Journal of Intelligent Decision Support System (IDSS) 6 (4) (2023), pp. 231-240

Published by: Institute of Computer Science (IOCS)

Journal of Intelligent Decision Support System (IDSS)

Journal homepage: www.idss.iocspublisher.org

Analysis Of Salary Of Permanent Employees And Contract Employees On The Medicom Campus Using The K – Means Algorithm

Leliana Harahap[1], Sartika Dewi Purba[2], Sutrisno Situmorang[3], Jonas Franky R Panggabean[4], Kamson Sirait[5]

[1,2,3,4,5] Akademi Informatika dan Komputer Medicom Medan

Article Info

Article history:

Received Nov 20, 2023
Revised Nov 29, 2023
Accepted Dec 12, 2023

Keywords:

Analysis;
Clustering;
Employees;
K-Means Algorithm;
Salary.

ABSTRACT

The Medicom campus is a place of work that provides jobs to the community. In work, employee status cannot be separated, namely permanent employees and contract employees. In employee status, employee salaries can be determined. In determining employee salaries, there are several problems that can disrupt employee performance at work. For this reason, a method is needed to determine employee salaries. One method that can be used is the K-Means Clustering Algorithm. Which is considered quite effective in determining the suitability of salaries for permanent employees and contract employees. By creating clusters to make it easier for finance workers to record and determine and adjust employee salaries based on their status.

Corresponding Author:

Sartika Dewi Purba,
Komputerisasi Akuntansi,
Akademi Informatika dan Komputer Medicom,
Jl.Darat No.74, Medan, Sumatera Utara, 20153, Indonesia
Email: sartikadewipurba85@gmail.com

Introduction

Work is an activity that can be done without needing to have any special expertise or skills. The aim of doing this work is to earn income to meet daily needs. Everyone can work to meet their needs, so work must be done by everyone.

In an organization or business, worker classification is needed to determine the skill/professional level of workers which is useful for improving the quality of workers in an organization or company. Ability is a skill that a person has that enables a person to be able or unable to do the job (Arifin, 2020).

From the explanation above, it can be concluded that determining employee salaries is a process of allocating resources according to their skills to achieve the best goals. Therefore, a comparison of the salaries of permanent employees and contract employees can be determined.

At the AMIK Medicom Campus, finding employee salaries is still done manually, so that errors can be found in the process of receiving employee salaries, which can make employees complain about receiving salaries, therefore a technique is needed to determine fixed salaries and employee contracts. One technique that can be used to overcome the above problem is the K-Means Clustering Algorithm. An algorithm is a sequence or steps to carry out calculations or can also be used and implemented to solve problems written sequentially (Rustam, 2020). The K-Means algorithm is an iterative clustering algorithm that partitions a data set into a number of clusters that have been determined at the beginning (Putri, 2020). As objects that have the same characteristics are grouped into one cluster and

objects that have different characteristics are grouped into another cluster (Syafnidawaty, 2020). One method of clustering that is more famous than other clustering algorithms is K-means because of the simplicity of the algorithm and its efficiency (Ansari, 2019)

The cluster algorithm method is suitable for use in segmentation, such as in the customer segmentation process for a supermarket carried out by sangkar (Rajagopal, 2019). This technique was also used in previous research to determine different surveys with algorithms based on spatial data sets (Vaghela, 2019). Data mining with a Distributed and energy efficient System using the Clustering Method for hierarchical Wireless Sensor NetworksHuang as research conducted by previous research (Sulastri H, 2021).

Based on analysis and testing with the rapid miner at the AMIK Medicom Campus, a cluster method can be produced so that it can help group employee salary data and the results of the testing can be obtained from the highest salary and lowest salary for employees using the K-means algorithm to make it easier to search for data. in helping to classify employee salaries.

Method

a. Knowledge Discovery In Database (KDD)

Knowledge discovery in databases (KDD) is a computer science terminology that refers to the process of searching for useful information from large collections of digital data, such as databases or datasets.

