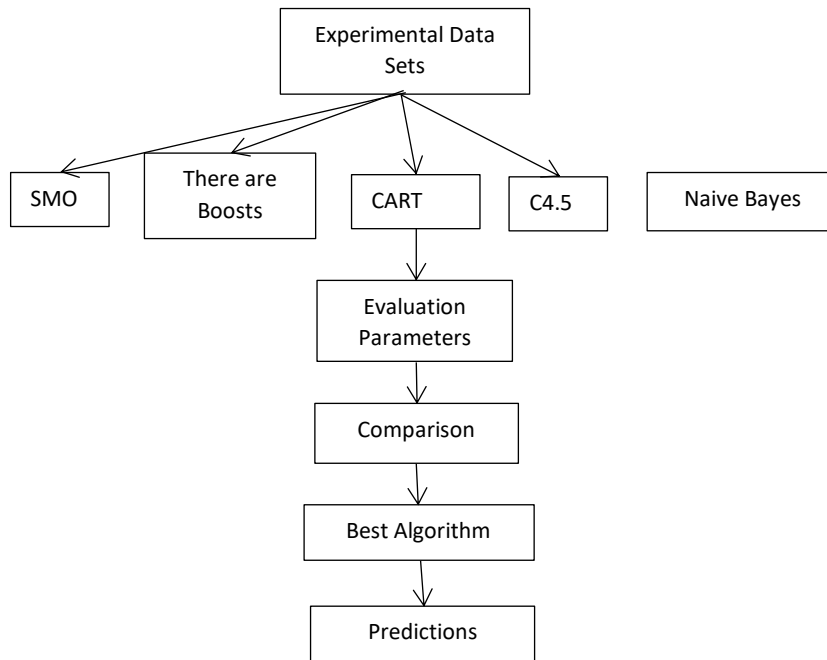


Figure 1. Research Flow

3.1. Data Set

The research materials used in this study are three different data sets taken from sinta.kemdikbud.or_publication.drtpm@kemdikbud.go.id

Table 1. Data Sets

Dataset Name	Number of National Journals	Number of Sinta Accredited Journals	Journal Characteristics	Characteristics of Data sets	Missing Value
1. WOS	10	214	Numeric	Multivariate	No
2. Bima	34	351	Numeric	Multivariate	No
3. Garuda	35	683	Nominal	Multivariate	Yes

Data sets 1 and 2 have the same data type, but the number of national journals and sinta accredited journals is different. Data sets 2 and 3 have the same number of national journals, with different sinta accredited journals.

3.2. Data Analysis Model

This study uses the SMO, Adaboost, CART, C4.5, and Naïve Bayes classification algorithms on the Weka tool.

3.3. Testing mode

The test method used is 10 fold cross validation.

3.4. Evaluation Parameters

Evaluation parameters that will be used are accuracy, built time, root mean square error.

3.5. Experiment

- 1) Experiment 1 compares the accuracy of the data set with the number of national and accredited journals different sinta.
- 2) Experiment 2 compares the accuracy of data sets with different data types (nominal numeric)
- 3) Experiment 3 compares the accuracy of the data set with the size of the data set with nominal data type

4) Experiment 4 compares the accuracy of data sets with the size of data sets with numeric data types

In this study the method used is the data mining method [5] as follows. (a) data collection stage, (b) data processing stage, (c) clustering stage and (d) analysis stage.

3.6. Clustering Method

The clustering method is one of the main data analysis methods to help identify groupings of data objects from datasets. Clustering is unsupervised classification and is the process of partitioning a set of data objects from one set into several classes. This can be done by applying various equations and steps regarding the distance algorithm, namely the Euclidean [21].

Partitioning the dataset into several similar subsets or groups such that the elements of a particular group have a shared set of properties with a high degree of similarity within a group and a low level of similarity between groups is also called unsupervised learning. If given a number of data points, each of which has a number of attributes, and by using one similarity measure, clusters can be found so that data points in one cluster have greater similarity. Data points in different clusters have little similarity. The similarity measure used is Euclidean distance if the attribute is continuous.

Cluster analysis is a multivariate technique whose main objective is to group objects based on their characteristics. Cluster analysis classifies objects so that each object that is most closely related to other objects is in the same cluster. The groups formed have high internal homogeneity and high external heterogeneity. The focus of cluster analysis is to compare objects based on sets of variables, this is why experts define set variables as a critical stage in cluster analysis. The cluster variable set is a set of variables that represent the characteristics used by the objects.

