# E-Archive Document Clustering Information System Using K-Means Algorithm


Aida Fitriyani
*Faculty of Computer Science*
*Universitas Bhayangkara Jakarta Raya*
Jakarta, Indonesia 17121
aida.fitriyani@dsn.ubharajaya.ac.id

Dwipa Handayani
*Faculty of Computer Science*
*Universitas Bhayangkara Jakarta Raya*
Jakarta, Indonesia 17121
dwipa.handayani@dsn.ubharajaya.ac.id

Achmad Noeman
*Faculty of Computer Science*
*Universitas Bhayangkara Jakarta Raya*
Jakarta, Indonesia 17121
achmad.noeman@dsn.ubharajaya.ac.id

Asep Ramdhani Mahbub
*Faculty of Computer Science*
*Universitas Bhayangkara Jakarta Raya*
Jakarta, Indonesia 17121
asep.ramdhani@dsn.ubharajaya.ac.id

Ratna Salkiawati
*Faculty of Computer Science*
*Universitas Bhayangkara Jakarta Raya*
Jakarta, Indonesia 17121
ratna.salkiawati@dsn.ubharajaya.ac.id

Ahmad Fathurrozi
*Faculty of Computer Science*
*Universitas Bhayangkara Jakarta Raya*
Jakarta, Indonesia 17121
ahmad.fathurrozi@dsn.ubharajaya.ac.id



2022 Seventh International Conference on Informatics and Computing (ICIC) | 979-8-3503-4571-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICIC56845.2022.10006935



***Abstract*— Archiving is an important activity in a company, because the archive is a form of decision making as physical evidence owned by the company. Records management system in a company or organization must be managed properly. Currently, the archive management system is still manual which has the risk of documents being easily damaged and lost either due to natural disasters or unwanted accidents. Therefore, it is necessary to manage archives that can use a computerized system to facilitate users and avoid data loss with various factors. The research method uses the K-Means Clustering algorithm to facilitate the process of grouping archives, and the prototype is the method that will be used for software development. The function of this application is to simplify the process of archiving and searching for archived data. Management carried out by the system includes archive storage, archive destruction, and archive file reports.**



**Keywords— Archives, K-Means Algorithm, Clustering**


## I. INTRODUCTION

Archives are the beginning of information in an organization, company and all of its stakeholders, this is because the archive is part of the evidence that will be responsible for its authenticity for all activities that have been carried out in the organization and one form of decision making as physical evidence owned by the company [1].

The importance of archives that act as decision making to be made into physical evidence, therefore archives that have been stored in an organization must be managed properly. Archiving is an activity to store data to be used as evidence of correct information and according to needs without any confusion or ambiguity in processing information [1]

The existence of the archive has its own very important place for all users, be it a company or organization, but that does not mean separating themselves from the company or organization [2]. At this time the archive management system is still manual which has the risk of documents being easily damaged and lost either due to natural disasters or unwanted accidents. Therefore, archive management is needed that can use a computerized system to facilitate users and avoid data loss with various factors.

The k-means clustering algorithm in data grouping is based on the distance between a large number of data to the cluster centroid point obtained through an iterative process quickly and efficiently, [3].

Prototype is a process of the initial stages in the design of a structured system by going through the stages in the manufacturing process, with the aim of developing the model into the final result of the final system. the stages that must be passed in the prototype method include: gathering requirements, designing and evaluating prototypes [6].

Clustering is a technique of grouping data by separating data with identical elements. In the IT world this classification method is carried out for grouping certain data in large quantities [7].