KDD has a handful of processes that include: (a) Data preparation, (b) Data selection, (c) Data cleaning, (d) Application of information to datasets, (f) Interpretation of results

The main goal of KDD is to extract high-level information from low-level data. In other words, the KDD process focuses on mapping low-level data and transforming it into a more concise, clear and meaningful form. Discover unexpected relationships and can summarize data in new ways so that it can be useful and understandable for users (R. Mythily, 2021).

This KDD method has been used in the data industry for more than a decade, to be precise since 1989 by Piatetsky -Shapiro. Initially, data mining and KDD were done manually. However, with data volumes increasing to terabytes, this process has become too difficult to do manually. As a result, there is now a lot of software and tools available that can help uncover hidden data and formulate assumptions and hypotheses—which are the foundation of the application of artificial intelligence. With the development of technology, knowledge discovery in databases now includes many multidisciplinary elements.

b. Definition of Data Mining

Data Mining is a process that uses computer engineering to be able to perform data analysis and knowledge extraction automatically according to Fajar Astuti Hermawati (Nurhayati, 2018). Data mining is a unit of process of dredging or collecting important information from large data units. The data mining process always uses statistical methods, mathematics, and even utilizes artificial intelligence technology (Adani, 2021). Data mining is a process using artificial intelligence, statistical techniques, mathematics, and machine learning to identify and extract useful information and related knowledge from various large databases (Gurunescu, 2017). Clustering is one of the methods of data mining (Gullo, 2018) and clustering has become a valid instrument for solving complex problems in computer science and statistics (Tagarelli, 2018).

c. Data Mining Stages

Stages carried out in the Data Mining process (Firdaus, 2019): (a) Data Cleaning, (b) Data Integration, (c) Data Transpormation, (d) Application of Data Mining Technique (f) Interpretation / Evaluation (g) Knowledge.

d. Definition of Clustering

Clustering or clustering is a method of grouping data. According to Tan, 2006 clustering is a process for grouping data into several clusters or groups so that data in one cluster has the maximum

level of similarity and data between clusters has a minimum similarity. Clustering is the process of partitioning a set of data objects into subsets called clusters. Objects in a cluster have similar characteristics to each other and are different from other clusters. A decision tree is an elaboration question, while the leaves reflect the classes or groups (Mubarok, 2019). Clustering is the process of grouping data points into two or more groups so that data points belonging to the same group are more similar to each other than to different groups, based only on the information available with the data points (N. Nidheesh, 2018).

e. K-Means algorithm

The main principle of this technique is to arrange k partitions/centroids/means from a set of data. The K-means algorithm starts by forming cluster partitions at the beginning, then this cluster partition is iteratively improved until there are no significant changes to the cluster partitions according to Witten, Eibe, & Hall (Butarbutar, 2019). The algorithm is very easy to implement and execute, fast and can be used according to Wu & Kumar's needs (Butarbutar, 2019)

The K-Means Clustering algorithm is one of the algorithms used in data clustering analysis. The definition is an algorithm that aims to group data into similar groups, based on the similarity of certain attributes or characteristics. In this algorithm, each group will be represented by a centroid which is the center point of the cluster. The use of the K-Means algorithm aims to make it easier to understand data systematically.

The research framework is the stages that must be carried out by researchers in conducting research. The following are the stages of the research framework for the K-Means algorithm method. The research methodology can be seen in Figure 1 below:

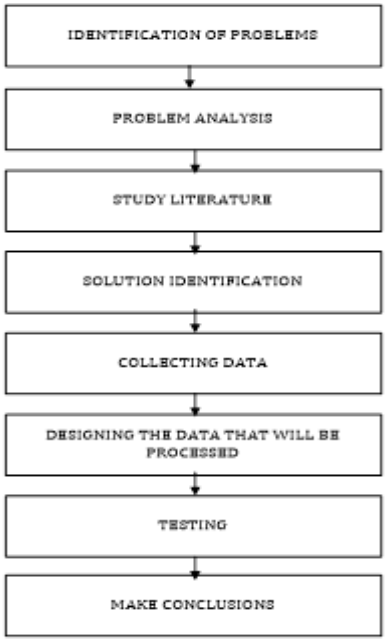

Figure 1. Research Stages

Results and Discussions

a. Data Analysis

When collecting data, the first thing to do is an interview technique while asking for data from the personnel department (N. Y. S. Munti, 2019). There are several factors that can influence employee salaries, one of which is internal factors which will be discussed in this research which will be grouped so that they can help the process of understanding employee salary data and data taken from the AMIK Medicom campus.

b. Clustering Analysis Using the K-Means Algorithm

Cluster analysis means creating groups of data objects based on existing information and their relationships. In the picture below are the algorithm stages:

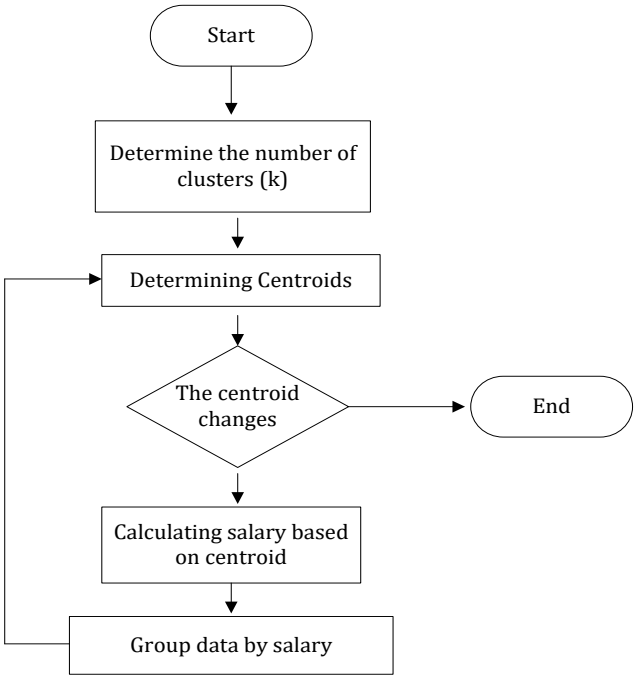

Figure 2. K-Means Process Flowchart Diagram

Figure 2 is a flowchart of the K-Means process. The initial step is to determine the number of Clusters (K), then determine the centroid. When testing the centroid, if the answer is Yes then the salary is calculated based on the centroid, if not then the process is end. And the final step is to group the data based on the highest and lowest salaries.

c. Data Determination

In data grouping, you must first determine the data that will be processed so that the goal is achieved.

d. Determine the Attributes Used

In this research, the factors that influence the salaries of permanent employees and contract employees are: (a) Employee position, (b) Employee status, (c) Employee basic salary, (d) Employee benefits, (f) Employee Overtime, (g) Total employee salary.

e. Data Coding

Employee salaries are influenced by several factors that have been determined as tools. There are 30 employee data, 19 permanent employees and 11 contract employees.

Table 1: Sample Initial Data for Permanent Employees and Contract Employees

ID	Name	Gender	Position	Status	Basic salary	Benefits	Over Time	Total Salary
07B10034711	SUYANTO	L	Staff Adm	TETAP	Rp 3.200.000	Rp 1.000.000	Rp100.000	Rp 4.300.000
97B00372445	DENI ARIANTO	L	Teknisi	TETAP	Rp 2.900.000	Rp 1.000.000	Rp100.000	Rp 4.000.000
93B00239668	SALMAN	L	Satpam	TETAP	Rp 3.000.000	Rp 1.000.000	Rp100.000	Rp 4.100.000