Cluster analysis is the process of partitioning a set of data objects into subsets. Each part is a cluster, so objects in a cluster are similar to one another. However, it is different from the objects in other clusters. The cluster set resulting from cluster analysis can be referred to as clustering. In this context, different clustering methods can result in different clustering of the same data set. Partitioning is not done by humans but by clustering algorithms; therefore, clustering is useful in leading to the discovery of previously unknown groups in the data.

The clustering method is a process for finding groups in data. The goal is not to predict the target class variable but to simply derive groupings in the data. For example, customers of a company can be grouped based on consumer behavior. The process of dividing data into meaningful groups is called clustering. In many cases, it is not known which group to look for, and thus the group is difficult to identify. The identified groups are referred to as clusters. Clustering data mining tasks can be used in two different classes: to describe a given data set and as a reprocessing step in other prediction algorithms. There is a difference between the clustering method and the clustering algorithm. A clustering method is a general strategy applied to solve a clustering problem. While the clustering algorithm is only an example of the method, all clustering algorithms can basically be categorized into two main categories. Namely partition and hierarchy. One of the algorithms included in the partition is K-Means.

3.7. Clustering K-Means

K-Means can also be interpreted as a clustering method, which is included in the partitioning approach. The K-Means algorithm is a centroid model. Centroid mode is a model that uses centroids to create clusters. A centroid is the midpoint of a cluster. Centroid is a value. The centroid is used to calculate the distance of a data object to the centroid. A data object is included in a cluster if it has the shortest distance to the cluster centroid. The K-Means algorithm can be interpreted as a simple learning algorithm to solve a grouping problem that aims to minimize double errors.

The K-Means algorithm is an algorithm that belongs to the partitioning clustering approach. Each cluster is connected by a centroid (center point). Each point is assigned to the cluster with the nearest centroid. The number of clusters K must be determined. The basic algorithm is very simple: choose K points as the initial centroid repeat.

1. Form K clusters by locating all the points closest to them
2. Repeat the centroid calculation for each cluster
3. Until the centroid does not change

K-Means can be applied to data represented in r -dimensional space. K-Means groups r -dimensional data sets, $X = \{x_i | i=1, \dots, N\}$, where $x_i \in R^d$ denotes the i th data as "data points". K-Means partitions X into K clusters. The K-Means algorithm groups all data points in X so that each point x_i only falls into one of the K partitions. What needs to be considered is which point is in which cluster, which is done by giving each point a cluster ID. Points with the same cluster ID are in the same cluster, while points with different cluster IDs are in different clusters. The parameter that must be entered when using the K-Means algorithm is the value of K . The value of K used is usually based on previously known information about how many data clusters actually appear in X . How many clusters are needed for its implementation, or the type of cluster sought by exploring or experimenting with several K values. How many K values

are selected does not need to be understood by K-Means when partitioning the X data set. Clustering K-Means is a prototype-based grouping method in which the data set is divided into k clusters. K-Means is one of the simplest and most commonly used clustering algorithms. In this technique, the user determines the number of clusters (k) that need to be grouped in a data set. The purpose of clustering K-Means is to find the prototype data points for each cluster. All of these data points are then assigned to the nearest prototype, which then forms a cluster. The prototype is referred to as the centroid, cluster center. The cluster center can be the mean of all data objects in the cluster, as in K-Means, or a representative data object, as in K-Medoid Clustering. The cluster centroid or data object does not have to be a real data point in the data set and can be an imaginary data point that represents the characteristics of all data points in the cluster.

The advantages of the K-Means clustering method include that even though we can allocate cluster membership absolutely to the data, this can be done at a better granularity level by providing membership percentages. The K-Means algorithm is based on a simple idea. Initially, it is determined how many clusters will be formed. Any object or first element in the cluster can be selected to serve as the cluster centroid point. The K-Means method algorithm will then repeat the following steps until stability occurs (no objects can be moved).

1. Determine the coordinates of the midpoint of each cluster
2. Determine the distance of each object to the center point coordinates
3. Grouping these objects based on their minimum distance
4. The flowchart of the K-Means method algorithm can be seen in Figure 2.

