Jurnal Informatika Universitas Pamulang **ISSN: 2541-1004**
Penerbit: Program Studi Teknik Informatika Universitas Pamulang **e-ISSN: 2622-4615**
Vol. 6, No. 4, Desember 2021 (818-825) **10.32493/informatika.v6i4.14122**

K-Means Algorithm in Grouping Central Civil Servants by Agencies and Groups in Pematangsiantar

Septri Wanti Siahaan[1], Indra Gunawan[2]

[1,2]Program Studi Teknik Informatika, STIKOM Tunas Bangsa Pematangsiantar, Jl. Jend. Sudirman Blok A, No. 1,2, dan 3, Kota Pematangsiantar, Sumatera Utara, Indonesia, 21143
e-mail: [1]septriwanti26@gmail.com, [2]indra@amiktunasbangsa.ac.id

Submitted Date: October 26th, 2021 Reviewed Date: January 08th, 2022
Revised Date: January 09th, 2022 Accepted Date: January 31st, 2022

Abstract

Civil Servants (PNS) are Indonesian citizens who have certain conditions and are given the mandate to be a State Civil Apparatus on a permanent basis and are not bound by a contract to carry out their assigned duties and positions. There is a view that states that civil servants have useful guarantees in the future and there is no contract system in work. Each government has civil servants with different groups. This study aims to group the central civil servants according to agencies and groups in Pematangsiantar. In completing this research, it was carried out using the k-means algorithm. The data used in the study came from the Central Statistics Agency (BPS) by processing the data into 3 clusters, namely high, medium, and low. The results of this study indicate that the k-means algorithm can be applied in determining the grouping of Central Civil Servants by Agencies and Groups in Pematangsiantar with the results that 1 item is high (C1), 4 items are medium (C2), and 21 items are low (C3).

Keywords: Grouping; Civil Servants; K-Means Algorithm; Data Mining; Institutions Government

1. Introduction

Every year the state of Indonesia holds a test for Candidates for Civil Servants. The number of applicants who register for the CPNS test is very large, because people have the view that civil servants have useful guarantees in the future and there is no contract system in work. The more people who register to enter the CPNS test, the tougher the competition faced by applicants.

Civil Servants (PNS) are Indonesian citizens who are appointed as government officials and work in government agencies. In carrying out each task as government officials, civil servants are divided into ranks and groups. The increase in the class of civil servants is influenced by education, performance, years of service and others. The higher the rank and class owned, the higher the salary earned by Civil Servants. Each division or government agency in Pematangsiantar City has Civil Servants consisting of groups: I, II, III, and IV.

Pematangsiantar City has a different number of civil servants in each government agency or division in Pematangsiantar. In civil servants, there are many fields or professions where in measuring employee performance, each field has its own assessment

The problem in this research is that each government agency in Pematangsiantar has a number of different fields or professions in each agency, therefore the method used in solving research problems is to group the Central PNS in Pematangsiantar city based on the agency and class by using K-Means Algorithm. After the implementation of the K-Means Algorithm in research problems, it will produce a grouping of Central Civil Servants according to Agencies and Groups in Pematangsiantar with the results being classified as high (C1), moderate (C2), and classified as low (C3). The K-Means algorithm is an algorithm found in data mining to perform the process of grouping data. The data in this study were obtained from the Central Statistics Agency (BPS).

2. Research Methods

This research uses research methods contained in data mining, namely the clustering. Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine

learning techniques to extract and identify useful information and related knowledge from large databases (Irnanda et al., 2019). Clustering is an activity (task) that purpose to group data that has similarities between one data and other data into clusters or groups so that data in one cluster has a maximum level of similarity (similarity) and data between clusters has a minimum similarity (Aditya et al., 2020). Clustering is a technique of one of the data mining functionality, the clustering algorithm is an algorithm for grouping a number of data into certain data groups (clusters) (Maulida, 2018). The algorithm used in the grouping method is the K-Means algorithm. The K-Means algorithm is a group analysis method that leads to the partitioning of N objects of observation into K groups, where each object of observation is a data group with the closest mean (Parlina et al., 2018). K-Means is one of the non-hierarchical data grouping methods that can partition data into two or more groups (Gustientiedina et al., 2019). The K-Means algorithm is an algorithm in data mining that is used in grouping data using a separator that aims to divide it into different blocks or clusters. According to (Nurzahputra et al., 2017) the clustering method used in this study is the K-Means method. The K-Means algorithm partitions data into groups so that data with the same characteristics are included in the same group and data with different characteristics are grouped into other groups (Nasution & Eka, 2018). The purpose of grouping this data is to minimize the objective function set in the grouping process, which generally tries to minimize variation within a group and maximize variation between groups (Rosmini et al., 2018). K-Means can also be interpreted as a clustering method that is included in the partitioning approach. The K-Means algorithm is a centroid model. Centroid model is a model that uses centroids to create clusters. The steps in the K-Means Algorithm, namely (Wardhani, 2016) :

1. Specify k as the total of clusters you want to form. Set the cluster center.
2. Calculating the distance of each data to the center of the cluster using the Euclidean equation.

