JURNAL TEKNOLOGI DAN OPEN SOURCE
Vol. 5, No. 2, Desember 2022, pp. 115~135
e-ISSN: 2622-1659, accredited Four Grade by Kemenristekdikti, Decree No: 36/E/KPT/2019
DOI: 10.36378/jtos.v3xx

Analysis Of Employee Discipline Based On Digital Attendance With The K-Means Algorithm Method

Yulya Muharmi[1], Sri Nadriati[2]
[1,2]Department of Information System, STMIK Dharmapala Riau, Indonesia

Article Info

Article history:

Received 1103, 2022
Revised 1116, 2022
Accepted 12 03, 2022

Keywords:

Discipline
Analysis
Data Mining
K-Means Algorithm
Rapidminer

ABSTRACT

Employee discipline is one of the most important factors for the progress of the company. PT. Sumatra Core Cellular (PT. SIS) Pekanbaru has implemented a digital attendance application, but the company has not evaluated the application to determine the level of employee discipline. Data mining is the process of extracting useful information from a large database population. One of the data mining methods is the K-Means algorithm. The data mining process uses the method of K-Means algorithm with 2 clusters namely discipline and less disciplined categories. The data used is attendance data of 159 employees, namely data on tardiness, non-attendance (TAP), attendance hours and 4 selected questionnaire questions. Tools for grouping with the Rapidminer application. Using the K-Means algorithm method, it is known that cluster 0 consists of 133 employees or 83.64% with a disciplined category and cluster 1 produces 26 employees or 16.35% with a less disciplined category. Judging from the accuracy of attendance hours, employees in cluster 0 are more likely to be present at 07.45 - 08.15 and in cluster 1 they are more likely to be present at 08.15 - 08.30. In terms of lateness and TAP, there is a lack of discipline in cluster 1. From the level of satisfaction with the application based on 4 selected questions, it can be concluded that the digital attendance application increases the discipline of the employees. The results of this analysis can be used as a reference for evaluating employee discipline, determining promotions and improving employee discipline in the future.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Yulya Muharmi
Department of Information System
STMIK Dharmapala Riau
Pekanbaru, Indonesia
Email: yulya.muharmi@lecturer.stmikdharmapalariau.ac.id
© The Author(s) 2022

1. Introduction

Human resources play a very important role in the organisation because human resources make the other resources work and run. The role of human resources is so important that this factor becomes the determining factor for the back and forth of the organization [1]. So it is not surprising that many companies place human resources or commonly called employees as the main assets of the company and rely heavily on human resources.[2] Utilizing information technology is a must for companies not to be outdated, so naturally currently many companies are competing to update new systems and technologies because to improve their business, especially in fields that are closely related to information technology.[3]

Employees are an intellectual resource for any organization, and are a major contributor to business success [4]. The use of human resources plays an important role that every individual in an organization or a company needs to realize to strive for various ways to make human resources so that the company can

continue to grow, one way to maintain the company's human resources continues to increase is to maintain the work discipline of each individual involved in the organization or company[5]. The quality of human resources will have a significant effect on the achievement of human resource performance and the company's performance in general [6].

Every organisation wants its employees to be able to work optimally, which affects an employee's performance, e.g. age, gender, attendance. Absenteeism itself is the degree to which employees are present in relation to responsibilities. Attendance in relation to responsibility of employees at work. Employees who show up on time, are not late and do not often ask permission when they enter work can be described as disciplined [7]. In general, absenteeism is believed to be a leading indicator of poor performance [8]. Excessive absenteeism can have a serious impact on any organization [9]. Some will see this as a scare tactic that suggests that there will be consequences if absenteeism continues to occur [10]. PT. SIS Pekanbaru has applied information technology in employee attendance in the form of digital attendance. However, the attendance data on the digital attendance application has not been evaluated to see the level of discipline of the employee.

Advances in scientific data collection methods have led to the large-scale collection of promising data on various areas of science and technology [11]. Data mining is the process of discovering interesting knowledge from large amounts of data stored in databases, data warehouses, or other information storage places [12]. The implementation of data mining in this company is able to help companies in managing data owned by companies that initially do not have value become important information to support decision making in the future by paying attention to data patterns and information that can be obtained through data extraction[13]. Data mining modeling in this study adapts to the needs of information and data characteristics in the data warehouse [14]. Data mining is commonly used for classification, clustering, prediction, estimation, association and description of data [15]. One of the data mining methods in grouping the level of discipline of PT. SIS Pekanbaru is a method of the K-Means algorithm.

The K-Means algorithm is a non-hierarchical clustering method that partitions data into clusters where data with the same characteristics will be grouped into one cluster, and data with different characteristics will be grouped in another cluster[16]. Data grouping with the K-Means algorithm is carried out by means of determining the number of clusters, calculating the closest distance to the center of the cluster. The closest distance data represents the members of the cluster, recalculated until the data does not move to another cluster, to minimize the purpose function[17].

The application of the K-Means Algorithm to disciplinary analysis has been carried out by Murni, et al [16], Murni, et al (2022) applying the K-Means Algorithm to group the level of discipline of employees at the Gotting Sidodadi Bandar Pasir village office with data on the highest, medium, and low levels of discipline. Atmojo, et al[7] conducting E-Attendance data analysis at SDN X Taman Fajar and Puskesmas Y using the K-Means Clustering Algorithm.The results of this study produced the results of the analysis of work discipline behavior patterns by looking at the percentage of late and the number of attendance taken from the E-Attendance data of SD Negeri X Taman Fajar, Purbolinggo District and Puskesmas Y located in Negara Nabung Village, East Lampung Regency so that it can be considered for agencies related to increasing the level of work discipline.

The data analysed were data on lateness, non-absence from home (TAP), attendance hours and 4 selected questionnaire questions. The results of the analysis of the grouping of staff discipline levels on PT. SIS using the Rapidminer application consists of two categories, namely the discipline category and the less disciplined category. This grouping aims to find out which employees fall into the discipline category and the less disciplined category. The results of this analysis can be used as a reference in decision making so that those who lack discipline can be given measures to make changes in the future, determine promotion and be valued for employees in the discipline category.

2. Research Method

Data Mining is a term used to describe the discovery of knowledge in a database. Data mining is a process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from various large databases[18]. Data mining is a part of the Knowledge Discovery in Database (KDD) process that serves to extract patterns or examples based on data using a specific algorithm. The KDD process is as follows [19]:

1. *DataSelection*

   In the initial data, the data will be selected first and choose which attributes to use in this study. The attributes used in determining the level of employee discipline are employee name, delay, non-return

absence (TAP), attendance hours and 4 selected questionnaire questions related to the use of digital attendance applications.

2. *Preprocessing/Cleaning*
   After the data selection was carried out at the previous phase, a prepocessing / cleaning process was carried out with the aim of eliminating duplicate data or invalid data. However, if there is no duplicate data or invalid data, it can proceed to the data transformation stage.
3. *Transformation Data*
   Data transformation is carried out because the data type is not numeric, so the data must be transformed first by doing the frequency on the data that appears the most by sorting the highest frequency to the lowest and doing the initials of the data. The data referred to in this study are data on delays, non-absence of return (TAP) and 4 selected questionnaire questions related to the use of this digital attendance application.
4. *Data Mining*
   The next phase is to carry out the data mining process by processing data using the K-Means Algorithm method which in this study will be grouped with 2 clusters which will later produce groupings with disciplinary and less disciplined categories.
5. *Interpretation/Evaluation*
   This phase is a testing stage by applying the Rapidminer application.This Rapidminer application helps in processing data so that the information conveyed is easier to understand and analysis is carried out to find out the level of discipline of employees at PT. SIS Pekanbaru.

Clustering is a useful tool in data science [20]. K-means clusteringis a type of unsupervised learning, which is used in unlabeled data(that is, data without defined categories or groups)[21]. The K-Means algorithm is an algorithm that groups data into several groups, where data in a group has characteristics that are different from data in other groups [22].