4. RESULTS AND DISCUSSION

Cluster analysis is an analysis of various objects based on their level of similarity. Clustering analysis is the pivot for data mining. K-Means is one of the simplest unsupervised learning algorithms for solving a well-known grouping problem. The K-Means algorithm aims to minimize the objective function, in this case, the double error function.

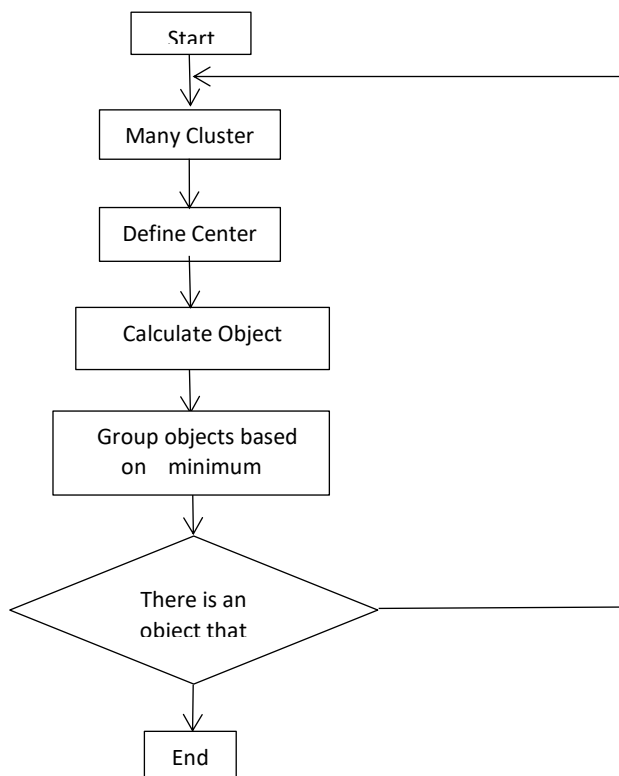


Figure 2 Flowchart of the K-Means Method Algorithm (Safriyanto, 2010: 5) [15]

Table 2. Author data

Author	Number of National Journals	Number of Sinta Accredited Journals
A	1	3
B	3	3
C	4	3
D	5	3
E	1	2
F	4	2
G	1	1
H	2	1

The K-Means algorithm has many real time applications, but its performance cannot be guaranteed to match that required by random initial centroids. The computational complexity of K-Means is quite high due to the need to define a large number of data points [21].

Sample case:

Sinta ABC's National and Accredited Journal Publications have data on authors who have obtained scores, in the form of the number of national and accredited Sinta journals owned by customers, which are presented in Table 2.

Clustering is expected to be able to produce groups of authors who meet the following characteristics:

1. Authors with almost the same number of national and accredited journals will be in the same author group.
2. Authors whose number of national journals and Sinta-accredited journals is quite different will be in different groups.

The following are the clustering steps using the K-Means algorithm:

- a. Step 1: Determine the desired number of clusters (eg k = 3)
- b. Step 2: Select the initial centroid randomly. In this step, 3 pieces of data will be randomly selected as centroids, for example: data (B, E, F)
 $M1 = (3,3)$, $M2 = (1,2)$, $M3 = (4,2)$
- c. Step 3: Calculate the distance to the centroid.

In this step, each data will be determined by the closest centroid, and the data will be determined as a member of the group closest to the centroid.

To calculate the distance to the centroid of each cluster on author A as follows:
 Data: (1,3), M1 centroid: (3,3), M2 centroid: (1,2), M3 centroid: (4,2)

$$DM1 = \sqrt{(1 - 3)^2 + (3 - 3)^2} = 2$$

$$DM2 = \sqrt{(1 - 1)^2 + (3 - 2)^2} = 1$$

$$DM3 = \sqrt{(1 - 4)^2 + (3 - 2)^2} = 3.162$$

Table 3. Distance Calculation Results

Author	Distance to centroid cluster1	Distance to centroid cluster 2	Distance to centroid cluster3	Closest distance
A	2	1	3.162	C2
B	0	2.236	1.414	C1
C	1	3.162	1	C3
D	2	4.123	1.414	C3
E	2.236	0	3	C2
F	1.414	3	0	C3
G	2.828	1	3.162	C2
H	2.236	1.414	2.236	C2

The results of calculations related to calculating the distance to the centroid can be seen in Table 3 where the calculation of the distance between the point and the centroid can be known so that the closest distance can be known.