$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \quad (1)$$

3. Group the data into clusters with the shortest distance using the equation.

$$Min \sum_{k=1}^{k} d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \quad (2)$$

4. Calculate the new cluster center using the equation.

$$C_{kj} = \frac{\sum_{i=1}^{p} x_{ij}}{P} \quad (3)$$

Where:
Xij E cluster to – k
p = number of cluster members to k
5. Repeat steps 2 to 4 until there is no more data moving to another cluster.

Basically, the utilization of the K-Means algorithm in the clustering process rely on which are available data and the conclusions to be reached. For this reason, the K-Means algorithm is used which contains the following rules:
a. The number of clusters that need to be entered.
b. Only has numeric type attribute (Agustin, 2015).

The data used in this research is sourced from the Central Statistics Agency (BPS). The tools used to help complete this research are using Microsoft Excel 20017 and Rapidminer software. Rapidminer is a tool used to assist data processing in data mining. Text analysis, extracting patterns from large datasets and combining them with statistics, artificial intelligence, and databases is the task of Rapidminer. The scheme in the K-Means Algorithm is as follows:

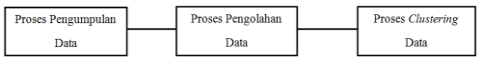

Figure 1. K-Means . Algorithm Schematic

In Figure 1, it is explained that the data collection process, namely in classifying data on Central Civil Servants by Agencies and Groups in Pematangsiantar, was obtained from the Central Statistics Agency (BPS). After the data collection process, the next process the data will be data processing by making it into 3 clusters. The data that has been obtained are as follows:

Table 1. Data on the Number of Civil Servants by Agencies and Groups in Pematangsiantar in 2019

Satuan/Unit Kerja/*Division*	**Golongan I**	**Golongan II**	**Golongan III**	**Golongan IV**	**Jumlah/Total**
LAPAS Kelas II Pematangsiantar	0	79	57	1	137
Imigrasi Pematangsiantar	0	1	42	1	44
BPS Kota Pematan-	0	2	11	2	15

Satuan/Unit Kerja/*Division*	Golongan I	Golongan II	Golongan III	Golongan IV	Jumlah/Total
gsiantar					
Kejari I Pematangsiantar	0	7	23	3	33
Pengadilan Negeri Kelas IB P.Siantar	0	3	26	4	33
Kantor Pertanahan Kota Pematangsiantar	1	7	18	2	28
BP DAS Asahan Barumun	0	4	32	1	37
KPPN Pematangsiantar	0	1	23	1	25
MIN Bah Kapul	0	3	6	8	17
MTsN.Pematangsiantar	0	7	28	11	46
Balai Diklat Kehutanan	0	14	32	10	56
KPP Pematangsiantar	0	38	48	3	89
KPKNL Pematangsiantar	0	12	17	2	31
KPBC Pematangsiantar	0	35	21	1	57
Pengadilan Agama Pematangsiantar	0	0	12	2	14
MAN Pematangsiantar	0	1	18	7	26
KPU Kota Pematangsiantar	0	2	9	0	11
Kanwil DJP Sumatera Utara II	0	37	51	13	101
LP Narkotik-a	0	46	12	2	60
Ditjen	0	1	23	2	26
Bimas Islam Kemenag Kota Pematangsiantar					
Ditjen Bimas Kristen Kemenag Kota Pematangsiantar	0	0	20	33	53
Ditjen Bimas Khatolik Kemenag Kota Pematangsiantar	0	1	6	5	1
Ditjen Bimas Budha Kemenag Kota Pematangsiantar	0	1	1	0	2
Ditjen Pendidikan Islam Kemenag Kota Pematangsiantar	0	0	28	18	46
Ditjen Penyelenggaraan Haji dan Umrah Kemenag Kota Pematangsiantar	0	0	3	0	3
Kantor Kementerian Agama Kota Pematangsiantar – SETJEN	0	3	9	3	15

In the data processing process, the data that has been obtained from the Central Statistics

Agency (BPS) will be processed according to the available data. After the data is processed, the clustering process will be carried out on which are available data. The data that has been processed are as follows:

