Journal of Artificial Intelligence and Engineering Applications
Website: https://ioinformatic.org/

15th October 2021. Vol. 1. No. 1

K-MEDOIDS ALGORITHM ANALYSIS IN PERMANENT WORKER GROUPING OF INDONESIAN CONSTRUCTION COMPANIES

Sonia Ananta Tarigan[1]*, Harly Okprana[2*], Ilham Syahputra Saragih[3]

[1] *STIKOM Tunas Bangsa Pematangsiantar, North Sumatra, Indonesia*
[1,2,3] *Jln. Sudirman Blok A No. 1-3 Pematangsiantar, Sumatera Utara*
[1] *soniaananta0@gmail.com*, [2] *harly@amiktunasbangsa.ac.id*, [3] *ilhamsaragih@amiktunasbangsa.ac.id*

Abstract

The construction companies are both those who run the construction work, both construction administrators and construction consultants who need the manpower for their operations. There is no way to determine the existence of a policy of the workers who have a work agreement with the business owner for a period of time. The company's long-term rating of construction workers in Indonesia from 2010-2018 is based on the need to provide information and input to the local government center at the construction site in Indonesia. One of the grouping methods that can be used is k - Medoids. The advantage of this method is to overcome sensitive to outlier. This method in its horn is represented by objects close to the center and thus capable of sterilizing a more precise value. Analysis of the data grouping shows that two cluster data produced one in the low and 33 in high cluster with total cost of 2.7557.

Keywords: *Construction Company, Analysis, K Medoids, Cluster*

1. Introduction

Large data has a lot of hidden information in it. With the development of science and technology, emerging disciplines that study large amounts of data in databases called Data Mining [1]. One of the stages in Data Mining is modeling. At this stage, many use statistical methods to obtain hidden information such as data grouping, forecasting and so on [2]. One of the methods commonly used to group data is cluster analysis. The purpose of this method is to group a set of data into several similar or similar groups. In the cluster or grouping process, several methods are used, one of which is the Partition Method [3],[4],[5]. This method partitions the data into several groups first and then determines the center point of the cluster with the size of the central symptom obtained from each group.

A construction company is a company that runs a business in the field of construction work, both construction implementers and construction consultants [6]. In carrying out its activities, the company uses a form of human resource strategy, which is to meet the company's human resource needs [7],[8]. The company uses two types of workers, namely permanent workers (who come from within the company) and out sourcee (workers who come from outside the company/contract workers). In the construction industry, workers are the main driver of construction activities, they are a very important component in the company.

The grouping of Permanent Workers of Construction Companies in Indonesia by province is based on the need to provide information and input to the central government of the local government on construction companies by province in Indonesia. In grouping large amounts of data, various data mining methods can be used [9], [10]. One of the methods used in the partition method is K-Means [11]. K-Means is widely used for cluster analysis because it is easy to understand and implement. However, the K-Means method has a weakness that is sensitive to outliers. Outlier is an observation that deviates far from a set of other observations [12]. The impact of outliers for K-Means is the formation of bad clusters, because outliers affect the distribution of data. The consequence due to the influence of the data distribution is the change in the cluster identity represented by the cluster average [13],[14]. So we need a method that can overcome the outlier problem, namely K-Medoids. In this method, each cluster is represented by an object close to the center of the cluster called a medoid [15].

2. Results and Discussion

To determine the existence of a data analysis process, first the data used must be valid. the data used in this study is valid because it is obtained from the official website of the National Statistics Agency, so that the data can be directly processed and tested for validity for a study.