The sequence of the clustering process with the K-Means algorithm first determines the number of clusters to be formed [23]. Next determine the center point (Centroid) at random [24]. Third, repeat or repeat the calculation of the distance between one data and another with centroids, determine new cluster members, and update the center point or centroid with new cluster members. The last is to check whether the value of the center point or centroid changes. If the centroid changes then repeats the distance calculation again, but if the centroid does not change then the calculation is completed[25]. While the tool used is the Rapidminer Application [26]. RapidMiner is a data mining software, which can be used as a standalone application for data analysisor integrated as a data mining engine into other products [27]. The flowchart of the k-means algorithm can be seen below:

Figure of Flowchart K-Means [28]

2.1. Result and Discussion

Data Analysis

Data processing for analysis of employee discipline using digital attendance data and data from questionnaires distributed to 159 employees PT. SIS Pekanbaru. The data processed are data on lateness, non-attendance (TAP) and attendance hours. The data is obtained using the k-means algorithm method and supported by the Rapidminer application so that later results can be obtained from the data grouping. Below is a selection of the data used for this study:

Table of Sample Data on the Level of Discipline in PT. SIS

NO	NAMA KARYAWAN	KETERLAMBATAN KARYAWAN	TAP KARYAWAN	JAM KEHADIRAN			P1	P2	P3	P4
				07:45-08:00	08:01-08:15	08:15-08:30				
1	AFDA ILHAM	2	3	28	130	48	1	1	1	1
2	AFFRATI ADILLAH	1	2	0	3	153	1	1	1	1
3	AGUS ARDI	1	1	175	29	7	1	1	1	1
4	AGUS WASITO	4	3	6	103	91	1	1	1	1
5	AILEEN SARA ROS	1	2	162	47	9	1	1	1	1
6	AL FAJRI	1	1	81	43	7	1	2	1	1
7	AN FEBRIS ADHIWI	5	4	33	65	59	1	1	1	1
8	ALI YANIS	3	5	20	110	38	1	1	1	1
9	ANCES LOVA LUF	1	1	3	17	5	1	1	1	1
10	ANDRE ERYANTO	2	1	35	167	8	1	1	1	1
11	ANGGA REVANDI	4	5	2	157	47	1	1	1	1
12	ANGGA SAPUTRA	3	4	43	136	18	1	1	1	1
13	ANGGI	5	4	84	114	21	1	1	1	2
14	ARIF PUTRA ADEL	5	2	83	91	12	1	1	2	2
15	ASRINI	2	3	0	6	209	1	1	2	1
16	BETTRIYA ASTUTI	3	3	74	128	7	1	2	2	2
17	BRIAN VANT VIAT	3	1	83	108	19	1	1	1	1
18	D.PANJI PUPUNG N	5	1	163	52	3	1	1	1	1
19	DASRI	2	2	56	146	13	1	1	1	1
20	DENI ALFIAN	4	1	15	148	39	1	1	1	1
21	DESKA EWA SAPU	1	1	6	15	3	1	1	1	1
22	DESSY SAPUTRI	4	2	1	45	160	1	1	1	1
23	DHENY PRATAMA	3	3	79	117	6	1	1	1	1
24	DIARESI SUSANTI	2	3	1	6	204	1	1	2	1
25	DWI ENDANG MAI	3	2	95	87	10	1	1	1	1
26	EFNI YUNITA	3	1	6	34	66	1	1	1	1
27	EKA WATI	3	2	0	40	92	1	2	2	2
28	EKO DOSI	1	1	207	11	0	1	1	1	1
29	EKO IRAWAN	4	1	10	158	38	1	1	1	1
30	EKO RUSTAMI AR	2	3	8	152	33	1	1	1	1

Data Transformation

The data is transformed because the data type is not numeric. Therefore, the data must first be transformed by sorting the frequency of most of the data from the highest to the lowest frequency and the initials of the data.

Table 1. Data Transformation Late

Terlambat	Frekuensi	Inisial
<39	49	1
>121-161	41	2
>81-120	25	3
>162-197	25	4
>40-80	19	5

In late data, transformations are carried out by frequency of the data and insial. It can be seen in the table that 49 employees who were late less than 39 times during the 9 months of application use were 49 employees who were given the initials 1. Initials are carried out from the highest frequency to the lowest.

Table 2. TAP Data Transformation

TAP (*Tidak absen Pulang*)	Frekuensi	Inisial
<34	86	1
>35-69	30	2
>70-104	24	3
>105-138	11	4
>138-165	8	5

In the data table, no absence home (TAP) was also transformed, it can be seen in the table that employees who are not absent home (TAP) are less than 34 times as many as 86 employees with an initial of 1. Initials are performed from highest to lowest frequency.

Table 3.Question Transformations (1)

P1	Frekuensi	Inisial
Ya	150	1
Tidak	9	2

In the table of question 1, a transformation is carried out with the highest frequency, namely 150 employees with the answer "Yes" and 9 employees with the answer "No". Question 1 is Does the Digital Attendance Application make it easier to fill in attendance?

Table 4. Question Transformation (2)

P2	Frekuensi	Inisial
Ya	115	1
Tidak	44	2

In question 2 table, a transformation was carried out with the highest frequency, namely 115 employees with the answer "Yes" and 44 employees with the answer "No". Question 2 is whether the existence of this Digital Attendance Application can reduce the level of fraud in filling attendance?

Table 5. Question Transformation (3)

P3	Frekuensi	Inisial
Ya	114	1
Tidak	45	2

In question 3 table, a transformation was carried out with the highest frequency, namely 114 employees with the answer "Yes" and 45 employees with the answer "No". Question 3 is whether the Digital Attendance Application is easy to use?

Table 6. Question Transformation (4)

P4	Frekuensi	Inisial
Ya	116	1
Tidak	43	2

In question table 4, a conversion was made with the highest frequency, namely 116 employees answering "yes" and 43 employees answering "no". Question 4 is: Does the digital presence application have a high level of security?

After the conversion of each attribute is completed, the next step is to transfer the initials of the data into the output data, then the data is ready for processing. Manual processing using Microsoft Excel involves 159 data from all PT. SIS. The data to be processed can be seen in the following table:

Table 7. Data on the Level of Employee Discipline at PT. SIS Pekanbaru

DATA KE-i	KETERLAMBATAN KARYAWAN	TAP KARYAWAN	JAM KEHADIRAN			P1	P2	P3	P4
			07:30-08:00	08:01-08:30	>08:30				
1	2	3	28	130	48	1	1	1	1
2	1	2	0	3	153	1	1	1	1
3	1	1	175	29	7	1	1	1	1
4	4	3	6	103	91	1	1	1	1
5	1	2	162	47	9	1	1	1	1
6	1	1	81	43	7	1	2	1	1
7	5	4	33	65	59	1	1	1	1
8	3	5	20	110	38	1	1	1	1
9	1	1	3	17	5	1	1	1	1
10	2	1	35	167	8	1	1	1	1
11	4	5	2	157	47	1	1	1	1
12	3	4	43	136	18	1	1	1	1
13	5	4	84	114	21	1	1	1	2
14	5	2	83	91	12	1	1	2	2
15	2	3	0	6	209	1	1	2	1
16	3	3	74	128	7	1	2	2	2
17	3	1	83	108	19	1	1	1	1
18	5	1	163	52	3	1	1	1	1
19	2	2	56	146	13	1	1	1	1
20	4	1	15	148	39	1	1	1	1
21	1	1	6	15	3	1	1	1	1
22	4	2	1	45	160	1	1	1	1
23	3	3	79	117	6	1	1	1	1
24	2	3	1	6	204	1	1	2	1
25	3	2	95	87	10	1	1	1	1
26	3	1	6	34	66	1	1	1	1
27	3	2	0	40	92	1	2	2	2
28	1	1	207	11	0	1	1	1	1
29	4	1	10	158	38	1	1	1	1
30	2	3	8	152	33	1	1	1	1

Clustering Process Using K-Means Algorithm

The data obtained will be processed through a clustering procedure using the K-Means algorithm so that the best grouping results are obtained, which will later help in determining employees with disciplined and less disciplined categories.