Table 2. Data on the Number of Civil Servants by Agencies and Groups in Pematangsiantar in 2019 that have been processed

Satuan/ Unit Kerja/*Di-vision*	**Golongan I**	**Golongan II**	**Golongan III**	**Golongan IV**	**Jumlah/Total**
LAPAS Kelas II Pematangsiantar	0	79	57	1	0
Imigrasi Pematangsiantar	0	1	42	1	0
BPS Kota Pematangsiantar	0	2	11	2	0
Kejari I Pematangsiantar	0	7	23	3	0
Pengadilan Negeri Kelas IB P.Siantar	0	3	26	4	0
Kantor Pertanahan Kota Pematangsiantar	1	7	18	2	1
BP DAS Asahan Barumun	0	4	32	1	0
KPPN Pematangsiantar	0	1	23	1	0
MIN Bah Kapul	0	3	6	8	0
MTsN.Pematangsiantar	0	7	28	11	0
Balai Diklat Kehutanan	0	14	32	10	0
KPP Pematangsiantar	0	38	48	3	0
KPKNL Pematangsiantar	0	12	17	2	0
KPBC Pematangsiantar	0	35	21	1	0
Pengadilan Agama Pematangsiantar	0	0	12	2	0
MAN Pematangsiantar	0	1	18	7	0
KPU Kota Pematangsiantar	0	2	9	0	0
Kanwil DJP Sumatera Utara II	0	37	51	13	0
LP Narkotika	0	46	12	2	0
Ditjen Bimas Islam Kemenag Kota Pematangsiantar	0	1	23	2	0
Ditjen Bimas Kristen Kemenag Kota Pematangsiantar	0	0	20	33	0
Ditjen Bimas Khatolik Kemenag Kota Pematangsiantar	0	1	6	5	0
Ditjen Bimas Budha Kemenag Kota Pematangsiantar	0	1	1	0	0
Ditjen Pendidikan Islam Kemenag	0	0	28	18	0

Satuan/ Unit Kerja/*Di-vision*	**Golo-ngan I**	**Golo-ngan II**	**Golo-ngan III**	**Golo-ngan IV**	**Juml-ah/Total**
Kota Pematan-gsiantar					
Ditjen Penyelen-ggaraan Haji dan Umrah Kemenag Kota Pematan-gsiantar	0	0	3	0	0
Kantor Kemente-rian Agama Kota Pematan-gsiantar – SETJEN	0	3	9	3	0

Furthermore, in the process of clustering data on Central Civil Servants by Agencies and Groups in Pematangsiantar, it is done by creating 3 clusters. The three (3) clusters used are high, medium, and low. The process of grouping the data is done by using the Euclidean Distance formula.

3. Result and Discussion

The calculation process on PNS data (Civil Civil Servants) according to Agencies and Groups in Pematangsiantar is applied using the K-Means Algorithm. After processing the data, the available data will enter the clustering or grouping process. Grouping of data is done by dividing it into 3 clusters. The 3 clusters are the level of civil servants according to agencies and groups in Pematangsiantar high (C1), the level of civil servants according to agencies and groups in Pematangsiantar moderate (C2), the level of civil servants according to agencies and low Pematangsiantar Group (C3). The data grouping is also done using Rapidminer tools. The initial centroid applied is in Table 3 and the calculation results in Iteration-1 are in Table 4.

Table 3. Initial Centroid

***Centroid* Awal**				
Lapas Kelas II Pematangsiantar	0	79	57	1
KPP Pematangsiantar	0	38	48	3
Ditjen Bimas Budha Kemenag Kota Pematangsiantar	0	1	1	1

Table 4. 1st iteration

Satuan/Unit Kerja/*Division*	**C1**	**C2**	**C3**	***Cluster***
LAPAS Kelas II Pematangsiantar	0	42,0238027 8	96,02603814	C1
Imigrasi Pematangsiantar	79,42921377	37,5366487 6	41,01219331	C2
BPS Kota Pematangsiantar	89,69949833	51,6333225 7	10,24695077	C3
Kejari I Pematangsiantar	79,64923101	39,8246155	23	C3
Pengadilan Negeri Kelas IB P.Siantar	82,13403679	41,3521462 6	25,3968502	C3
Kantor Pertanahan Kota Pematangsiantar	81,89627586	43,1624837 1	18,16590212	C3
BP DAS Asahan Barumun	79,0569415	37,6297754 4	31,1608729	C3
KPPN Pematangsiantar	85,08818954	44,6989932 8	22,02271555	C3
MIN Bah Kapul	91,79324594	54,8999089 3	9,643650761	C3
MTsN.Pemata-ngsiantar	78,26237921	37,7491721 8	29,7657 5213	C3
Balai Diklat Kehutanan	70,22107946	29,6816441 6	35,07135583	C2
KPP Pematangsiantar	42,02380278	0	59,89156869	C2
KPKNL Pematangsiantar	78,03845206	40,4722126 9	19,5192213	C3
KPBC Pematangsiantar	56,85068161	27,2396769 4	39,45883931	C2
Pengadilan Agama	90,92304438	52,3545604	11,2249 7216	C3