2.1. Data Processing

In this study using data sourced from the Central Statistics Agency which can be seen in Table 1 below:

Table 1. Data

Province	Number of Permanent Workers in Construction Companies (Persons)								
	2010	2011	2012	2013	2014	2015	2016	2017	2018
Aceh	7696	8058	8450	8786	8983	9143	9252	9524	9614
North Sumatra	31321	32494	33542	34813	35827	36464	37167	38168	39051
West Sumatra	8028	8247	8416	8532	8827	8948	9080	9282	9362
Riau	18996	19794	20830	21730	22564	23074	23534	24092	24573
Jambi	6615	6774	7017	7204	7322	7503	7592	7758	7818
South Sumatra	14120	14725	15511	15961	16702	16971	17260	17571	18132
Bengkulu	5324	5510	5660	5812	5969	6030	6090	6217	6260
Lampung	10982	11365	11610	11817	12011	12165	12323	12515	12573
Kep. Bangka Belitung	1744	1772	1825	1847	1874	1895	1916	1941	1964
Kep. Riau	13851	14414	15489	16353	16700	17330	17559	17783	18121
Dki Jakarta	354131	367163	376207	382088	391225	396664	405632	415421	425445
West Java	79793	82047	86765	89076	91246	93526	95052	96939	99191
Central Java	32947	34447	35790	37306	38802	40054	40683	41144	42089
Di Yogyakarta	5801	6046	6265	6388	6546	6620	6676	6750	6812
East Java	115525	120476	123471	127841	130019	133344	135662	138296	141397
Banten	50232	51997	53164	54315	54699	55563	55920	56267	57035
Bali	10370	10758	11193	11515	11835	12099	12267	12427	12659
West Nusa Tenggara	5693	5787	5933	6061	6217	6302	6397	6533	6588
East Nusa Tenggara	10490	10735	11222	11442	11655	11809	11934	12128	12189
West Kalimantan	12425	12945	13281	13627	13850	14076	14292	14540	14700
Central Kalimantan	9108	9480	9716	9922	10229	10365	10467	10738	10877
South Kalimantan	6374	6641	6814	6991	7106	7221	7309	7459	7491
East Kalimantan	21299	22507	23708	24952	26357	27699	28285	28982	29606
North Kalimantan	3052	3252	3564	3749	3995	4244	4480	4722	4815
North Sulawesi	5419	5475	5614	5763	5822	5881	5978	6060	6152
Central Sulawesi	8802	9025	9231	9367	9470	9577	9653	9774	9823
South Sulawesi	18279	19406	20231	20864	21461	21970	22271	22551	22884
Southeast Sulawesi	6062	6238	6405	6561	6714	6861	6969	7083	7234
Gorontalo	2345	2411	2509	2602	2711	2782	2805	2834	2869
West Sulawesi	2098	2145	2190	2230	2286	2325	2325	2350	2348
Maluku	3448	3594	3704	3791	3873	3913	4193	4285	4277
North Maluku	2887	2959	3022	3073	3123	3223	3283	3359	3413
West Papua	7342	7953	8597	8721	8975	9155	9242	9381	9456
Papua	28417	29830	31582	31653	32421	32676	33234	33971	34274

From the data obtained, it is known that there are 9 columns. And obtained the values of max a and min a as follows:

Table 2. Max a and Min a Data for Each Column

Column to -	MIN	MAX
Column to 1	1744	354131
Column to 2	1772	367163
Column to 3	1825	376207
Column to 4	1847	382088
Column to 5	1874	391225
Column to 6	1895	396664
Column to 7	1916	405632
Column to 8	1941	415421
Column to 9	1964	425445

So that the normalization calculation is obtained as follows:

$$v^1 = \frac{7696-1744}{354131+1744}(354131-1744) + 1744 = 0{,}1689 \quad (1)$$

Table 3. Max a and Min a Data for Each Column

Provinsi	2010	2011	2012	2013	2014	2015	2016	2017	2018
Aceh	0.01689	0.01720	0.01770	0.01825	0.01826	0.01836	0.01817	0.01834	0.01806
North Sumatra	0.08393	0.08408	0.08472	0.08670	0.08720	0.08757	0.08732	0.08761	0.08758
West Sumatra	0.01783	0.01772	0.01761	0.01758	0.01786	0.01787	0.01775	0.01775	0.01747
Riau	0.04896	0.04932	0.05076	0.05229	0.05314	0.05365	0.05355	0.05357	0.05339
Jambi	0.01382	0.01369	0.01387	0.01409	0.01399	0.01421	0.01406	0.01407	0.01382
South Sumatra	0.03512	0.03545	0.03656	0.03712	0.03808	0.03819	0.03801	0.03780	0.03818
Bengkulu	0.01016	0.01023	0.01024	0.01043	0.01052	0.01047	0.01034	0.01034	0.01014
Lampung	0.02622	0.02625	0.02614	0.02622	0.02604	0.02602	0.02578	0.02557	0.02505
Kep. Bangka Belitung	0	0	0	0	0	0	0	0	0
Kep. Riau	0.03436	0.03460	0.03650	0.03815	0.03808	0.03910	0.03875	0.03831	0.03815
Dki Jakarta	1	1	1	1	1	1	1	1	1
West Java	0.22149	0.21970	0.22688	0.22940	0.22954	0.23211	0.23070	0.22975	0.22959
Central Java	0.08855	0.08942	0.09072	0.09325	0.09485	0.09666	0.09603	0.09481	0.09475
Di Yogyakarta	0.01151	0.01170	0.01186	0.01194	0.01200	0.01197	0.01179	0.01163	0.01145
East Java	0.32289	0.32487	0.32492	0.33135	0.32912	0.33298	0.33129	0.32977	0.32925
Banten	0.13760	0.13746	0.13713	0.13799	0.13567	0.13595	0.13377	0.13139	0.13004
Bali	0.02448	0.02459	0.02502	0.02543	0.02558	0.02585	0.02564	0.02536	0.02525
West Nusa Tenggara	0.01121	0.01099	0.01097	0.01108	0.01115	0.01116	0.01110	0.01111	0.01092
East Nusa Tenggara	0.02482	0.02453	0.02510	0.02523	0.02512	0.02511	0.02481	0.02464	0.02415
West Kalimantan	0.03031	0.03058	0.03060	0.03098	0.03076	0.03086	0.03066	0.03047	0.03007
Central Kalimantan	0.02090	0.02110	0.02108	0.02124	0.02146	0.02146	0.02118	0.02128	0.02105
South Kalimantan	0.01314	0.01333	0.01333	0.01353	0.01344	0.01349	0.01336	0.01335	0.01305
East Kalimantan	0.05549	0.05675	0.05845	0.06076	0.06288	0.06536	0.06532	0.06540	0.06527
North Kalimantan	0.00371	0.00405	0.00464	0.00500	0.00545	0.00595	0.00635	0.00673	0.00673
North Sulawesi	0.01043	0.01013	0.01012	0.01030	0.01014	0.01010	0.01006	0.00996	0.00989
Central Sulawesi	0.02003	0.01985	0.01978	0.01978	0.01951	0.01946	0.01916	0.01894	0.01856
South Sulawesi	0.04692	0.04826	0.04916	0.05001	0.05031	0.05085	0.05042	0.04985	0.04940
Southeast Sulawesi	0.01225	0.01222	0.01223	0.01240	0.01243	0.01258	0.01252	0.01244	0.01244
Gorontalo	0.00171	0.00175	0.00183	0.00199	0.00215	0.00225	0.00220	0.00216	0.00214
West Sulawesi	0.00100	0.00102	0.00097	0.00101	0.00106	0.00109	0.00101	0.00099	0.00091
Maluku	0.00484	0.00499	0.00502	0.00511	0.00513	0.00511	0.00564	0.00567	0.00546
North Maluku	0.00324	0.00325	0.00320	0.00322	0.00321	0.00336	0.00339	0.00343	0.00342
West Papua	0.01589	0.01692	0.01809	0.01808	0.01824	0.01839	0.01815	0.01799	0.01769
Papua	0.07569	0.07679	0.07948	0.07839	0.07846	0.07797	0.07757	0.07746	0.07630

For the selection of each medoid, it was chosen randomly, assuming DKI Jakarta and Papua as the initial medoids. The following is an initial medoid table.