1. *Determining the number of clusters*

   Determination of the number of clusters is carried out to find out the results of grouping employee data. So in this study the number of clusters used is as many as 2*clusters (k=2),* So that later it will be known which clusters belong to the category of discipline and the category of lack of discipline.

2. *Determining the centroid*

   The initial center of the cluster or centroid is determined randomly or randomly, where the cluster value 0 is taken from the 16th row and the cluster value 1 on the 15th row. Here are the initial centroid values on employee data groupings:

   Cluster 0 : (3 ; 3 ; 74 ; 128 ; 7 ; 1 ; 2 ; 2 ; 2)

   Cluster 1 : (2 ; 3 ; 0 ; 6 ; 209 ; 1 ; 1 ; 2 ; 1)

3. *Calculating the distance from the centroid*

   After determining the initial center of the centroid, the calculation of the distance between the centroid point and the point of each object is carried out using Euclidian Distance.The following is the calculation of centroid 0 with data of the 16th row and centroid 1 in the data of the 15th row.

Calculation of the distance from the 1st data to the center of the cluster:

Centroid0= $\sqrt{(3-3)^2+(3-3)^2+(74-74)^2+(128-128)^2(7-7)^2+(1-1)^2}+(2-2)^2+(2-2)^2+(2-2)^2=0{,}0$

Centroid 1 = $\sqrt{(2-3)^2+(3-3)^2+(0-74)^2+(6-128)^2(209-7)^2+(1-1)^2}+(2-1)^2+(2-2)^2+(2-1)^2$= 247,3

Calculation of the distance from the 2nd data to the center of the cluster:

Centroid 0 = $\sqrt{(3-2)^2+(3-3)^2+(74-0)^2+(128-6)^2(7-209)^2+(1-1)^2}+(2-1)^2+(2-2)^2+(2-1)^2$= 247,3

Centroid 1 = $\sqrt{(2-2)^2+(3-3)^2+(0-0)^2+(6-6)^2(209-209)^2+(1-1)^2}+(1-1)^2+(2-2)^2+(1-1)^2$= 0,0

For the overall result of the calculation of the distance between centroid points, it can be seen in table.8 below.

4. Assign each object to the nearest centre of gravity to assign the objects by grouping them based on the minimum distance of the object to the centre of the cluster by specifying the code "1" if the result of the cluster is close to zero. The calculation is performed continuously until the 159th data is compared with the cluster centre. After the calculation is done, you will get data as in the following table:

Table 8.Results of Distance Calculation and Grouping of 0th Iteration Data

Iterasi-0				50,4	219,4	1	0
Cluster 0	*Cluster 1*	*Cluster 0*	*ClustCr 1*	74,6	220,9	1	0
61,7	205,1	1	0	64,9	228,9	1	0
206,0	56,1	0	1	65,5	217,6	1	0
141,5	268,3	1	0	192,8	302,1	1	0
110,9	152,9	1	0	196,9	65,6	0	1
119,7	260,6	1	0	76,6	257,6	1	0
85,3	220,8	1	0	248,5	4,7	0	1
91,4	164,6	1	0	78,5	214,3	1	0
64,9	201,2	1	0	89,2	177,7	1	0
131,8	204,3	1	0	95,8	259,3	1	0
55,2	259,9	1	0	64,6	263,6	1	0
87,4	221,5	1	0	99,5	212,7	1	0
33,9	235,0	1	0	85,1	216,2	1	0
22,3	232,5	1	0	137,0	194,1	1	0
38,5	230,1	1	0	67,3	183,6	1	0
247,3	0,0	0	1	29,2	224,4	1	0
0,0	247,3	1	0	80,4	174,2	1	0
25,1	231,1	1	0	8,5	245,2	1	0
117,1	266,7	1	0	132,0	272,1	1	0
26,2	247,3	1	0	109,2	177,6	1	0
70,1	222,0	1	0	127,3	226,2	1	0
132,0	206,3	1	0	88,2	242,6	1	0
188,8	62,7	0	1	139,6	274,6	1	0
12,2	244,5	1	0	65,5	244,3	1	0
243,0	5,1	0	1	103,6	238,3	1	0
46,2	234,9	1	0	62,8	231,1	1	0
130,2	145,9	1	0	120,0	211,3	1	0
143,0	121,9	0	1	126,8	224,6	1	0
177,3	294,2	1	0	75,1	252,4	1	0
77,2	229,0	1	0	119,0	193,8	1	0
74,9	228,8	1	0	32,3	229,2	1	0
246,8	7,7	0	1	179,3	73,3	0	1
90,1	279,2	1	0	119,6	263,0	1	0
29,6	245,6	1	0	127,5	202,8	1	0
107,4	147,6	1	0	82,2	188,1	1	0
37,0	223,8	1	0	238,6	14,0	0	1
51,0	256,2	1	0	86,8	217,5	1	0
150,1	277,7	1	0	154,0	281,1	1	0
66,5	255,8	1	0	51,9	260,6	1	0
88,0	257,2	1	0	252,8	6,6	0	1

116.3	261.5	1	0
141.5	208.2	1	0
18.3	243.8	1	0
49.3	245.4	1	0
10.2	241.9	1	0
89.5	254.7	1	0
61.7	246.1	1	0
131.7	210.2	1	0
86.7	210.3	1	0
51.2	248.2	1	0
128.2	220.5	1	0
202.5	58.2	0	1
116.7	260.5	1	0
204.0	58.8	0	1
225.9	30.6	0	1
80.8	214.1	1	0
63.1	254.1	1	0
6.5	252.0	1	0
127.2	196.7	1	0
94.4	263.9	1	0
195.6	72.2	0	1
242.0	6.9	0	1
17.8	249.6	1	0
141.2	208.1	1	0
97.6	172.1	1	0
102.0	254.2	1	0
74.9	246.2	1	0
11.6	236.2	1	0
122.4	217.3	1	0
141.5	208.6	1	0
76.0	241.1	1	0
96.6	238.8	1	0
177.6	286.5	1	0
134.8	274.6	1	0
83.7	245.1	1	0
79.5	224.0	1	0
125.3	212.2	1	0
64.5	254.6	1	0
84.7	224.3	1	0
30.8	253.7	1	0
111.1	261.1	1	0

169.1	83.7	0	1
26.1	245.1	1	0
65.5	208.4	1	0
64.8	213.0	1	0
141.6	122.3	0	1
29.1	258.6	1	0
58.5	233.8	1	0
251.8	5.3	0	1
87.5	222.9	1	0
258.8	12.2	0	1
254.1	8.0	0	1
138.9	267.2	1	0
72.9	255.1	1	0
109.1	207.7	1	0
24.8	234.6	1	0
166.6	87.3	0	1
57.0	240.4	1	0
41.7	238.4	1	0
71.7	217.8	1	0
71.9	237.9	1	0
165.8	277.4	1	0
83.9	178.4	1	0
247.7	5.5	0	1
75.4	247.6	1	0
57.9	247.5	1	0
50.9	259.3	1	0
120.1	183.7	1	0
60.2	259.7	1	0
123.1	196.2	1	0
95.0	271.7	1	0
161.4	282.6	1	0
147.1	207.1	1	0
97.0	262.4	1	0
149.7	134.5	0	1
42.0	224.8	1	0
141.7	117.2	0	1
128.4	123.7	0	1
107.6	233.2	1	0

In Table 8. above, you can see the result of calculating the distance between the centres of gravity of clusters 1 to 159. The data on cluster0 and cluster 1 are compared, and the data with the smallest distance is given the number "1". After performing the 0th iteration, perform the 1st iteration to compare the position of the results of cluster 0 and cluster 1. If the position does not change, the next iteration does not need to be performed again. To perform the 1st iteration, first define a new cluster. The results of the new cluster are determined as follows:

Cluster 0 :
1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21, 23, 25, 26, 28, 29, 30, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 46, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 73, 74, 75, 77, 78, 79, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 93, 96, 97, 98, 99, 100, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 123, 124, 125, 127, 128, 130, 133, 134, 135, 136, 138, 139, 140, 141, 142, 143, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 156, 159.