Therefore, with a web-based E-Archive information system, the archiving process will be easier in grouping using the K-Means Clusterization algorithm technique, so it is hoped that it can help and simplify the grouping process in the archiving system. The K-Means method uses a working method by grouping data (clusters) based on predetermined characteristics [7]

## II. METHODOLOGY

K-Means can be defined as a grouping of data that is unsupervised learning which functions to group data into data clusters [9]. The steps in implementing the K-Means Clustering algorithm [3]:

1. Determination of the number of clusters k
2. Determination of the value of the center that is used as a basis later. Done at random or self-chosen.
3. Data allocation which is seen from the proximity of the data to the central value with the data placement formula based on the following:

$$Di = \sqrt{((Xi - K)2 + (Yi - S)2)}$$

Symbols :

D*i* : Distance i to centroid
*i* : number of data objects
X,Y : The coordinates of the data object, the X coordinates of the data object variable 1, the Y coordinates of the data object variable 2, and so on if there are more than 2 sample variables as follows:

(Xi-K)2+(Yi-S)2+(Zi-T)2+…
K,S : Centroid coordinates, K Central coordinates for object variable X,S for object Y.

4. The search for a new central value by calculating the average of the data that has been in the cluster at stage 3. At the cluster or grouping stage to determine which data belongs to which grouping, the provisions for grouping are given, namely:
   a. The input requirements for cluster 1 are obtained from the formula: If Di ⩽ Dy And Di ⩽ Dz)
      If the conditions are met then it is included in the cluster group 1
   b. The requirements for entering cluster 2 can be seen in the formula: If (Dy ⩽ Di And Dy ⩽ Dz)
      if the conditions are met then it is included in the cluster group 2
   c. The requirements for entering cluster 3 can be seen in the formula: If (Dz ⩽ Dy And Dz ⩽ Di)
      If the conditions are met then it is included in the cluster group 3.
      Where:
      *i, y, z* = number of data objects

      D*i* = distance of object I with centroid cluster 1
      D*y* = distance of object I with centroid cluster 2
      D*z* = distance of object I with centroid cluster 3

      Calculation of the new centroid value based on the example in the table below
      Centroid 1 = (data1 +data 2+ data 7)/3
      Centroi 2 = (data2 +data 4)/2
      Centroid 3 = (data5 +data 6)/2

Do the 3rd step again using the latest main value (which was calculated in Step 4). Do this until the new main value with the previous one is the same or does not change, then clustering can be stopped.

TABLE I. CLUSTER DATA COLLECTION

| Data | *Cluster 1* | *Cluster 2* | *Cluster 3* |
|---|---|---|---|
| 1 | x | | |
| 2 | | x | |
| 3 | x | | |
| 4 | | x | |
| 5 | | | x |
| 6 | | | x |
| 7 | x | | |

K-Means Algorithm is used to group archiving data owned by companies or organizations. The K-Means formula will be applied to grouping files based on the specified specifications. The application of the k-means clustering method is a calculation for analyzing the data obtained in the implementation of the system. In the application of this method, the author will use archived data within 1 month, here are the data documents that are archived:

TABLE II. ARCHIVE DOCUMENT DATA

| No | Department | No | Date |
|---|---|---|---|
| 1 | PPIC | MTO Platform, Ladder and Handrail IKPT R.1 | 15-Jun-21 |
| 2 | PPIC | JBT-00032-MTO Jembatan A50-N R.0 & PR 173951 | 15-Jun-21 |
| 3 | Engineering | Cutting plan jembatan A-50N R.0 | 15-Jun-21 |
| 4 | Engineering | Terlampir cutting plan jembatan B-60 2 UNIT R.0 | 15-Jun-21 |
| 5 | Marketing | SPK No S-JBT-00034-VI-2021 Jembatan PT. Triwira Bangun Sejahtera | 14-Jun-21 |
| 6 | Marketing | Desain Drawing Ladder Platform Dearator | 15-Jun-21 |
| 7 | Marketing | Desain Drawing Ladder Platform Transfer Air Receiver | 15-Jun-21 |
| 8 | PPIC | JBT-00033-MTO Jembatan B-60 2 UNIT R.0 & PR 173947 | 15-Jun-21 |
| 9 | Engineering | Report list JPO Manggarai R.2 | 14-Jun-21 |
| 10 | Engineering | Cutting plan tambahan pipe support jpo | 16-Jun-21 |
| 11 | PPIC | MTO Palu-3 Wika R. 2 | 14-Jun-21 |
| 12 | PPIC | MTO Palu-3 Wika R. 3 | 16-Jun-21 |
| 13 | Marketing | SPK S-JBT-00036-VI-2021 Jembatan AG20S PT. Wiratama Globalindo Ja | 16-Jun-21 |
| 14 | Marketing | Desain Drawing Ladder Platform Instrument | 16-Jun-21 |
| 15 | Engineering | JBT-00031-Report List Jembatan C-24.4 | 16-Jun-21 |
| 16 | Engineering | JBT-00031-Welded Beam Jembatan C-24.4 R.0 | 7-Jun-21 |
| 17 | Engineering | Welded jembatan C.24 R.1 | 8-Jun-21 |
| 18 | Marketing | SPK S-KUM-00027-VI-2021 JPO Manggarai Pipe Support PT NCK | 16-Jun-21 |
| 19 | PPIC | MTO conveyor TLS-5 R.18 | 16-Jun-21 |
| 20 | PPIC | MTO Incenerator R.5 | 16-Jun-21 |
| 21 | PPIC | JBT-00031-MTO Jembatan C-24.4 R.0 | 8-Jun-21 |
| 22 | PPIC | JBT-00031-MTO Jembatan C-24.4 R.1 | 16-Jun-21 |
| 23 | Engineering | KUM-00015-Report List Limestone - Palu3 R.0 | 17-Jun-21 |
| 24 | Marketing | Design Drawing Buchket Wheel Structure | 17-Jun-21 |
| 25 | Marketing | Design Drawing Buchket Wheel Structure | 18-Jun-21 |
| 26 | Engineering | report Palet Pipe Sleeve. | 15-Jun-21 |
| 27 | Marketing | SPK No S-JBT-00037-VI-2021 PT. Wiratama Globalindo Jaya | 21-Jun-21 |
| 28 | Engineering | KUM-00015-Report List Fire Station & Fire Pump Shelter-Palu3 R.1 | 21-Jun-21 |
| 29 | Engineering | Cutting Plan Limestone -Palu3 R.0 | 17-Jun-21 |
| 30 | Engineering | KUM-00015-Report List Fire Station & Fire Pump Shelter-Palu3 R.0 | 14-Jun-21 |
| 31 | Engineering | KUM-00015-Report List Workshop & Warehouse-Palu3 R.1 | 21-Jun-21 |
| 32 | PPIC | MTO Linkset Jembatan B-60 | 22-Jun-21 |
| 33 | PPIC | MTO Jembatan AG-20S | 22-Jun-21 |
| 34 | Marketing | Design Drawing Buchket Wheel Structure | 22-Jun-21 |
| 35 | Marketing | Design Drawing Continuous Ship Unloader Shop Drawing and Document | 22-Jun-21 |
| 36 | Marketing | SPK No C-KUM-00028-VI-2021 PT. Nindya Karya (persero) Tbk | 23-Jun-21 |
| 37 | Marketing | SPK No S-JBT-00038-VI-2021 PT. Wiratama Globalindo Jaya | 23-Jun-21 |
| 38 | Marketing | SPK No S-JBT-00039-VI-2021 PT. | 23-Jun-21 |