Table 4. Early Medoids

Initial	2010	2011	2012	2013	2014	2015	2016	2017	2018
Dki Jakarta (C1)	1	1	1	1	1	1	1	1	1
Papua (C2)	0,0757	0,0768	0,0795	0,0784	0,0785	0,0780	0.0776	0.0775	0.0763

Calculate the value of the closest distance (cost) with the Euclidian Distance equation. To calculate the distance between the centroid point with the point of each object using Euclidian Distance. The formula for calculating distance uses equation (2.1). Then the calculation to calculate the distance of each object with the initial medoid at Cost 1 is as follows:

$$\text{Aceh (C1)} = \sqrt{\begin{matrix}((0.01689-1)^{\wedge}2+(0.01720-1)^{\wedge}2+(0.01770-1)^{\wedge}2+\\(0.01825-1)^2+(0.01826-1)^2+(0.01836-1)^2\\+(0.01817-1)^{\wedge}2+(0.01834-1)^{\wedge}2+(0.01806-1)^{\wedge}2)\end{matrix}} = 2{,}94626$$

$$\text{North Sumatra (C2)} = \sqrt{\begin{matrix}((0.08393-1)^{\wedge}2+(0.08408-1)^{\wedge}2+(0.08472-1)^{\wedge}2+\\(0.08670-1)^2+(0.08720-1)^2+(0.08757-1)^2\\+(0.08732-1)^{\wedge}2+(0.08761-1)^{\wedge}2+(0.08758-1)^{\wedge}2)\end{matrix}} = 2{,}7411$$

$$\text{West Sumatra (C3)} = \sqrt{\begin{matrix}((0.01783-1)^{\wedge}2+(0.01772-1)^{\wedge}2+(0.01761-1)^{\wedge}2+\\(0.01758-1)^2+(0.01786-1)^2+(0.01787-1)^2\\+(0.01775-1)^{\wedge}2+(0.01775-1)^{\wedge}2+(0.01747-1)^{\wedge}2)\end{matrix}} = 2{,}94686$$

So on until the 24th data (Papua Province). While the calculation to calculate the distance of each object with the initial medoid at Cost 2 is as follows:

$$\text{Aceh (C1)} = \sqrt{\begin{matrix}((0.01689 - 0{,}0757)\string^2 + (0.01720 - 0{,}0768)\string^2 + (0.01770 - 0{,}0795)\string^2 + \\ (0.01825 - 0{,}0784)^2 + (0.01826 - 0{,}0785)^2 + (0.01836 - 0{,}0780)^2 \\ +(0.01817 - 0.0776)\string^2 + (0.01834 - 0.0775)\string^2 + (0.01806 - 0.0763)\string^2)\end{matrix}} = 2{,}94626$$

$$\text{North Sumatra (C2)} = \sqrt{\begin{matrix}((0.08393 - 0{,}0757)\string^2 + (0.08408 - 0{,}0768)\string^2 + (0.08472 - 0{,}0795)\string^2 + \\ (0.08670 - 0{,}0784)^2 + (0.08720 - 0{,}0785)^2 + (0.08757 - 0{,}0780)^2 \\ +(0.08732 - 0.0776)\string^2 + (0.08761 - 0.0775)\string^2 + (0.08758 - 0.0763)\string^2)\end{matrix}} = 2{,}7411$$

$$\text{West Sumatra (C3)} = \sqrt{\begin{matrix}((0.01783 - 0{,}0757)\string^2 + (0.01772 - 0{,}0768)\string^2 + (0.01761 - 0{,}0795)\string^2 + \\ (0.01758 - 0{,}0784)^2 + (0.01786 - 0{,}0785)^2 + (0.01787 - 0{,}0780)^2 \\ +(0.01775 - 0.0776)\string^2 + (0.01775 - 0.0775)\string^2 + (0.01747 - 0.0763)\string^2)\end{matrix}} = 2{,}94686$$