Cluster 1 :
2, 15, 22, 24, 27, 31, 45, 47, 72, 76, 80, 92, 94, 95, 101, 102, 122, 126, 129, 131, 132, 137, 144, 155, 157, 158.

5. Iterate, then determine the position of the new centroid by calculating the average of the data on the same centroid.

*Centroid*1

$$= \left(\frac{2+1+4+3+1+4+1+2+4+5+2+3+1+4+2+5+3+1+1+1+5+1+4+5+1+4+3+1+4+1+4+1}{133},\right.$$

$$\frac{2+1+4+3+1+4+1+2+4+5+2+3+1+4+2+5+3+1+1+1+5+1+4+5+1+4+3+1+4+1+4+1}{133},$$

$$\frac{2+1+5+4+1+2+5+2+1+2+1+1+1+2+4+2+2+1+1+1+2+2+1+2+2+3+3+3+1+1+1}{133},$$

$$\frac{2+1+2+43+6+1+1+15+12+109+31+96+91+6+48+35+162+8+73+4+0+0+13+136+154+1+0+1+16+196+3+192}{133},$$

$$\frac{2+1+157+136+15+45+0+196+150+75+176+32+2+176+134+23+52+3+0+4+2+12+167+55+51+45+1+0+163+16+191+18}{133},$$

$$\frac{2+1+47+18+3+160+205+5+42+20+10+21+2+36+2+36+25+147+9+1+151+137+206+30+19+3+135+220+214+31+13+10+2}{133},$$

$$\frac{2+1+1+1+1+1+1+1+1+1+1+1+1+1+1+2+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1}{133},$$

$$\frac{2+1+1+1+1+1+1+2+1+1+1+2+1+1+1+2+2+2+1+1+1+1+1+1+1+1+1+1+1+2+1+1}{133},$$

$$\frac{2+1+1+1+1+1+1+1+1+1+1+1+2+1+1+1+2+2+2+1+1+1+1+1+1+1+1+1+1+2+1+1+1}{133},$$

$$\frac{2+1+1+1+1+1+1+1+1+1+1+1+2+1+1+1+2+2+2+1+1+1+1+1+1+1+1+1+1+2+1+1+1}{133},$$

)

= (0.632 ; 0.474 ; 10.962 ; 15.789 ; 14.481 ; 0.256 ; 0.293 ; 0.286 ; 0.286)

Centroid 2 =

$$\left(\frac{2+1+4+1+1+5+3+1+2+4+3+5+2+3+3+5+2+4+1+4+2+3+3+3+1+4+2+1+2+4+2+2+1+2+1+3+4+5+1+1+2+2+1+4+1+2+3+1+2+3+2+3+1+4+1+4+2+1+5+1+1+5+1+2+5}{26},\right.$$

$$\frac{3+2+3+2+1+4+5+1+1+5+4+2+3+3+1+1+2+1+1+2+3+2+1+2+1+1+3+5+2+5+2+2+1+4+1+1+1+2+1+2+2+5+1+2+1+1+1+1+2+5+4+1+1+1+1+2+3+1+1+1+1+2+1+4+2}{26},$$

$$\frac{28+0+6+162+81+33+20+3+35+2+43+83+0+74+83+163+56+15+6+1+1+95+6+0+207+10+8+1+15+17+43+37+183+39+12+109+218+8+16+0+58+12+145+31+96+0+39+53+26+69+174+0+91+6+21+115+111+29+87+129+21+48+35}{26},$$

$$\frac{130+3+103+47+43+65+110+17+167+157+136+91+6+128+108+52+146+148+15+45+6+87+34+40+11+158+152+0+196+93+129+163+25+170+68+92+150+75+0+3+177+2+51+116+64+176+32+13+106+126+102+133+42+135+2+176+163+33+79+17+2+77+22+134+23}{26},$$

$$\frac{48+153+91+9+7+59+38+5+8+47+18+12+209+7+19+3+13+39+3+160+204+10+66+92+0+38+33+205+5+91+27+10+1+14+1+11+42+20+0+144+17+208+8+70+2+10+21+15+60+27+66+11+3+87+2+36+23+2+20+0+2+4+17+25+147}{26},$$

$$\frac{1+1+1+

Table 10.Results of Distance Calculation and Grouping of 1st Iteration Data

Iterasi-1			
Cluster 0	*Cluster 1*	*Cluster 0*	*Cluster 1*
120,28	144,43	1	0
247,58	139,56	0	1
164,75	209,29	1	0
116,21	165,58	1	0
154,34	190,83	1	0
75,54	200,25	1	0
70,16	181,01	1	0
97,67	165,52	1	0
12,54	253,00	1	0
147,16	153,26	1	0
145,30	154,06	1	0
124,54	146,99	1	0
122,71	137,43	1	0
104,31	159,54	1	0
260,93	195,11	0	1
129,00	142,34	1	0
117,13	143,25	1	0
156,78	190,07	1	0
137,81	138,05	1	0
134,58	150,11	1	0
12,63	253,81	1	0
215,24	148,81	0	1
122,31	147,50	1	0
258,54	190,07	0	1
110,29	160,41	1	0
54,94	219,18	1	0
214,42	82,05	0	1
196,64	235,94	1	0
144,19	151,04	1	0
137,53	157,07	1	0
263,60	191,49	0	1
158,61	180,53	1	0
137,80	137,98	1	0
109,06	163,97	1	0
118,35	145,24	1	0
145,54	149,60	1	0
172,84	217,30	1	0
150,81	154,73	1	0
138,04	176,18	1	0
81,33	178,45	1	0
137,06	150,36	1	0
114,77	161,12	1	0
132,82	143,99	1	0
208,15	248,65	1	0
241,89	130,22	0	1
152,71	161,34	1	0
263,64	194,38	0	1
59,93	200,37	1	0
114,67	156,86	1	0
143,00	179,08	1	0
146,38	161,53	1	0
86,90	199,42	1	0
52,26	209,44	1	0
11,75	253,65	1	0
104,89	146,92	1	0
118,74	140,84	1	0
101,63	156,44	1	0
130,88	139,06	1	0
165,54	200,27	1	0
140,05	153,44	1	0
82,19	235,15	1	0
150,58	161,77	1	0
168,61	207,40	1	0
147,84	148,71	1	0
106,20	205,17	1	0
118,55	157,59	1	0
23,37	242,79	1	0
78,36	235,88	1	0
133,47	167,23	1	0
12,57	235,34	1	0
124,39	140,66	1	0
211,87	135,00	0	1
152,24	194,90	1	0
9,09	247,70	1	0
81,09	182,45	1	0
258,84	182,40	0	1
59,22	208,03	1	0