| No | Department | No | Date |
|---|---|---|---|
| | | Wiratama Globalindo Jaya | |
| 39 | Engineering | JBT-00035-Cutting Plan Linkset Jembatan B-60 R.0 | 23-Jun-21 |
| 40 | Engineering | JBT-00036-Cutting Plan Jembatan AG-20S R.0 | 23-Jun-21 |
| 41 | PPIC | MTO Palu-3 Wika R. 4 | 24-Jun-21 |
| 42 | Marketing | SPK S-JBT-00042-VI-2021 Jembatan BG25 PT. Wiratama Globalindo Jay | 24-Jun-21 |
| 43 | Marketing | Desain Drawing Ladder Platform Service Air Server | 24-Jun-21 |
| 44 | Marketing | Desain Drawing Ladder Platform Ammonia Tank | 24-Jun-21 |
| 45 | Marketing | Desain Drawing Ladder Platform Condensate Holding Storage Tank | 24-Jun-21 |
| 46 | Marketing | SPK S-PJL-00057-VI-2021 PT. Wiratama Globalindo Jaya | 24-Jun-21 |
| 47 | Marketing | SPK S-JBT-00041-VI-2021 PT. Wiratama Globalindo Jaya | 24-Jun-21 |
| 48 | Marketing | SPK S-JBT-00043-VI-2021 PT. Wiratama Globalindo Jaya | 24-Jun-21 |
| 49 | Marketing | SPK S-PJL-00059-VI-2021 PT. Wiratama Globalindo Jaya | 25-Jun-21 |
| 50 | Engineering | KUM-00018-Report List Reclaimer-L15-Bucket R.0 | 3-Jun-21 |
| 51 | Engineering | KUM-00018-Report List Reclaimer-L16-Central Transfer and Stockyard C | 3-Jun-21 |
| 52 | Engineering | KUM-00018-Report List Reclaimer-Tripper and Upper R.4 | 3-Jun-21 |
| 53 | Engineering | KUM-00018-Report List Reclaimer-Tripper Tail R.0 | 3-Jun-21 |
| 54 | Engineering | KUM-00018-Report List Reclaimer-Tripper and Upper R.3 | 2-Jun-21 |
| 55 | PPIC | MTO Reclaimer R.6 | 10-Jun-21 |
| 56 | PPIC | MTO Reclaimer R.7 | 16-Jun-21 |
| 57 | PPIC | MTO Reclaimer R.8 | 18-Jun-21 |
| 58 | PPIC | MTO Reclaimer R.9 | 25-Jun-21 |
| 59 | Marketing | Design Drawing Continuous Ship Unloader Shop Drawing and Document | 26-Jun-21 |
| 60 | Marketing | R.1 SPK No S-JBT-00037-VI-2021 PT. Wiratama Globalindo Jaya | 23-Jun-21 |
| 61 | Marketing | R.2 SPK No S-JBT-00037-VI-2021 PT. Wiratama Globalindo Jaya | 28-Jun-21 |
| 62 | Marketing | SPK S-KUM-00029-VI-2021 PT WIJAYA KARYA Add Palu 3 | 28-Jun-21 |
| 63 | Marketing | SPK No S-JBT-00040-VI-2021 PT. Wiratama Globalindo Jaya | 28-Jun-21 |
| 64 | Engineering | KUM-00024-Report List Anchor Plate IKPT R.1 | 7-Jun-21 |
| 65 | Engineering | KUM-00025-Report List Anchor Plate R.1 | 7-Jun-21 |
| 66 | Engineering | KUM-00025-Report List Anchor Plate R.2 | 28-Jun-21 |
| 67 | Engineering | report list coal crushing R.5 | 4-Jun-21 |
| 68 | PPIC | MTO Jembatan CG-12 R.0 | 29-Jun-21 |
| 69 | PPIC | MTO Jembatan BG-12 R.0 | 29-Jun-21 |
| 70 | PPIC | MTO Platform, Ladder and Handrail IKPT R.2 | 29-Jun-21 |
| 71 | Marketing | SPK S-PJL-00056-VI-2021 PT WGJ Pekerjaan Kolom SPBU | 24-Jun-21 |
| 72 | PPIC | MTO RDMP RU-V R.6 | 30-Jun-21 |
| 73 | PPIC | MTO Platform, Ladder and Handrail IKPT R.3 | 30-Jun-21 |
| 74 | PPIC | MTO Add. Local Platform R.3 | 30-Jun-21 |
| 75 | PPIC | MTO Jembatan CG-30 R.0 | 29-Jun-21 |
| 76 | PPIC | MTO Jembatan CG-30 R.1 | 30-Jun-21 |
| 77 | PPIC | MTO Jembatan CG-12 R.1 | 30-Jun-21 |
| 78 | PPIC | MTO Jembatan BG-12 R.1 | 30-Jun-21 |

1. Pre-processing stage, at this stage the data will be pre-processed to get the data needed to perform calculations. The data obtained is the date of the archive, the number of records and the number of departments.