So on until the 24th data (Papua Province). The overall results can be seen in table 5 below :

Table 5. Nearest Distance Value (Cost) in the First Iteration

No	Cost 1	Cost 2	At close range	Cluster
1	2.94626	0.17898	0.17898	2
2	2.74110	0.02667	0.02667	2
3	2.94686	0.17959	0.17959	2
4	2.84379	0.07668	0.07668	2
5	2.95813	0.19086	0.19086	2
6	2.88850	0.12126	0.12126	2
7	2.96904	0.20177	0.20177	2
8	2.92224	0.15498	0.15498	2
9	3.00000	0.23273	0.23273	2
10	2.88801	0.12080	0.12080	2
11	0.00000	2.76730	0	1
12	2.31698	0.45048	0.45048	2
13	2.72033	0.04767	0.04767	2
14	2.96472	0.19744	0.19744	2
15	2.01454	0.75284	0.75284	2
16	2.59435	0.17314	0.17314	2
17	2.92426	0.15700	0.15700	2
18	2.96677	0.19950	0.19950	2
19	2.92550	0.15822	0.15822	2
20	2.90824	0.14097	0.14097	2
21	2.93643	0.16916	0.16916	2
22	2.96000	0.19273	0.19273	2
23	2.81479	0.04880	0.04880	2
24	2.98380	0.21654	0.21654	2
25	2.96962	0.20236	0.20236	2
26	2.94164	0.17438	0.17438	2
27	2.85161	0.08436	0.08436	2
28	2.96283	0.19556	0.19556	2
29	2.99395	0.22667	0.22667	2
30	2.99698	0.22971	0.22971	2
31	2.98434	0.21708	0.21708	2
32	2.99009	0.22282	0.22282	2
33	2.94686	0.17958	0.17958	2
34	2.76730	0	0	2
Amount	**94,53984**	**8,88863**	**6,121335**	
Total Cost	**103,42847**			

After getting the results of the distance from each object (cost) in the 1st iteration, then proceed to the 2nd iteration. Initialization of new medoid candidates (non-medoid) in the 2nd iteration can be seen in table 6. below:

Table 6. Second Medoids

Initial	2010	2011	2012	2013	2014	2015	2016	2017	2018
North Sumatra	0.08393	0.08408	0.08472	0.08670	0.08720	0.08757	0.08732	0.08761	0.08758
Banten	0.13760	0.13746	0.13713	0.13799	0.13567	0.13595	0.13377	0.13139	0.13004

Then the calculation to calculate the distance of each object with the initial medoid at Cost 1 is as follows:

$$\text{Aceh (C1)} = \sqrt{\begin{array}{c}((0.01689 - 0.0839)^\wedge 2 + (0.01720 - 0.08408)^\wedge 2 + (0.01770 - 0.08472)^\wedge 2 + \\ (0.01825 - 0.08670)^2 + (0.01826 - 0.08720)^2 + (0.01836 - 0.08757)^2 \\ +(0.01817 - 0.08732)^\wedge 2 + (0.01834 - 0.08761)^\wedge 2 + (0.01806 - 0.08758)^\wedge 2)\end{array}} = 0{,}20518$$

$$\text{North Sumatra (C2)} = \sqrt{\begin{array}{c}((0.08393 - 0.08393)^\wedge 2 + (0.08408 - 0.08408)^\wedge 2 + (0.08472 - 0.08472)^\wedge 2 + \\ (0.08670 - 0.08670)^2 + (0.08720 - 0.08720)^2 + (0.08757 - 0.08757)^2 \\ +(0.08732 - 0.08732)^\wedge 2 + (0.08761 - 0.08761)^\wedge 2 + (0.08758 - 0.08758)^\wedge 2)\end{array}} = 0$$