180,76	215,31	1	0	166,35	204,48	1	0
142,20	155,26	1	0	131,92	169,54	1	0
265,81	199,35	0	1	140,84	153,44	1	0
155,47	186,59	1	0	35,10	245,09	1	0
19,07	262,82	1	0	145,73	156,11	1	0
131,73	139,57	1	0	148,54	148,93	1	0
142,36	143,87	1	0	138,99	143,27	1	0
129,52	137,94	1	0	147,77	189,76	1	0
159,12	164,64	1	0	210,76	124,49	0	1
129,05	157,75	1	0	128,64	142,07	1	0
31,35	251,02	1	0	95,49	173,02	1	0
138,27	152,86	1	0	128,51	144,53	1	0
140,69	150,18	1	0	213,63	80,08	0	1
65,44	241,24	1	0	138,52	143,49	1	0
244,28	137,21	0	1	140,04	143,20	1	0
149,44	193,15	1	0	265,19	198,31	0	1
246,05	137,50	0	1	64,93	208,94	1	0
254,48	165,56	0	1	269,11	206,38	0	1
136,72	151,44	1	0	267,01	200,41	0	1
146,92	154,07	1	0	167,21	202,64	1	0
131,81	144,03	1	0	145,67	161,76	1	0
9,95	245,48	1	0	78,75	210,04	1	0
161,55	172,52	1	0	124,64	141,95	1	0
246,05	123,87	0	1	210,96	120,69	0	1
255,23	191,88	0	1	143,17	143,77	1	0
129,78	143,61	1	0	137,07	138,32	1	0
18,53	262,73	1	0	131,00	152,36	1	0
105,86	170,72	1	0	148,13	148,29	1	0
144,79	178,73	1	0	185,05	222,65	1	0
148,15	155,86	1	0	114,96	150,82	1	0
123,79	139,34	1	0	263,24	193,39	0	1
57,30	237,56	1	0	146,92	157,87	1	0
21,21	262,36	1	0	142,23	149,75	1	0
149,98	152,06	1	0	146,65	150,89	1	0
131,21	176,63	1	0	17,15	231,91	1	0
188,97	239,75	1	0	147,72	155,61	1	0
				14,07	241,30	1	0
				165,10	175,49	1	0
				181,49	224,88	1	0
				22,46	267,87	1	0
				159,91	174,70	1	0
				234,65	61,60	0	1
				128,73	136,64	1	0
				183,11	123,65	0	1
				188,28	92,65	0	1
				101,12	209,15	1	0

On the table. 10 is the result of the 1st iteration which when compared to the 0th iteration of each clutser does not change position so that the loop is stopped. Then the calculation results are obtained only until the 1st iteration.

The results of the new cluster of the 1st iteration are contained in the following sequence of data:
Cluster 0 :
1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21, 23, 25, 26, 28, 29, 30, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 46, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 73, 74, 75, 77, 78, 79, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 93, 96, 97, 98, 99, 100, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 123, 124, 125, 127, 128, 130, 133, 134, 135, 136, 138, 139, 140, 141, 142, 143, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 156, 159.

Cluster 1 :
2, 15, 22, 24, 27, 31, 45, 47, 72, 76, 80, 92, 94, 95, 101, 102, 122, 126, 129, 131, 132, 137, 144, 155, 157, 158
In the 0th iteration and the 1st iteration the cluster result does not change then the iteration is not done again. It can be concluded that cluster 0 consists of 133 data and cluster 1 consists of 26 data seen in table 11 below:

Table 11 : Cluster Final Results

Cluster	Data KC-i	KCtCrlambatan Karyawan	TAP Karyawan	Jam Kchadiran			P1	P2	P3	P4
				07:45-08:00	08:01-08:15	08:15-08:30				
Cluster 0	1	2	3	28	130	48	1	1	1	1
	3	1	1	175	29	7	1	1	1	1
	4	4	3	6	103	91	1	1	1	1
	5	1	2	162	47	9	1	1	1	1
	6	1	1	81	43	7	1	2	1	1
	7	5	4	33	65	59	1	1	1	1
	8	3	5	20	110	38	1	1	1	1
	9	1	1	3	17	5	1	1	1	1
	10	2	1	35	167	8	1	1	1	1
	11	4	5	2	157	47	1	1	1	1
	12	3	4	43	136	18	1	1	1	1
	13	5	4	84	114	21	1	1	1	2
	14	5	2	83	91	12	1	1	2	2
	16	3	3	74	128	7	1	2	2	2
	17	3	1	83	108	19	1	1	1	1
	18	5	1	163	52	3	1	1	1	1
	19	2	2	56	146	13	1	1	1	1
	20	4	1	15	148	39	1	1	1	1
	21	1	1	6	15	3	1	1	1	1
	23	3	3	79	117	6	1	1	1	1
	25	3	2	95	87	10	1	1	1	1
	26	3	1	6	34	66	1	1	1	1
	28	1	1	207	11	0	1	1	1	1
	29	4	1	10	158	38	1	1	1	1
	30	2	3	8	152	33	1	1	1	1
	32	2	2	15	196	5	1	2	1	1
	33	2	1	53	147	15	1	1	1	1
	34	4	5	17	93	91	1	2	2	2
	35	2	2	43	129	27	1	1	1	1
	36	2	2	37	163	10	1	2	1	2
	37	1	1	183	25	1	1	2	2	2
	38	2	4	23	170	14	1	1	1	1
	39	1	1	138	68	1	1	2	2	2
	40	3	1	39	92	11	1	1	1	1
	41	4	1	12	150	42	1	1	1	1
	42	5	2	109	75	20	1	1	1	1
	43	2	3	22	145	43	1	1	1	1
	44	1	1	218	0	0	1	1	1	1
	46	2	2	16	177	17	1	1	2	1

	48	1	1	59	51	8	1	1	1	1
	49	4	2	12	116	70	1	2	2	2
	50	1	1	145	64	2	1	1	1	1
	51	2	1	31	176	10	1	1	1	1
	52	3	1	96	32	21	1	2	2	2
	53	1	1	52	46	3	1	1	1	1
	54	1	1	0	13	15	1	2	2	2
	55	2	2	39	106	60	1	1	1	1
	56	3	5	53	126	27	1	1	1	1
	57	2	4	26	102	66	1	1	1	1
	58	3	1	69	133	11	1	1	1	2
	59	1	1	174	42	3	1	1	1	1
	60	4	1	0	135	87	2	2	2	2
	61	1	1	91	2	2	1	1	1	1
	62	4	2	6	176	36	1	1	1	1
	63	1	1	178	35	2	1	1	1	1
	64	2	3	21	163	23	1	2	1	1
	65	1	1	115	33	2	1	1	1	1
	66	5	1	111	79	20	1	1	1	1
	67	1	1	29	17	0	1	2	2	2
	68	1	1	87	2	2	2	2	2	2
	69	5	2	129	77	4	1	1	1	1
	70	1	1	21	22	17	2	2	2	2
	71	2	4	48	134	25	1	1	1	1
	73	1	1	160	45	4	1	1	1	1
	74	1	1	15	15	7	1	2	2	2
	75	2	1	9	92	42	1	2	1	1
	77	1	1	63	42	4	1	1	1	1
	78	1	1	191	28	4	1	1	1	1
	79	2	2	41	168	9	1	1	1	1
	81	3	2	162	52	9	1	2	2	2
	82	1	1	8	3	1	1	2	2	2
	83	3	2	60	138	13	1	1	1	1
	84	2	3	35	156	18	1	2	2	2
	85	2	2	68	132	14	2	2	2	2
	86	2	1	3	180	23	1	1	1	1
	87	5	1	119	86	9	1	1	1	1
	88	1	1	36	2	2	1	2	2	2
	89	2	5	4	148	54	2	1	2	1
	90	2	3	31	155	13	1	1	1	1
	91	1	1	73	0	1	1	1	1	1
	93	1	1	157	46	5	1	1	1	1
	96	4	1	8	148	49	1	1	1	1