From Table 3, the customer membership is obtained as follows.

Cluster 1 = {B}, Cluster 2 = {A,E,G,H}, Cluster 3 = {C,D,F}

In this step, the ratio between the BCV (Between Cluster Variation) and WCV (Within Cluster Variation) is also calculated:

Because centroid $M1=(3,3)$, $M2=(1,2)$, $M3=(4,2)$

$$d(m1,m2) = \sqrt{(3 - 1)^2 + (3 - 2)^2} = 2.236$$

$$d(m1,m3) = \sqrt{(3 - 4)^2 + (3 - 2)^2} = 1.414$$

$$d(m2,m3) = \sqrt{(1 - 4)^2 + (2 - 2)^2} = 3$$

$$BCV = d(m1,m2) + d(m1,m3) + d(m2,m3) = 2.236 + 1.414 + 3 = 6,650$$

In this case $d(m_i,m_j)$ represents the Euclidean Distance from m to m_j Meanwhile, to calculate the WCV, namely by selecting the smallest distance between the data and the centroid in each cluster can be seen in Table 4.

Table 4. Distance Between Data and Centroid

Author	Distance to smallest centroid
A	1
B	0
C	1
D	1.414
E	0
F	0
G	1
H	1.414

Table 5. Distribution of Cluster

Cluster 1		
Author	Number of National Journals	Number of Sinta Accredited Journals
B	3	3
Mean	3	3
Cluster 2		
Author	Number of National Journals	Number of Sinta Accredited Journals
A	1	3
E	1	2
G	1	1
H	2	1
Mean	1.25	1.75
Cluster 3		
Author	Number of National Journals	Number of Sinta Accredited Journals
C	4	3
D	5	3
F	4	2
Mean	4.33	2.67

$$WCV=1^2+0^2+1^2+1.414^2+0^2+0^2+1^2+1.414^2 = 7$$

So that the ratio of $BCV/WCV = 6,650/7 = 0,950$

Because this step is iteration 1 then proceed to the next step.

d. Step 4: Update the centroid by calculating the average value in each cluster. Table 5 shows the results of the Cluster division.

So we get a new centroid, namely: $m1=(3,3)$, $m2=(1.25,1.75)$, $m3=(4.33,2.67)$

e. Step 3: (iteration-2) return to step 3, if there is still data moving clusters or if the centroid value is above the threshold value, or if the value in the objective function used is still above the threshold. Furthermore, in this step, data is placed again in the nearest centroid, the same as in step-3, to calculate the distance to the centroid of each cluster in author A as follows:

Data : (1,3), $m1=(3,3)$, $m2=(1.25,1.75)$, $m3=(4.33,2.67)$

$$dm1 = \sqrt{(1 - 3)^2 + (3 - 3)^2} = 2$$

$$dm2 = \sqrt{(1 - 1.25)^2 + (3 - 1.75)^2} = 1.275$$

$$dm3 = \sqrt{(1 - 4.33)^2 + (3 - 2.67)^2} = 3.350$$

Calculation results related to the centroid distance to the customer's point can be seen in Table 6 regarding the shortest distance.

From table 6, the author membership is obtained as follows: Cluster 1= {B}, Cluster 2= {A,E,G,H}, Cluster 3= {C,D,F}

In this step, the ratio between the BCV (Between Cluster Variation) and WCV (Within Cluster Variation) is also calculated:

$$BCV= d(m1,m2) + d(m1,m3) + d(m2,m3) = 6,741$$

$$WCV=1.275^2+0^2+0.471^2+0.745^2+0.354^2+0.745^2+0.791^2+1.064^2 = 4.833$$

$$\text{So big ratio } BCV/WCV = 6.741/4.833 = 1.394$$

When compared, the current ratio (1,394) is greater than the previous ratio (0,950) therefore the algorithm continues to the next step.

f) Step 4 – iteration 3

In this step, update the centroid again. The results related to the distribution of clusters can be seen in Table 7 of each cluster.