90B02063963	TOMSON MANALU	L	TU	KONTRAK	Rp 2.850.000	Rp 650.000	Rp 65.000	Rp 3.565.000
96B00394094	LAMRIA NABABAN	P	Supir	KONTRAK	Rp 2.550.000	Rp 650.000	Rp 65.000	Rp 3.265.000
89B00272022	PRANTO SINAGA	L	Kebersihan	TETAP	Rp 2.700.000	Rp 1.000.000	Rp100.000	Rp 3.800.000
98B10027110	ANJUR PANEM PARDEDE	L	IT	TETAP	Rp 3.200.000	Rp 1.000.000	Rp100.000	Rp 4.300.000
89B00276437	PUKKA SIHOMBING	L	Staff Adm	TETAP	Rp 3.200.000	Rp 1.000.000	Rp100.000	Rp 4.300.000
91B02272091	PIATOR MAHARAJA	L	Teknisi	TETAP	Rp 2.900.000	Rp 1.000.000	Rp100.000	Rp 4.000.000
92B02299274	ANDI	L	Satpam	KONTRAK	Rp 2.850.000	Rp 650.000	Rp 65.000	Rp 3.565.000
98B10028548	SAMIN	L	TU	TETAP	Rp 2.900.000	Rp 1.000.000	Rp100.000	Rp 4.000.000
97B00234520	EKA SRI AGUSTINI	P	Supir	TETAP	Rp 3.100.000	Rp 1.000.000	Rp100.000	Rp 4.200.000
97B00234447	MUHAMMAD IRSYAD NASUTION	L	Kebersihan	TETAP	Rp 2.700.000	Rp 1.000.000	Rp100.000	Rp 3.800.000
94B00657933	YOSEP SIALLAGAN	L	IT	TETAP	Rp 3.200.000	Rp 1.000.000	Rp100.000	Rp 4.300.000
97B00372262	SUTARJO	L	Staff Adm	KONTRAK	Rp 2.750.000	Rp 650.000	Rp 65.000	Rp 3.465.000
97B00234579	EDI SISWANTO	L	Teknisi	KONTRAK	Rp 2.350.000	Rp 650.000	Rp 65.000	Rp 3.065.000
91B02272372	JUNAIDI PURBA	L	Satpam	TETAP	Rp 3.000.000	Rp 1.000.000	Rp100.000	Rp 4.100.000
95B00470102	RALIONO	L	TU	KONTRAK	Rp 2.850.000	Rp 650.000	Rp 65.000	Rp 3.565.000
90B02195476	RUSMEN	L	Supir	TETAP	Rp 3.100.000	Rp 1.000.000	Rp100.000	Rp 4.200.000
95B00445740	ANTON SINAGA	L	Kebersihan	KONTRAK	Rp 2.650.000	Rp 650.000	Rp 65.000	Rp 3.365.000
98B10027920	SUARSINI	P	IT	TETAP	Rp 3.200.000	Rp 1.000.000	Rp100.000	Rp 4.300.000
90B02063831	ABDUL MUIS	L	TU	TETAP	Rp 2.900.000	Rp 1.000.000	Rp100.000	Rp 4.000.000
94B00657271	HOTMAN GARINGGING	L	Supir	KONTRAK	Rp 2.350.000	Rp 650.000	Rp 65.000	Rp 3.065.000
96B00392353	SUSILAWATI	P	Kebersihan	KONTRAK	Rp 2.850.000	Rp 650.000	Rp 65.000	Rp 3.565.000
83B02001687	DEWI MARPAUNG	P	IT	KONTRAK	Rp 2.550.000	Rp 650.000	Rp 65.000	Rp 3.265.000
98B10030056	RAMLAN SAING	L	Staff Adm	TETAP	Rp 3.200.000	Rp 1.000.000	Rp100.000	Rp 4.300.000
7B0 800623	ARIANTO	L	Staff Adm	TETAP	Rp 3.200.000	Rp 1.000.000	Rp100.000	Rp 4.300.000
97B00357925	DEWI KARTIKA	P	Staff Adm	KONTRAK	Rp 2.850.000	Rp 650.000	Rp 65.000	Rp 3.565.000
94B00656612	FIKSON PAKPAHAN	L	IT	TETAP	Rp 3.200.000	Rp 1.000.000	Rp100.000	Rp 4.300.000
98B10029660	DAKWA	L	Kebersihan	TETAP	Rp 2.700.000	Rp 1.000.000	Rp100.000	Rp 3.800.000