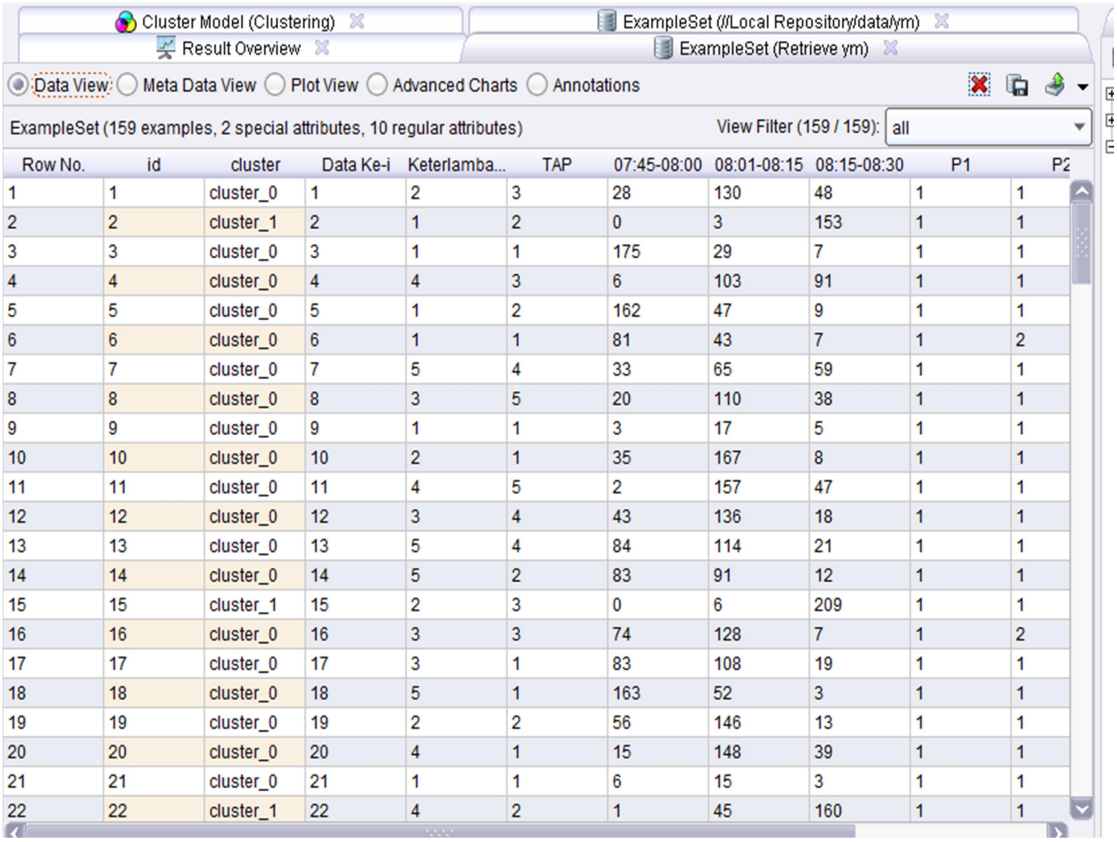
	97	2	1	27	169	16	1	1	1	1
	98	3	1	72	132	3	1	1	1	2
	99	1	1	4	22	13	1	2	2	2
	100	4	1	2	188	18	1	1	1	1
	103	3	1	89	119	5	1	2	1	1
	104	1	1	5	5	1	1	2	2	2
	105	2	4	2	106	69	1	1	1	1
	106	5	1	149	59	10	1	1	1	1
	107	4	3	16	171	27	1	1	1	1
	108	3	3	69	125	17	1	2	2	2
	109	1	1	66	6	2	1	1	1	1
	110	1	1	14	0	1	1	2	2	2
	111	4	1	13	167	30	1	1	1	1
	112	5	2	136	55	19	1	1	1	1
	113	5	1	199	2	3	1	2	2	2
	114	1	3	175	39	0	1	2	2	2
	115	5	1	132	68	12	1	1	1	1
	116	4	2	7	154	41	1	1	1	1
	117	1	1	42	7	1	1	2	2	2
	118	2	4	27	171	17	1	1	1	1
	119	4	3	7	161	47	1	1	1	1
	120	2	1	55	152	9	1	1	1	1
	121	1	2	154	51	3	1	1	1	1
	123	3	1	95	113	10	1	2	2	2
	124	3	1	15	110	29	1	2	2	2
	125	2	3	23	140	45	1	1	1	1
	127	3	1	101	120	0	1	1	2	1
	128	2	3	27	154	30	1	1	1	1
	130	1	1	69	41	0	1	1	1	1
	133	1	1	177	35	11	1	2	2	2
	134	4	1	22	177	21	1	1	1	1
	135	3	4	89	21	22	1	2	2	2
	136	3	2	53	133	19	1	1	1	1
	138	2	2	29	158	25	2	2	1	1
	139	2	1	43	150	24	1	1	1	1
	140	2	4	13	144	41	1	1	1	1
	141	4	1	16	163	31	1	2	2	2
	142	1	1	196	16	13	1	1	1	1
	143	2	1	20	116	70	1	2	1	2
	145	4	1	17	173	27	1	1	2	1
	146	4	1	31	164	21	1	1	1	1
	147	2	2	38	164	7	1	2	2	2
	148	5	1	14	26	27	1	1	2	1

	149	2	1	31	170	10	1	2	2	2
	150	5	2	4	27	14	1	1	1	1
	151	4	1	3	191	10	1	1	1	1
	152	1	1	192	18	2	1	1	1	1
	153	1	1	0	1	2	1	2	2	2
	154	4	1	1	190	22	1	2	1	1
	156	2	3	43	139	33	2	2	2	2
	159	1	1	111	27	5	1	1	1	1
Cluster 1	2	1	2	0	3	153	1	1	1	1
	15	2	3	0	6	209	1	1	2	1
	22	4	2	1	45	160	1	1	1	1
	24	2	3	1	6	204	1	1	2	1
	27	3	2	0	40	92	1	2	2	2
	31	1	5	1	0	205	1	1	1	1
	45	1	2	8	3	144	1	1	2	2
	47	2	5	0	2	208	1	1	1	1
	72	5	2	35	23	147	2	2	2	2
	76	5	3	1	2	196	1	1	2	1
	80	3	3	1	1	213	1	1	1	1
	92	1	1	4	4	151	1	1	1	1
	94	1	1	7	0	151	1	1	1	1
	95	5	2	1	1	179	2	1	1	1
	101	5	2	0	2	137	1	1	1	1
	102	1	2	0	12	206	1	1	1	1
	122	4	3	1	45	135	1	1	1	1
	126	2	4	4	38	91	1	1	1	1
	129	1	2	0	2	212	1	1	1	1
	131	3	3	0	1	220	1	1	1	1
	132	1	3	1	0	214	1	1	1	1
	137	4	1	0	45	131	1	1	1	2
	144	5	3	0	2	207	1	1	1	1
	155	3	2	0	17	75	1	1	1	1
	157	4	1	1	81	119	1	1	1	1
	158	2	5	38	45	98	1	1	1	1

Table 11 is the result of manual clustering with Microsoft Excel. The table consists of 2 clusters, namely cluster 0 and cluster 1. Cluster 0 has a total number of 133 employees, while in cluster 1 the total number of employees is 26. Judging by the number of each data cluster, cluster 0 is the cluster with the highest amount of data and cluster 1 is the cluster with the least amount of data. To find out the results of the analysis with the categories of discipline and lack of discipline in PT. SIS Pekanbaru then carried out the data processing using Rapidminer.