TABLE III. PRE PROCESSING

| No | Archive Date | Number of Archives | Number of Departments |
|---|---|---|---|
| 1 | 2-Jun-21 | 1 | 1 |
| 2 | 3-Jun-21 | 4 | 1 |
| 3 | 4-Jun-21 | 1 | 1 |
| 4 | 7-Jun-21 | 3 | 1 |
| 5 | 8-Jun-21 | 2 | 2 |
| 6 | 10-Jun-21 | 1 | 1 |
| 7 | 14-Jun-21 | 4 | 3 |
| 8 | 15-Jun-21 | 8 | 3 |
| 9 | 16-Jun-21 | 10 | 3 |
| 10 | 17-Jun-21 | 3 | 2 |
| 11 | 18-Jun-21 | 2 | 2 |
| 12 | 21-Jun-21 | 3 | 3 |
| 13 | 22-Jun-21 | 4 | 1 |
| 14 | 23-Jun-21 | 6 | 3 |
| 15 | 24-Jun-21 | 9 | 2 |
| 16 | 25-Jun-21 | 2 | 2 |
| 17 | 26-Jun-21 | 1 | 1 |
| 18 | 28-Jun-21 | 4 | 3 |
| 19 | 29-Jun-21 | 4 | 1 |
| 20 | 30-Jun-21 | 6 | 1 |

2. The stages in determining the centroid randomly, so here will be a calculation to identify the nearest centroid of an object. Three sample samples for the calculation process in determining the centroid value. The following is sample data:

TABLE IV. CLUSTER CENTER

| Amount | Central Cluster | Clusters (Random) | |
|---|---|---|---|
| C1 | 3-Jun-21 | 4 | 1 |
| C2 | 16-Jun-21 | 10 | 3 |
| C3 | 30-Jun-21 | 6 | 1 |

At this stage, the data will be calculated in centroids, to determine the distance in each cluster. The following is a table for calculating the closest distance to each cluster

$$d(x,y) = \sqrt{\sum_{i-1}^{n} (y1 - xi)^2}$$

a. C1

$$d(b_1,c_1) = (x_1-c_{1x})^2+(y_1-c_{1y})^2 = \sqrt{(1-4)^2+(1-1)^2}$$
$$= \sqrt{(3)^2+0}$$
$$= \sqrt{9}$$
$$= 3$$

b. C2

$$d(b_1,c_1) = (x_1-c_{2x})^2+(y_1-c_{2y})^2 = \sqrt{(1-10)^2+(1-3)^2}$$
$$= \sqrt{(-9)2+(-2)2}$$
$$= \sqrt{81+4}$$
$$= \sqrt{85}$$
$$= 9.21954446$$

c. C3

$$d(b_1,c_1) = (x_1-c_{2x})^2+(y_1-c_{2y})^2 = \sqrt{(1-6)^2+(1-1)^2}$$
$$= \sqrt{(-5)2+0}$$
$$= \sqrt{25}$$
$$= 5$$