Table 2: Table of Total Costs for Permanent Employees and Contract Employees

Code	Position	Basic Salary		Benefits		Overtime / Hours	
		Permanent	Contract	Permanent	Contract	Permanent	Contract
1	Staff Adm	Rp3.200.000	Rp2.850.000	Rp 1.000.000	Rp 650.000	Rp 100.000	Rp 65.000
2	Teknisi	Rp2.900.000	Rp2.550.000	Rp 1.000.000	Rp 650.000	Rp 100.000	Rp 65.000
3	Satpam	Rp3.000.000	Rp2.650.000	Rp 1.000.000	Rp 650.000	Rp 100.000	Rp 65.000
4	TU	Rp2.900.000	Rp2.550.000	Rp 1.000.000	Rp 650.000	Rp 100.000	Rp 65.000
5	Supir	Rp3.100.000	Rp2.750.000	Rp 1.000.000	Rp 650.000	Rp 100.000	Rp 65.000

6	Kebersihan	Rp2.700.000	Rp2.350.000	Rp 1.000.000	Rp 650.000	Rp 100.000	Rp 65.000
7	IT	Rp3.200.000	Rp2.850.000	Rp 1.000.000	Rp 650.000	Rp 100.000	Rp 65.000

f. Data Transformation

Data transformation is carried out because the data type is not numeric, so the data must be transformed first by frequencying the data that appears the most by sorting the frequency from highest to lowest and initializing the data.

Table 3. Sample of employee data for the Clustering process

ID	Name	Gender	Position	Status	Basic salary	Benefits	Over Time
07B10034711	SUYANTO	1	1	1	11	21	31
97B00372445	DENI ARIANTO	1	2	1	12	22	32
93B00239668	SALMAN	1	3	1	13	23	33
90B02063963	TOMSON MANALU	1	4	2	111	221	331
96B00394094	LAMRIA NABABAN	2	5	2	112	222	332
89B00272022	PRANTO SINAGA	1	6	1	16	26	36
98B10027110	ANJUR PANEM PARDEDE	1	7	1	17	27	37
89B00276437	PUKKA SIHOMBING	1	1	1	11	21	31
91B02272091	PIATOR MAHARAJA	1	2	1	12	22	32
92B02299274	ANDI	1	3	2	117	227	337
98B10028548	SAMIN	1	4	1	14	24	34
97B00234520	EKA SRI AGUSTINI	2	5	1	15	25	35
97B00234447	MUHAMMAD IRSYAD NASUTION	1	6	1	16	26	36
94B00657933	YOSEP SIALLAGAN	1	7	1	17	27	37
97B00372262	SUTARJO	1	1	2	115	225	335
97B00234579	EDI SISWANTO	1	2	2	116	226	336
91B02272372	JUNAIDI PURBA	1	3	1	13	23	33
95B00470102	RALIONO	1	4	2	111	221	331
90B02195476	RUSMEN	1	5	1	15	25	35
95B00445740	ANTON SINAGA	1	6	2	113	223	333
98B10027920	SUARSINI	2	7	1	17	27	37
90B02063831	ABDUL MUIS	1	4	1	14	24	34
94B00657271	HOTMAN GARINGGING	1	5	2	116	226	336
96B00392353	SUSILAWATI	2	6	2	117	227	337
83B02001687	DEWI MARPAUNG	2	7	2	114	224	334
98B10030056	RAMLAN SAING	1	1	1	11	21	31
97B00358006	ARIANTO	1	1	1	11	21	31
97B00357925	DEWI KARTIKA	2	1	2	117	227	337
94B00656612	FIKSON PAKPAHAN	1	7	1	17	27	37
98B10029660	DAKWA	1	6	1	16	26	36