Satuan/Unit Kerja/*Division*	**C1**	**C2**	**C3**	***Cluster***
Pematangsiantar		5		
MAN Pematangsiantar	87,412 81371	47,80 16736 1	18,384 77631	C3
KPU Kota Pematangsiantar	90,741 39078	53,16 01354 4	8,0622 57748	C3
Kanwil DJP Sumatera Utara II	44,090 81537	10,48 80884 8	62,968 24597	C2
LP Narkotika	55,812 18505	36,89 17334 9	46,368 09248	C2
Ditjen Bimas Islam Kemenag Kota Pematangsiantar	85,094 0656	44,66 54228 7	22,090 72203	C3
Ditjen Bimas Kristen Kemenag Kota Pematangsiantar	92,919 31984	55,92 85258 2	38,091 99391	C3
Ditjen Bimas Khatolik Kemenag KotaPematang-siantar	93,279 15094	56,00 89278 6	7,0710 67812	C3
Ditjen Bimas Budha Kemenag Kota Pematangsiantar	96,026 03814	59,89 15686 9	0	C3
Ditjen Pendidikan Islam Kemenag KotaPematang-siantar	85,854 52813	45,48 62616 6	32,465 36616	C3
Ditjen Penyelenggaraan Haji dan Umrah Kemenag Kota Pematangsiantar	95,697 43988	58,97 45707 9	2,2360 67977	C3
Kantor Kementerian Agama Kota Pematangsiantar –SETJEN	89,911 06717	52,40 22900 3	8,7749 64387	C3

The distance from the calculation results will be compared and the closest distance between the data and the cluster center is chosen, this distance indicates that the data is in one group with the nearest cluster center. From the iteration-1 calculation results, the closest distance to the data center is 1 item in cluster 1 (C1), 6 items in cluster 2 (C2), and 19 items in cluster 3 (C3) and will produce a new centroid. So, in finding the new centroid using the formula, namely:

$$\frac{x_1 + x_2 + x_3 + ... + x_n}{n} \quad (4)$$

The K-Means Algorithm process will stop iterating until the data clustering value is the same as the previous data grouping value. Next, look for the middle value or centroid, the same process is carried out by finding the closest distance. The process of searching for the shortest distance, grouping the data in the last iteration stops at Iteration-4 and can be seen in Table 5.

Table 5. 4th iteration

Satuan/Unit Kerja/*Division*	**C1**	**C2**	**C3**	***Cluster***
LAPAS Kelas II Pematangsiantar	0	46,79 81035 9	85,399 91218	C1
Imigrasi Pematangsiantar	79,429 2138	39,23 08870 7	24,808 16398	C3
BPS Kota Pematangsiantar	89,699 4983	43,13 42381 4	7,8067 27868	C3
Kejari I Pematangsiantar	79,649 231	33,57 17515 2	6,9674 24201	C3
Pengadilan Negeri Kelas IB P.Siantar	81,896 2759	35,46 21276 9	5,3240 02254	C3
Kantor Pertanahan Kota Pematangsiantar	79,056 9415	35,21 45211 5	15,021 48461	C3
BP DAS Asahan Barumun	85,088 1895	39,47 23004 1	7,4124 8946	C3
KPPN Pematangsiantar	91,793 2459	45,11 72084 7	11,976 85268	C3
MIN Bah Kapul	78,262 3792	32,98 57924	12,117 96187	C3
MTsN.Pemata-ngsiantar	70,221 0795	25,56 48684 7	18,331 53021	C3
Balai Diklat Kehutanan	42,023 8028	15,13 48108 7	46,113 39285	C2
KPP Pematangsiantar	78,038 4521	31,50 49599 3	9,4839 33783	C3
KPKNL Pematangsiantar	56,850 6816	13,19 32748	32,210 94534	C2
KPBC Pematangsiantar	90,923 0444	44,37 97532 7	7,5924 30441	C3
Pengadilan Agama Pematangsiantar	87,412 8137	40,91 53088 7	2,6542 41888	C3