Table 6 Centroid Distance Calculation Results

Author	Distance to centroid cluster1	Distance to centroid Cluster2	Distance to centroid Cluster3	Distance closest
A	2	1.275	3.350	C2
B	0	1.768	1.374	C1
C	1	3.021	0.471	C3
D	2	3.953	0.745	C3
E	2.236	0.354	3.399	C2
F	1.414	2.813	0.745	C3
G	2.828	0.791	3.727	C2
H	2.236	1.061	2.867	C2

Table 7 Cluster Division

<i>Cluster 1</i>		
Author	Number of National Journals	Number of Sinta Accredited Journals
B	3	3
Mean	3	3
<i>Cluster 2</i>		
Author	Number of National Journals	Number of Sinta Accredited Journals
A	1	3
E	1	2
G	1	1
H	2	1
Mean	1.25	1.75
<i>Cluster 3</i>		
Author	Number of National Journals	Number of Sinta Accredited Journals
C	4	3
D	5	3
F	4	2
Mean	4.33	2.67

Table 8. Centroid Distance Calculation Results

Author	Distance to centroid cluster1	Distance to centroid Cluster2	Distance to centroid Cluster3	Distance closest
A	2	1.275	3.350	C2
B	0	1.768	1.374	C1
C	1	3.021	0.471	C3
D	2	3.953	0.745	C3
E	2.236	0.354	3.399	C2
F	1.414	2.813	0.745	C3
G	2.828	0.791	3.727	C2
H	2.236	1.061	2.867	C2

g) The third step of iteration 3

To calculate the distance to the centroid of each cluster on author A as follows:

Data Author A: (1,3), m1=(3,3), m2=(1.25,1.75), m3=(4.33,2.67)

$$dm1 = \sqrt{(1 - 3)^2 + (3 - 3)^2} = 2$$

$$dm2 = \sqrt{(1 - 1.25)^2 + (3 - 1.75)^2} = 1.275$$

$$dm3 = \sqrt{(1 - 4.33)^2 + (3 - 2.67)^2} = 3.350$$

Table 8 shows the results of calculating the distance to the midpoint of all customers.

From the table 8, the author membership is obtained as follows:

Cluster 1= {B}, Cluster 2={A,E,G,H}, Cluster 3 = {C,D,F}

In this step, the ratio between the BCV (Between Cluster Variation) and WCV (Within Cluster Variation) is also calculated:

$$BCV = d(m1,m2) + d(m1,m3) + d(m2,m3) = 6,741 \quad WCV = 1.275^2 + 0^2 + 0.471^2 + 0.745^2 + 0.354^2 + 0.745^2 + 0.791^2 + 1.064^2 = 4.833$$

So big ratio $BCV/WCV = 6.741/4.833 = 1.394$

When compared, the current ratio (1,394) is no longer greater than the previous ratio (1,394) therefore the algorithm will be stopped.

5. CONCLUSION AND SUGGESTION

5.1 Conclusion

From the results of research and discussion regarding the implementation of the K-Means Clustering method for lecturers based on national journal publications and sinta accreditation at the STIA Satya Negara Palembang Administrative Sciences Study Program during the period January 2019 to December 2022, the following conclusions can be drawn: The system for implementing the K-Clustering method K-Means in determining the grouping of lecturers related to the publication results of national journals and sinta accredited journals can be created by utilizing the Matlab software. The data is processed through several stages, starting with calculating the weight value of each national journal publication and sinta accredited journal. Then the results of the weights are recapitulated based on each lecturer so that the total value of the lecturers' weight in publishing research results in published journals can be known, both in the form of national publications and sinta-accredited publications. Then the data that has been recapitulated is stored in excel form so that it can be displayed in Matlab by importing data. By using the system that has been created, the data is displayed into the system, and then the data is clustered so that cluster results are obtained from all lecturers in the Administrative Science Study Program.

From this research, it can be seen that the best classification algorithm used on numeric data types is C4.5, while for nominal data is SMO. The best classification algorithms to use on small datasets or data sets with a small number of instances are Naïve Bayes and SMO, while the best classification algorithms to use on big datasets are SMO and C4.5.

Based on the results of the implementation of the system by applying the K-Means Clustering method, it can be concluded that regarding the results of the clustering, 4 groups were made, namely lecturers who had a high national weight and a high sinta accredited weight of 0, lecturers with a high national weight and a low sinta accredited weight of 5 lecturers, lecturers with a low national weight and a high sinta accredited weight of 9 lecturers, and lecturers with a low national weight and a low sinta accredited weight of 20 lecturers.

5.2 Suggestion

Suggestions that can be made for this research are as follows.