TABLE V. CLUSTER RESULTS

| Archive Center | C1 | C2 | C3 |
|---|---|---|---|
| Date | 3-Jun-21 | 16-Jun-21 | 30-Jun-21 |
| 2-Jun-21 | 3 | 9.219544457 | 5 |
| 3-Jun-21 | 0 | 6.32455532 | 2 |
| 4-Jun-21 | 3 | 9.219544457 | 5 |
| 7-Jun-21 | 1 | 7.280109889 | 3 |
| 8-Jun-21 | 2.236067977 | 8.062257748 | 4.123105626 |
| 10-Jun-21 | 3 | 9.219544457 | 5 |
| 14-Jun-21 | 2 | 6 | 2.828427125 |
| 15-Jun-21 | 4.472135955 | 2 | 2.828427125 |
| 16-Jun-21 | 6.32455532 | 0 | 4.472135955 |
| 17-Jun-21 | 1.414213562 | 7.071067812 | 3.16227766 |
| 18-Jun-21 | 2.236067977 | 8.062257748 | 4.123105626 |
| 21-Jun-21 | 2.236067977 | 7 | 3.605551275 |
| 22-Jun-21 | 0 | 6.32455532 | 2 |
| 23-Jun-21 | 2.828427125 | 4 | 2 |
| 24-Jun-21 | 5.099019514 | 1.414213562 | 3.16227766 |
| 25-Jun-21 | 2.236067977 | 8.062257748 | 4.123105626 |
| 26-Jun-21 | 3 | 9.219544457 | 5 |
| 28-Jun-21 | 2 | 6 | 2.828427125 |
| 29-Jun-21 | 0 | 6.32455532 | 2 |
| 30-Jun-21 | 2 | 4.472135955 | 0 |

4. The stage in this step determines the grouping of the clusters by determining the value of the closest distance from each cluster. There are values such as d(b1,c1) and d(b1,c2) > d(b1,c2) then the closest distance will be C2. The following is a table of each grouping against each cluster:

TABLE VI. RESULTS OF CLUSTER GROUPING

| Archive Center / Date | Shortest Distance |
|---|---|
| 2-Jun-21 | C1 |
| 3-Jun-21 | C1 |
| 4-Jun-21 | C1 |
| 7-Jun-21 | C1 |
| 8-Jun-21 | C1 |
| 10-Jun-21 | C1 |
| 14-Jun-21 | C1 |
| 15-Jun-21 | C2 |
| 16-Jun-21 | C2 |
| 17-Jun-21 | C1 |
| 18-Jun-21 | C1 |
| 21-Jun-21 | C1 |
| 22-Jun-21 | C1 |
| 23-Jun-21 | C3 |
| 24-Jun-21 | C2 |
| 25-Jun-21 | C1 |
| 26-Jun-21 | C1 |
| 28-Jun-21 | C1 |
| 29-Jun-21 | C1 |
| 30-Jun-21 | C3 |

4. The stage in this step will create a new cluster to get the smallest data, the calculation is done by using a way to find the minimum value and the value of the square of the smallest value. The following is the formula for finding the minimum distance value in the cluster:

$$d_{euclidean}(x, y) = \sum_{i=1} (xi - yi)^2$$

a. min (C1, C2, C3)

min (3:9.2129544457:5)

= 3

5. The steps in the next step are to determine the ratio between the tween cluster variation (BVC) and within cluster variation (WCV) quantities.

BCV = cluster 1 to cluster 2 + cluster 1 to cluster 3 + cluster2 to cluster 3 BCV value = d(m1,m2) + d(m1,m3) + d(m2,m3)
= 6.32455532+2+4.472135955
= 12.79669128

Meanwhile, to find out the WCV, that is by squaring the calculation results in each cluster.
WCV = min (C1,C2,C3)
= min (3:9.2129544457:5)
= (32) = 9

TABLE VII. WCV CALCULATION RESULTS

| | |
|---|---|
| 3 | 9 |
| 0 | 0 |
| 3 | 9 |
| 1 | 1 |
| 2.236067977 | 5 |
| 3 | 9 |
| 2 | 4 |
| 2 | 4 |
| 0 | 0 |
| 1.414213562 | 2 |
| 2.236067977 | 5 |
| 2.236067977 | 5 |
| 0 | 0 |
| 2 | 4 |
| 1.414213562 | 2 |
| 2.236067977 | 5 |
| 3 | 9 |
| 2 | 4 |
| 0 | 0 |
| 0 | 0 |
| WCV | 77 |

7. At this stage is to add up the BCV and WCV to find the ratio. The ratio is the result of the calculation conclusion from the data that has been processed. The ratio is BCV/WCV
R= 12.79669128/77 R= 0.166190796