Satuan/Unit Kerja/*Division*	**C1**	**C2**	**C3**	***Cluster***
MAN Pematangsiantar	90,741 3908	44,35 72147 5	10,523 54503	C3
KPU Kota Pematangsiantar	44,090 8154	19,90 13190 5	47,901 40917	C2
Kanwil DJP Sumatera Utara II	55,812 185	22,30 61090 3	43,244 01693	C2
LP Narkotika	85,094 0656	39,38 98781 4	6,8223 89611	C3
Ditjen Bimas Islam Kemenag Kota Pematangsiantar	92,919 3198	49,88 04821 5	27,590 66871	C3
Ditjen Bimas Kristen Kemenag Kota Pematangsiantar	93,279 1509	46,61 61184 6	11,993 53993	C3
Ditjen Bimas Khatolik Kemenag KotaPematang-siantar	96,026 0381	49,90 55357 7	17,842 2252	C3
Ditjen Bimas Budha Kemenag Kota Pematangsiantar	85,854 5281	41,49 17160 4	16,347 6298	C3
Ditjen Pendidikan Islam Kemenag KotaPematang-siantar	81,896 2759	35,46 21276 9	5,3240 02254	C3
Ditjen Penyelenggaraan Haji dan Umrah Kemenag Kota Pematangsiantar	95,697 4399	49,43 24033 4	16,153 79212	C3
Kantor Kementerian Agama Kota Pematangsiantar –SETJEN	89,911 0672	43,30 19918 7	9,1621 50403	C3

From the iteration-4 calculation, the closest distance to the data center is 1 item in cluster 1 (C1), 4 items in cluster 2 (C2), and 21 items in cluster 3 (C3).

4. Conclusion

Based on the calculation of all the iterations that have been carried out, the authors conclude that by using the K-Means Algorithm, it can complete and find out the clustering or grouping of data on Central Civil Servants (PNS) by Agencies and Groups in Pematangsiantar.

From the results of the analysis, it is known that the level of Central Civil Servants (PNS) by Agencies and Groups in Pematangsiantar is classified as high, the level of Central Civil Servants (PNS) by Agencies and Groups in Pematangsiantar is classified as moderate, and the level of Civil Servants (PNS) Centers According to Agencies and Groups in Pematangsiantar which are classified as grouping results are 1 item in cluster 1 (C1), 4 items in cluster 2 (C2), and 21 items in cluster 3 (C3).

References

Aditya, A., Jovian, I., & Sari, B. N. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019. *Jurnal Media Informatika Budidarma*, *4*(1), 51. https://doi.org/10.30865/mib.v4i1.1784

Agustin, F. E. M. (2015). Implementasi Algoritma K-Means Untuk Menentukan Kelompok Pengayaan Materi Mata Pelajaran Ujian Nasional (Studi Kasus: Smp Negeri 101 Jakarta). *Jurnal Teknik Informatika*, *8*(1), 73–78. https://doi.org/10.15408/jti.v8i1.1938

Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *5*(1), 17–24. https://doi.org/10.25077/teknosi.v5i1.2019.17-24

Irnanda, K. F., Windarto, A. P., Damanik, I. S., & ... (2019). Penerapan K-Means pada Proporsi Individu dengan Keterampilan (Teknologi Informasi dan Komunikasi) TIK Menurut Wilayah. *... Teknologi Informasi ...*, *c*, 452–456.

Maulida, L. (2018). Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan K-Means. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *2*(3), 167. https://doi.org/10.14421/jiska.2018.23-06

Nasution, Y. R., & Eka, M. (2018). Penerapan Algoritma K-Means Clustering Pada Aplikasi Menentukan Berat Badan Ideal. *Jurnal Ilmu Komputer Dan Informatika*, *02*(April), 77–81.

Nurzahputra, A., Muslim, M. A., & Khusniati, M. (2017). Penerapan Algoritma K-Means Untuk Clustering Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa. *Techno.Com*, *16*(1), 17–24. https://doi.org/10.33633/tc.v16i1.1284

Parlina, I., Windarto, A. P., Wanto, A., & Lubis, M. R. (2018). Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Asesment Center. *Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Asessment Center Untuk*

Clustering Program Sdp, *3*(1), 87–93.

Rosmini, R., Fadlil, A., & Sunardi, S. (2018). Implementasi Metode K-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah. *It Journal Research and Development*, *3*(1), 22–31. https://doi.org/10.25299/itjrd.2018.vol3(1).1773

Wardhani, A. K. (2016). Implementasi Algoritma K-Means untuk Pengelompokkan Penyakit Pasien pada Puskesmas Kajen Pekalongan. *Jurnal Transformatika*, *14*(1), 30–37.