Table 4. Clustering Code

Position	Code	Basic Salary		Benefits		Over Time	
		Permanent	Contract	Permanent	Contract	Permanent	Contract
Staff Adm	1	11	111	21	221	31	331
Teknisi	2	12	112	22	222	32	332
Satpam	3	13	113	23	223	33	333
TU	4	14	114	24	224	34	334
Supir	5	15	115	25	225	35	335
Kebersihan	6	16	116	26	226	36	336
IT	7	17	117	27	227	37	337

g. Clustering Process Using the K-Means Algorithm

The data that has been determined will be processed using a clustering process using the K-Means algorithm so that the best grouping results are obtained which will later help in understanding employee data.

There is some data used to carry out the clustering process, namely: (a) Determine the number of clusters from employee salaries of 5 clusters (k = 4), (b) Determine the centroid, taken at random

Center calculates the distance between centroid points using the points of each object using Euclidian Distance. There is an initial centroid calculation manually. The calculation is as follows:

$$D\,(i,j) = \sqrt{(X1i - X1j)^{\,2} + (X2i - X2j)^2 + .. + (Xkn - Xkj)^{\,2}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

Where:
x = cluster center
y = data
Where x is the first cluster center data and y data is the cluster data that will be tested using cluster center data.

Table 5 employee centroids

97B00234447	MUHAMMAD IRSYAD NASUTION	1	6	1	16	26	36
94B00657933	YOSEP SIALLAGAN	1	7	1	17	27	37

After analyzing the data as explained in the previous chapter, the next step is to carry out implementation. Implementation is carried out using the Rapid Miner 6.2 application which will help make it easier to group salaries of permanent and contract employees which will help in the decision making process.

a. Clustering Test Results
Clustering test results using the Rapid Miner 7.5 application with 5 clusters

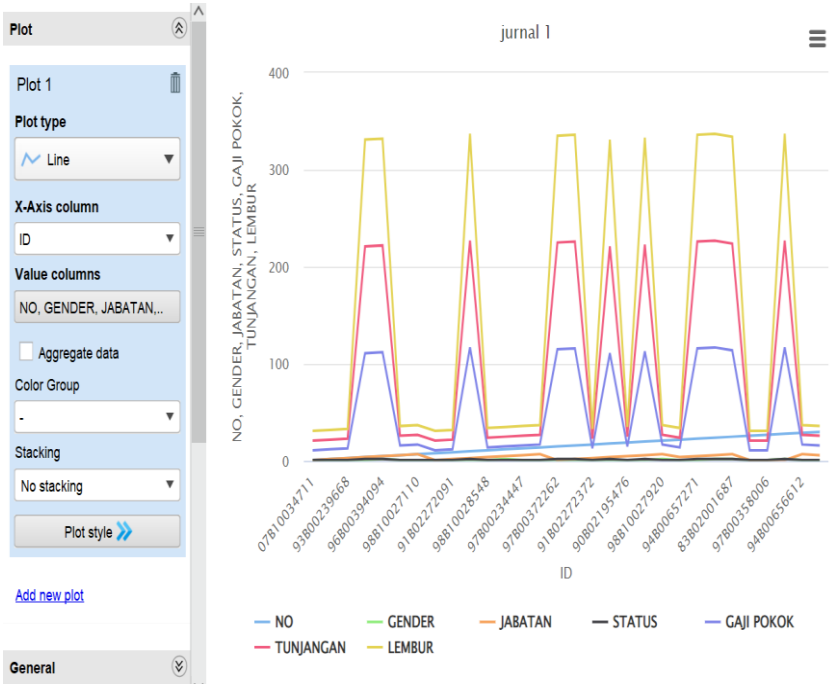

Figure 3. Rapid miner results with line diagram

Figure 3 Rapid Miner results with a line diagram which is the result of clustering testing using the Rapid Miner application with attributes Number, Gender, Position, Status, Allowances and overtime.