Clustering Test Results Using RapidMiner

The results of the analysis of employee discipline levels using digital attendance applications with the K-Means Clustering method can be seen in the following figure:



Cluster Model (Clustering) | ExampleSet (//Local Repository/data/ym)
Result Overview | ExampleSet (Retrieve ym)
Data View | Meta Data View | Plot View | Advanced Charts | Annotations
ExampleSet (159 examples, 2 special attributes, 10 regular attributes) View Filter (159 / 159): all

Row No.	id	cluster	Data Ke-i	Keterlamba...	TAP	07:45-08:00	08:01-08:15	08:15-08:30	P1	P2
1	1	cluster_0	1	2	3	28	130	48	1	1
2	2	cluster_1	2	1	2	0	3	153	1	1
3	3	cluster_0	3	1	1	175	29	7	1	1
4	4	cluster_0	4	4	3	6	103	91	1	1
5	5	cluster_0	5	1	2	162	47	9	1	1
6	6	cluster_0	6	1	1	81	43	7	1	2
7	7	cluster_0	7	5	4	33	65	59	1	1
8	8	cluster_0	8	3	5	20	110	38	1	1
9	9	cluster_0	9	1	1	3	17	5	1	1
10	10	cluster_0	10	2	1	35	167	8	1	1
11	11	cluster_0	11	4	5	2	157	47	1	1
12	12	cluster_0	12	3	4	43	136	18	1	1
13	13	cluster_0	13	5	4	84	114	21	1	1
14	14	cluster_0	14	5	2	83	91	12	1	1
15	15	cluster_1	15	2	3	0	6	209	1	1
16	16	cluster_0	16	3	3	74	128	7	1	2
17	17	cluster_0	17	3	1	83	108	19	1	1
18	18	cluster_0	18	5	1	163	52	3	1	1
19	19	cluster_0	19	2	2	56	146	13	1	1
20	20	cluster_0	20	4	1	15	148	39	1	1
21	21	cluster_0	21	1	1	6	15	3	1	1
22	22	cluster_1	22	4	2	1	45	160	1	1

Figure 3. Cluster Results View (ExampleSet Retrieve in Data View)

Figure 3 is a display of disciplinary data in the data view of 159 employee data processed with Rapidminer. The results of the data view processing show the grouping of the employee data according to the individual clusters, namely cluster 0 and cluster 1.

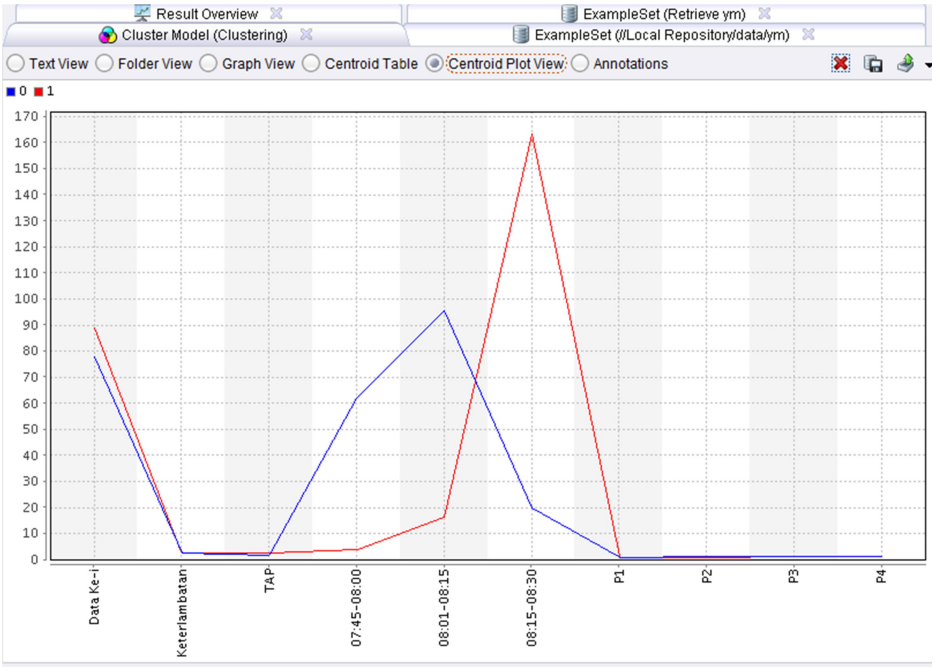


Figure 4. Centroid Plot View

Figure 4 shows the results of cluster 0, which is blue, and cluster 1, which is red. It can be seen that in cluster 0, the hour with the highest attendance from 08.01-08.15 is blue, while in cluster 1, the hour with the highest attendance from 08.15-08.30 is red. It can be concluded that cluster 0 is a disciplined category and cluster 1 is a less disciplined category.

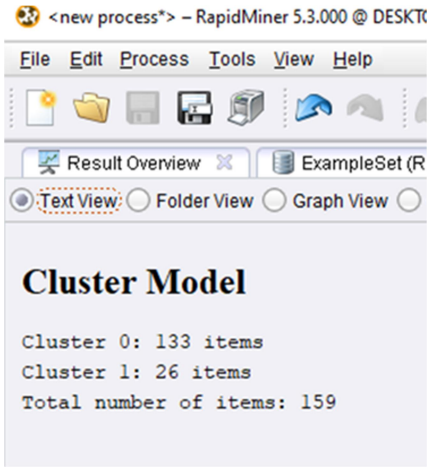
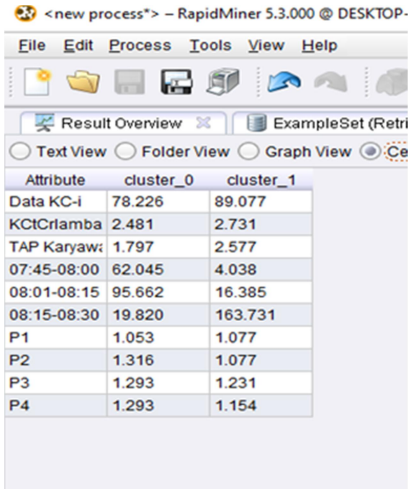


Figure 5. Cluster ModelView (Text View)

Figure 5 is a grouping in cluster 0 with a total of 133 employees and 26 employees in cluster 1 with a total of 159 employee data in PT. SIS Pekanbaru. The information obtained from the cluster analysis results can be seen in the following focus table.



Attribute	cluster_0	cluster_1
Data KC-i	78.226	89.077
KCtCrlamba	2.481	2.731
TAP Karyaw	1.797	2.577
07:45-08:00	62.045	4.038
08:01-08:15	95.662	16.385
08:15-08:30	19.820	163.731
P1	1.053	1.077
P2	1.316	1.077
P3	1.293	1.231
P4	1.293	1.154

Figure 6. Centroid Table View

Figure 6 shows the grouping of each cluster with the results of the data processing for each attribute. The number of clusters can be determined by the characteristics in the data. Data with the same characteristics are grouped in the same cluster and data with other characteristics are grouped in other clusters[23]. A cluster is a collection of objects that are "the same" between them and "different" in objects from other clusters [29]. In the digital attendance application data, tardiness, not absent from home and attendance hours are included in the same characteristics, while the four selected questions of the questionnaire contain different groups of characteristics grouped in other clusters, so that one gets two characteristics in the employee discipline data. Of these two characteristics, it is used as a reference in determining the number of clusters, so that there are 2 clusters used in the data processing with the categories of discipline and lack of discipline.

The results of the data processing revealed that cluster 0 consists of 133 employees or 83.64% with the discipline category and cluster 1 produces 26 employees or 16.35% with the less disciplined category. In the centroid table for cluster 0 and cluster 1, there is a significant difference in the attendance time of cluster 0 between 07:45 and 08:15 while in cluster 1 the attendance is highest between 08:15and 08:30. For the other

attributes, cluster 0 remains in the discipline category and cluster 1 in the less disciplined category. From the level of satisfaction with the use of digital attendance application based on 4 selected questions, it can be concluded that the use of this digital attendance application increases the discipline of the employees. The results of this analysis can be used by PT. SIS as a reference for evaluating employee discipline, determining employee promotion and improving employee discipline in the future.

2.2. Conclusion

Analysis of employee discipline based on digital attendance using the k-means algorithm method at PT. SIS Two groups were formed in Pekanbaru, namely cluster 0 with 133 employees or 83.64% with discipline category and cluster 1 with 26 employees or 16.35% with less disciplined category. The application of digital attendance monitoring suggests that the employees of PT. SIS are more disciplined than the less disciplined. And with this application, the level of cheating in completing the attendance sheet can be reduced and the lateness data can be easily viewed so that it can be used as an evaluation in the next month. The results of this analysis can be used by PT. SIS as a reference for assessing employee discipline, determining employee development and improving employee discipline in the future.

References

[1] Y. Maryani, M. Entang, and M. Tukiran, "The Relationship between Work Motivation , Work Discipline and Employee Performance at the Regional Secretariat of Bogor City (disiplin kerja)," *Int. J. Soc. Manag. Stud.*, vol. 02 No. 02, no. 02, pp. 1–16, 2021, [Online]. Available: https://ijosmas.org/index.php/ijosmas/article/view/14.