After the results of the ratio are known, in the calculation process to find the average value of the data that has been grouped, it is known that the average data of the departments that often carry out the archive process is as shown in the table below:

Department = $\frac{1+4+1+3+2+1+4+3+2+3+4+2+1+4+4}{20}$

TABLE VIII. RESULTS OF THE AVERAGE SCORE

| | C1 | | C2 | | C3 | |
|---|---|---|---|---|---|---|
| 2-Jun-21 | 1 | 1 | | | | |
| 3-Jun-21 | 4 | 1 | | | | |
| 4-Jun-21 | 1 | 1 | | | | |
| 7-Jun-21 | 3 | 1 | | | | |
| 8-Jun-21 | 2 | 2 | | | | |
| 10-Jun-21 | 1 | 1 | | | | |
| 14-Jun-21 | 4 | 3 | | | | |
| 15-Jun-21 | | | 1 | 1 | | |
| 16-Jun-21 | | | 1 | 1 | | |
| 17-Jun-21 | 3 | 2 | | | | |
| 18-Jun-21 | 2 | 2 | | | | |
| 21-Jun-21 | 3 | 3 | | | | |
| 22-Jun-21 | 4 | 1 | | | | |
| 23-Jun-21 | | | | | 1 | 1 |
| 24-Jun-21 | | | 1 | 1 | | |
| 25-Jun-21 | 2 | 2 | | | | |
| 26-Jun-21 | 1 | 1 | | | | |
| 28-Jun-21 | 4 | 3 | | | | |
| 29-Jun-21 | 4 | 1 | | | | |
| 30-Jun-21 | | | | | 1 | 1 |
| | 2.6 | 1.666666667 | 1 | 1 | 1 | 1 |

## III. Conclusions

The design of this e-archive information system will facilitate the process of digitally storing archives on the other hand, streamlining the use of archives or space. The k-means algorithm method helps the process of grouping (clustering) archive documents to be done faster. It is hoped that in the future there will be regular maintenance or maintenance of system applications as well as backing up data to avoid damage to the database.

## References


[1] Almahdi, G. F. S., & Pahlevi, T. (2020). Management of Electronic Filing Systems as a Determinant of Employee Work Productivity in Solokuro District, Lamongan Regency. Journal of Office Administration Education (JPAP) (Pengelolaan Sistem Kerasipan Elektronik Sebagai Determinan Produktivitas Kerja Pegawai di Kecamatan Solokuro Kabupaten Lamongan. Jurnal Pendidikan Administrasi Perkantoran), 8(2), 295–304. https://journal.unesa.ac.id/index.php/jpap/article/view/8425/4085

[2] Bahrudin, R. M., Ridwan, M., & Darmojo, H. S. (2019). Application of the Helpdesk Ticketing System in Handling Complaints on the Use of Web-Based Information Systems (Penerapan Helpdesk Ticketing System Dalam Penanganan Keluhan Penggunaan Sistem Informasi Berbasis Web). Jutis, 7(1), 71–82.

[3] Eric Fammaldo and Lukman Hakim. (2018). Application of the K-Means Clustering Algorithm for Grouping the Level of Family Welfare for the Smart Indonesia Card Program. Scientific Journal of Applied Information Technology (Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Tingkat Kesejahteraan Keluarga Untuk Program Kartu Indonesia Pintar. Jurnal Ilmiah Teknologi Informasi Terapan), V(1), 24–32. http://journal.widyatama.ac.id/index.php/jitter/article/view/249/213

[4] Firmansyah, Y., Maulana, R., & Fatin, N. (2020). Website-Based Citizen Complaint Information System (Sistem Informasi Pengaduan Warga Berbasis Website) (Case Study: Kelurahan Siantan Tengah, North Pontianak). Scholar's Journal, XIX(April), 397–404. https://jurnal.dcc.ac.id/index.php/JC/article/view/328