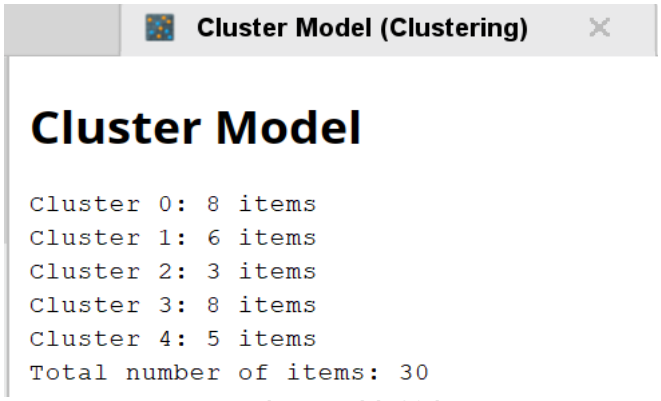

Figure 4. Cluster model of 5 clusters

Figure 4 explains the cluster model which consists of 5 clusters, namely: data cluster 0 (gender) = 8 items, cluster 1 (position) = 6 items, cluster 2 (status)) = 3 items, cluster 3 (allowances) = 8 items and cluster 4 (overtime) = 5 items. The test results can be obtained from the employee's highest salary and lowest salary. Using the K-Means algorithm can make data searching easier.

Figure 5. Root set cluster

Figure 5 Root set cluster obtained from rapidminer which produces Cluster 0=8 items, Cluster 1=6 items, Cluster 2=3 items, Cluster 3=8 items and Cluster 8=5 items with a total of 30 items.

Cluster Model (Clustering)

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4
NO	21.125	10.500	6.333	23.875	4.600
GENDER	1.375	1.167	1.333	1.125	1
JABATAN	4	5.833	4	4.250	1.800
STATUS	2	1	2	1	1
	114.875	15.833	113.333	14.250	11.800
TUNJANGAN	224.875	25.833	223.333	24.250	21.800
LEMBUR	334.875	35.833	333.333	34.250	31.800

Figure 6. Centroid value

Figure 6 centroid value with attributes Number, Gender, Position, Status, Allowances and overtime with 5 clusters, namely cluster 0, cluster 1, cluster 2, cluster 3 and cluster 4

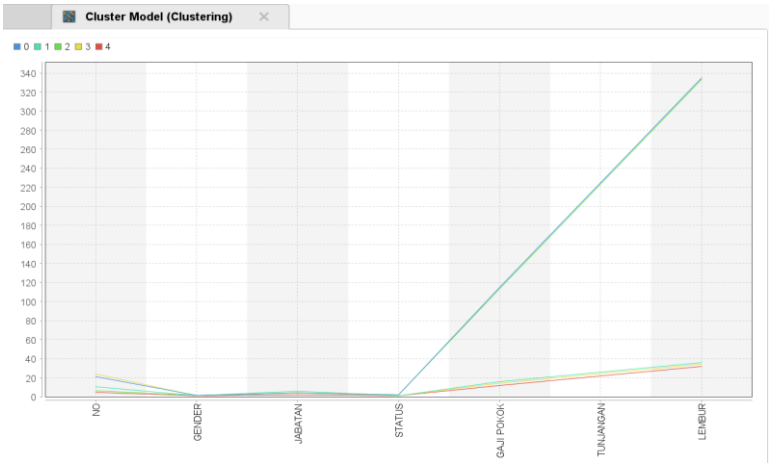

Figure 7. Clustering plot

Figure 7 Clustering plot is a graph of grouping data based on gender, position, status, basic salary, allowances and employee overtime.