[2] Z. Tupti and M. Arif, "The Influence of Discipline and Motivation on Employee Performance," *Int. J. Econ. Technol. Soc. Sci.*, vol. 1, no. 1, pp. 61–69, 2020.

[3] B. Huda, S. Shofia Hilabi, and M. Rahayuningsih, "Android Based Employee Absence and Leaving Application Information System," *Buana Inf. Technol. Comput. Sci. (BIT CS)*, vol. 2, no. 1, pp. 11–16, 2021, doi: 10.36805/bit-cs.v2i1.1243.

[4] S. Younis, "Know Your Stars Before They Fall Apart : A Social Network Analysis of Telecom Industry to Foster Employee Retention Using Data Mining Technique," pp. 16467–16487, 2021, doi: 10.1109/ACCESS.2021.3050327.

[5] J. A. Tambuwun, W. Manoppo, and D. Keles, "Pengaruh Kedisiplinan Kerja terhadap Kinerja Karyawan PT. Hasjrat Abadi Manado," *Productivity*, vol. 2, no. 3, pp. 224–227, 2021.

[6] N. D. P. Zamhir Basem, Zulher, Muhammad Yusril, "Analysis of Discipline , Organizational Commitment , Work Environment and Their Effect on Employee Performance," vol. 4, no. 2, pp. 11–22, 2022.

[7] R. S. T. Atmojo, "Analisis Data E-Absensi untuk Menganalisis Perbandingan Pola Disiplin Kerja menggunakan Algoritma Clustering K-Means," *Electrician*, vol. 13, no. 1, p. 19, 2019, doi: 10.23960/elc.v13n1.2088.

[8] A. Al-rasheed, "Identification of important features and data mining classification techniques in predicting employee absenteeism at work Identification of important features and data mining classification techniques in predicting employee absenteeism at work," no. October, pp. 4587–4596, 2021, doi: 10.11591/ijece.v11i5.pp4587-4596.

[9] A. Asiri and M. Abdullah, "Employees Absenteeism Factors Based on Data Analysis and Classifi cation," vol. 12, no. January, pp. 119–127, 2019, doi: 10.21786/bbrc/12.1/14.

[10] M. C. Kocakulah, A. G. Kelley, K. M. Mitchell, and M. P. Ruggieri, "Absenteeism Problems And Costs: Causes, Effects And Cures," *Int. Bus. Econ. Res. J.*, vol. 15, no. 3, pp. 89–96, 2016, doi: 10.19030/iber.v15i3.9673.

[11] K. A. A. Nazeer and M. P. Sebastian, "Improving the Accuracy and Efficiency of the k-means Clustering Algorithm," *Proc. World Congr. Eng.*, vol. I, no. July 2009, p. 6, 2009.

[12] M. O. Asanbe and W. F. William, "Teachers ' Performance Evaluation in Higher Educational Institution using Data Mining Technique," vol. 10, no. 7, pp. 10–15, 2016.

[13] C. Devi, O. Soleman, N. Pramaita, and M. Sudarma, "Classification Of Loyality Customer Using K-Means Clustering, Studi Case : PT. Sucofindo (Persero) Denpasar Branch," *Int. J. Eng. Emerg. Technol.*, vol. 5, no. 2, pp. 160–167, 2020.

[14] P. Edastama, A. Dudhat, and G. Maulani, "Use of Data Warehouse and Data Mining for Academic Data A Case Study at a National University," vol. 1, no. 2, pp. 206–215, 2021.

[15] P. Nova Agustina, "JURNAL RESTI Perbandingan Algoritma K - Means Dengan Algoritma Fuzzy C - Means," vol. 2, no. 3, pp. 621–626, 2018.

[16] N. R. Dewi Murni, Bachtiar Efendi, "IMPLEMENTATION OF EMPLOYEE DISCIPLINE CLUSTERING AT GOTTING SIDODADI VILLAGE OFFICE BANDAR PASIR MANDOGE USING K-MEANS KANTOR DESA GOTTING SIDODADI BANDAR PASIR MANDOGE MENGGUNAKAN ALGORITMA K-MEANS," vol. 3, no. 2, pp. 295–304, 2022.

[17] Q. I. Mawarni and E. S. Budi, "Implementasi Algoritma K-Means Clustering Dalam Penilaian Kedisiplinan Siswa," vol. 3, 2022, doi: 10.30865/json.v3i4.4242.

[18] S. Darma *et al.*, "Penerapan Metode K-Means Dalam Pengolompokan Jumlah Wisatawan Asing Di Indonesia," *Pros. Semin. Nas. Ris. Dan Inf. Sci.*, vol. 2, pp. 255–261, 2020, [Online]. Available: https://www.bps.go.id.

[19] N. L. P. P. Dewi, I. N. Purnama, and N. W. Utami, "Penerapan Data Mining Untuk Clustering Penilaian Kinerja

Dosen Menggunakan Algoritma K-Means (Studi Kasus: STMIK Primakara)," *J. Ilm. Teknol. Inf. Asia*, vol. 16, no. 2, p. 105, 2022, doi: 10.32815/jitika.v16i2.761.

[20] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[21] A. Sarker, S. M. Shamim, M. Shahiduz, Z. M. Rahman, M. Shahiduz Zama, and M. Rahman, "Employee's Performance Analysis and Prediction using K-Means Clustering & Decision Tree Algorithm," *Int. Res. J. Softw. Data Eng. Glob. J. Comput. Sci. Technol.*, vol. 18, no. 1, p. 7, 2018, [Online]. Available: https://computerresearch.org/index.php/computer/article/view/1660/1644.

[22] Y. Christian, U. I. Batam, and K. Riau, "Application of K-Means Algorithm for Clustering the Quality of Lecturer Learning at Batam International University," vol. 3, no. 36, pp. 191–199, 2020.

[23] A. T. Rahman, Wiranto, and A. Rini, "Coal Trade Data Clustering Using K-Means (Case Study Pt. Global Bangkit Utama)," *ITSMART J. Teknol. dan Inf.*, vol. 6, no. 1, pp. 24–31, 2017.

[24] P. M. A. Putra and I. G. A. G. A. Kadyanan, "Implementation of K-Means Clustering Algorithm in Determining Classification of the Spread of the COVID-19 Virus in Bali," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 10, no. 1, p. 11, 2021, doi: 10.24843/jlk.2021.v10.i01.p03.

[25] I. M. Titik Susilowati, Dedy Sugiarto2, "Uji Validasi Algoritme Self-Organizing Map (SOM) dan K-Means untuk Pengelompokan Pegawai," *Uji Validasi Algoritm. Self-Organizing Map dan K-Means untuk Pengelompokan Pegawai*, vol. 1, no. 10, 2020.

[26] N. K. Wardhani, W. Gata, U. M. Buana, and U. N. Mandiri, "HUMAN RESOURCES MANAGER ' S DECISION MAKING TOWARDS," vol. 3, no. 2, pp. 336–345, 2022.

[27] R. Oktarina and Junita, "Determine the clustering of cities in Indonesia for disaster management using K-Means by excel and RapidMiner," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 794, no. 1, 2021, doi: 10.1088/1755-1315/794/1/012094.

[28] R. Hidayat and H. Kusniyati, "Analisis Clustering Dalam Pengelompokan Penjualan Menggunakan Algoritma K-Means Pada Cafe 47°Coffee Clustering Analysis in Sales Grouping Using The K-Means Algorithm at Cafe 47°Coffee," vol. 7, no. 2, pp. 420–434, 2022, [Online]. Available: www.jurnal.unimed.ac.id.

[29] N. Wakhidah, "Clustering Menggunakan K-Means Algorithm," *J. Transform.*, vol. 8, no. 1, p. 33, 2010, doi: 10.26623/transformatika.v8i1.45.