- 1) This system allows it to be developed using software other than Matlab software to make it easier and more usable by the general public. This is because Matlab software requires lots of memory on the computer.
- 2) For further research, it is hoped that there will be more input variables in detail to be used as data that is implemented into the system. Where there is a calculation to get the weight value before obtaining the weight value between the national journal and the journal sinta accredited.

REFERENCES

- [1] A. Alamsyah, A., & I.H. Muna, (2016). Metode Fuzzy Inference System untuk Penilaian Kinerja Pegawai Perpustakaan dan Pustakawan. *Scientific Journal of Informatics*, 3(1), 88-98.
- [2] M. I. Artawan. (2002). *Strategi Meningkatkan Mutu Pendidikan di Perguruan Tinggi*. Fakultas Ekonomi Marwadewa, Bali
- [3] M.R. Asmawi, (2005). Strategi Meningkatkan Lulusan Bermutu di Perguruan Tinggi. *Makara Sosial Humaniora*, 66-71.
- [4] Dinas Kebudayaan. (2014). Pedoman Operasional Penilaian Angka Kredit Kenaikan Pangkat/Jabatan Akademik Dosen.
- [5] Direktorat Jenderal Pendidikan Tinggi. (2014). *Pedoman Akreditasi Terbitlah Berkala ilmiah*. Jakarta: Ditlitabmas.
- [6] A. Efferi, (2015). Manajemen Perguruan Tinggi dan Kinerja Dosen. *Jurnal Fakultas Agama Islam*, 3(1).
- [7] J. Han, M. Kamber, M & J. Pei, (2011). *Data mining: concepts and techniques*. Elsevier.
- [8] F.A. Hermawati, (2013). *Data Mining*. Yogyakarta: ANDI.

- [9] V. Kotu, & B. Deshpande, (2014). *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*. Morgan Kaufmann.
- [10] Lembaga Penelitian dan Pengabdian Kepada Masyarakat. (2016). *Buku Data Penelitian dan Pengabdian Kepada Masyarakat 2015*. Semarang: Universitas Negeri Semarang.
- [11] N.S. Rasid, & N. Ahmad, (2018). Grouping Students Academic Performance Using One-Way Clustering. *Int. J. Sci. Commer, Hummanit*, 2(3).
- [12] O.J. Oyelade, O.O. Oladipupo, & I.C. Obagbuwa, (2010). Application of k Means Clustering algorithm for prediction of Students Academic Performance. *International Journal of Computer Science and Information Security*. 7(1), 292-295.
- [13] J. Pal, & V. Bhattacharjee, (2015). Fuzzy Rule Based Classifier For Software Quality Data. *Journal of Theoretical and Applied Information Technology*, 72(1).
- [14] E. Prasetyo, (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta. Penerbit ANDI Yogyakarta, 218.
- [15] A. Safriyanto, (2010). Perancangan Aplikasi K-Means Untuk Pengelompokan Mahasiswa STMIK Elrahma Yogyakarta Berdasarkan Frekuensi Kunjungan Ke Perpustakaan Dan IPK. *Jurnal Teknologi Informasi dan Ilmu Komputer (FAHMA)*.
- [16] M. Shovon, H.Islam, & M. Haque, (2012). An Approach of Improving Students Academic Performance by using k means clustering algorithm and Decision tree. *International Journal of Advanced Computer Science and Application*, 3(8), 146-149.
- [17] A. Singla, & M. Karambir, (2012). Comparative Analysis & Evaluation of Euclidean Distance Function and Manhattan Distance Function Using K- Means Algorithm. *International Journal of Advanced Research in Computer Science and Software Engineering (IJARSSE)*, 2(7), 298-300.
- [18] N. Soni, & A. Ganatra, (2012). Categorization of several Clustering algorithms from different perspective: a review. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(8), 63-68.
- [19] E. Sugiharti, & M.A. Muslim, (2016). On-line Clustering of Lecturers Performance of Computer Science Department of Semarang State University Using K-Means Algorithm. *Journal of Theoretical and Applied Information Technology*, 83(1).
- [20] A. Trisnaningih, (2011). Faktor-Faktor Yang Mempengaruhi Kinerja Dosen Akuntansi. *Jurnal Akuntansi dan Auditing*, 8(1), 83-94.
- [21] B. Venkateswarlu, & P.G. Raju, (2013). Mine Blood Donors Information through Improved K-Means Clustering. *International Journal of Computational Science and Information Technology*, 1(3), 9-15