Conclusions

Based on the results of the analysis and testing of the problem, it can be concluded that the cluster method can help group employee salaries using 5 clusters of the 30 items used, with data cluster 0 (gender) = 8 items, cluster 1 (position) = 6 items, cluster 2 (status)) = 3 items, cluster 3 (allowances) = 8 items and cluster 4 (overtime) = 5 items. The test results can be obtained from the employee's highest salary and lowest salary. Using the K-Means algorithm can make data searching easier. Using the rapid miner application can help classify employee salaries. For development in future research, it is hoped that researchers will have a better understanding of the factors that influence the high and low salaries of employees, so that the understanding of employee data is broader. In further research development, it can be carried out with even larger amounts of data. The RapidMiner application is software that is very easy to use and can be used to help group data. RipadMiner software is used in a wider field because using this software in daily work applications can help work activities.

Acknowledgements

Thanks to AMIK Medicom, Foundation, Director, Management and all AMIK Medicom employees at Jalan Darat No.74 Medan

References

Adani, M. R. (2021). Mengenal Konsep Data Mining Beserta Contoh Implementasinya.

Ansari, T. H. (2019). An analysis of MapReduce efficiency in document clustering using parallel K-means algorithm. *Futur. Comput. Informatics J*, 1-10.

Arifin, M. (2020). Peran Lingkungan Kerja, Kemampuan Kerja, Dan Karakteristik Individu Terhadap Kinerja Karyawan Pada Pt. Buana Finance Tbk. Cabang Manado. *Jurnal EMBA: Jurnal Riset Ekonomi, Manajemen, Bisnis Dan Akuntansi*, 31-40.

Butarbutar, N. e. (2019). Komparasi Kinerja Algoritma Fuzzy C-Means Dan K-Means Dalam Pengelompokan Data Siswa Berdasarkan Prestasi Nilaiakademik Siswa. *Jurasik (Jurnal Riset Sistem Informasi & Teknik Informatika)*, 46-55.

Firdaus, D. (2019). Penggunaan Data Mining dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer . *Jurnal Format Volume 6*.

Gullo, F. (2018). From patterns in data to knowledge discovery: What data mining can do. *Phys. Procedia*.

Gurunescu, F. (2017). Data Mining Consepts, Models and Techniques. *Verlag Berlin Heidelberg*.

Mubarok, M. I. (2019). *Algoritma C4.5.*

N. Nidheesh, K. A. (2018). An enhanced deterministic K-Means clustering algorithm for cancer subtype prediction from gene expression data. *Comput. Biol. Med*, 213-221.

N. Y. S. Munti, G. W. (2019). Analisis Dan Penerapan Data Mining Untuk Menentukan Gaji Karyawan Tetap Dan Karyawan Kontrak Menggunakan Algoritma K-Means Clustering. *JITI*.

Nurhayati. (2018). Metode Rough Set untuk melihat perilaku suami yang menjadi akseptor KB vasetomi. *Jurnal Teknologi dan Sistem Informasi Ramesh Singh Yadava1*.

Putri, R. U. (2020). Kemampuan Kerja Karyawan dan Tunjangan Kesejahteraan Terhadap Prestasi Kerja Karyawan Pada Bengkel Betawi Diesel. *Mabiska Jurnal*, 1-15.

R. Mythily, A. B. (2021). Clustering models for data stream mining. *Procedia Comput. Sci*, 619–626.

Rajagopal, S. (2019). Customer data Clustering using data mining technique. *Enterprise DW/BI Consultant Tata Consultancy Services*.

Rustam, S. (2020). Penerapan Optimasi Jumlah Kluster Pada K-Means Untuk Pengelompokan Kelas Mata Kuliah Konsentrasi Mahasiswa Semester Akhir. *Jurnal Sistem Informasi dan Teknik Komputer*.

Sulastri H, G. A. (2021). Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia. *Jurnal Teknologi dan Sistem Informasi*, 299-305.

Syafnidawaty. (2020). *K-MEANS CLUSTERING.*

Tagarelli, A. A. (2018). Data Mining: Clustering. *Ref. Modul. Life Sci*.

Vaghela, N. M. (2019). Survey on Different Density Based Algorithms on Spatial Dataset. *International Journal of Advance Research in Computer Science and Management* .