[5] Irawan, M. D., & Simargolang, S. A. (2018). Implementation of E-Archive in the Informatics Engineering Study Program. Journal of Information Technology (Implementasi E-Arsip Pada Program Studi Teknik Informatika. Jurnal Teknologi Informasi), 2(1), 67. https://doi.org/10.36294/jurti.v2i1.411

[6] Noeman, A., & Handayani, D. (2019). Design of Information System Document Monitoring Sampling Product with Prototype Method (Perancangan Sistem Informasi Document Monitoring Sampling Product Dengan Metode Prototype). Facts Exacta, 12(3), 219–229. https://doi.org/10.30998/factorexacta.v12i3.4678

[7] Rahayu, S. B., & Aisyah, S. (2018). Analysis of Archives Handling in the General and Protocol Section at the East Jakarta Mayor's Office (Analisis Penanganan Arsip Pada Bagian Umum Dan Protokol Di Kantor Walikota Jakarta Timur), 2(1), 33–44.

[8] Seprilya, N., & Damayanti, F. (2020). Management Analysis of "Filing System (Correspondence) in Improving Communication Effectiveness at PT Dirgantara Indonesia." (Analisis Pengelolaan "Filing System (Tata Persuratan) Dalam Meningkatkan Efektivitas Komunikasi di PT Dirgantara Indonesia) Journal of Secretary & Business Administration (JSAB), 4(2), 157. https://doi.org/10.31104/jsab.v4i2.193

[9] Sugiono, Nurdiani, S., Linawati, S., Safitri, R. A., & Saputra, E. P. (2019). Grouping Student Behavior in E-Learning Lectures with K-Means Clustering. Journal of Scientific Studies, Bhayangkara University (Pengelompokan Perilaku Mahasiswa Pada Perkuliahan E-Learning dengan K- Means Clustering. Jurnal Kajian Ilmiah Universitas Bhayangkara Jakarta Raya), Greater Jakarta, 19(2), 126–133. https://jurnal.ubharajaya.ac.id/index.php/kajian-Sains/article/view/410/pdf.

[10] Tousalwa, C. C., & Nanulaitta, D. T. (2020). Archive Destruction at the Amahusu Village Office (Pemusnahan Arsip Pada Kantor Desa Amahusu), Nusaniwe District, Ambon City. 03(02), 259–268.

[11] Wisswani, N. W., Ngurah, I. G., Catur, B., Winetra, I. W. C., Elektro, J. T., Bali, P. N., Jimbaran, K. B., South, K., & Tuban, K. P. (2019). Record Center Archives of Decree (Sk) For Universities (Record Center Arsip Surat Keputusan (Sk) Untuk Universitas). 11(1), 1–8. https://jurnal.polibatam.ac.id/index.php/JI/article/view/1166/689

[12] Sinaga, A.K. Jain, "Data clustering: 50 years beyond k-means," Pattern Recognition Letters, vol. 31, pp. 651–666, 2010.

[13] M. Alhawarat and M. Hegazi, "Revisiting k-means and topic modeling, a comparison study to cluster arabic documents," IEEE Access, vol. 6, pp. 42740-42749, 2018

[14] Y. Meng, J. Liang, F. Cao, Y. He, "A new distance with derivative information for functional k-means clustering algorithm," Information Sciences, vol. 463–464, pp. 166–185, 2018

[15] J. Zhu, Z. Jiang, G.D. Evangelidis, C. Zhang, S. Panga, Z. Li, "Efficient registration of multi-view point sets by k-means clustering," Information Sciences, vol. 488, pp. 205–218, 2019

[16] J. Wu, J. Chen, H.Xiong, M. Sie, "External validation measures for k-means clustering: a data distribution perspective," Expert Syst. Appl., vol. 36, pp. 6050- 6061, 2009

[17] Z. Lv, T. Liu, C. Shi, J.A. Benediktsson, H. Du, "Novel land cover change detection method based on k-means clustering and adaptive majority voting using bitemporal remote sensing images," IEEE Access, vol.7, pp. 34425-34437, 2019