$$\text{West Sumatra (C3)} = \sqrt{\begin{array}{c}((0.01783 - 0.08393)^\wedge 2 + (0.01772 - 0.08408)^\wedge 2 + (0.01761 - 0.08472)^\wedge 2 + \\ (0.01758 - 0.08670)^2 + (0.01786 - 0.08720)^2 + (0.01787 - 0.08757)^2 \\ +(0.01775 - 0.08732)^\wedge 2 + (0.01775 - 0.08761)^\wedge 2 + (0.01747 - 0.08758)^\wedge 2)\end{array}} = 0{,}20581$$

So on until the 24th data (Papua Province). While the calculation to calculate the distance of each object with the initial medoid at Cost 2 is as follows:

$$\text{Aceh (C1)} = \sqrt{\begin{array}{c}((0.01689 - 0.13760)^\wedge 2 + (0.01720 - 0.13746)^\wedge 2 + (0.01770 - 0.13713)^\wedge 2 + \\ (0.01825 - 0.13799)^2 + (0.01826 - 0.13567)^2 + (0.01836 - 0.13595)^2 \\ +(0.01817 - 0.13377)^\wedge 2 + (0.01834 - 0.13139)^\wedge 2 + (0.01806 - 0.13004)^\wedge 2)\end{array}} = 0{,}35203$$

$$\text{North Sumatra (C2)} = \sqrt{\begin{array}{c}((0.08393 - 0.13760)^\wedge 2 + (0.08408 - 0.13746)^\wedge 2 + (0.08472 - 0.13713)^\wedge 2 + \\ (0.08670 - 0.13799)^2 + (0.08720 - 0.13567)^2 + (0.08757 - 0.13595)^2 \\ +(0.08732 - 0.13377)^\wedge 2 + (0.08761 - 0.13139)^\wedge 2 + (0.08758 - 0.13004)^\wedge 2)\end{array}} = 0{,}14722$$

$$\text{West Sumatra (C3)} = \sqrt{\begin{array}{c}((0.01783 - 0.13760)^\wedge 2 + (0.01772 - 0.13746)^\wedge 2 + (0.01761 - 0.13713)^\wedge 2 + \\ (0.01758 - 0.13799)^2 + (0.01786 - 0.13567)^2 + (0.01787 - 0.13595)^2 \\ +(0.01775 - 0.13377)^\wedge 2 + (0.01775 - 0.13139)^\wedge 2 + (0.01747 - 0.13004)^\wedge 2)\end{array}} = 0{,}35261$$

So on until the 24th data (Papua Province). The overall results can be seen in table 7 as follows:

Table 7. Nearest Distance Value (Cost) in the Second Iteration

No	Cost 1	Cost 2	At close range	Cluster
1	0,20518	0.35203	0,20518	1
2	0	0.14722	0	1
3	0.20581	0.35261	0.20581	1
4	0.10270	0.24976	0.10270	1
5	0.21707	0.36388	0.21707	1
6	0.14741	0.29435	0.14741	1
7	0.22798	0.37479	0.22798	1
8	0.18122	0.32798	0.18122	1
9	0.25894	0.40575	0.25894	1
10	0.14691	0.29389	0.14691	1
11	2.74110	2.59435	2.59435	2
12	0.42423	0.27794	0.27794	2
13	0.02119	0.12683	0.02119	1
14	0.22366	0.37046	0.22366	1
15	0.72661	0.58001	0.58001	2
16	0.14722	0	0	2
17	0.18320	0.33004	0.18320	1
18	0.22572	0.37252	0.22572	1
19	0.18446	0.33124	0.18446	1

20	0.16720	0.31400	0.16720	1
21	0.19537	0.34219	0.19537	1
22	0.21895	0.36575	0.21895	1
23	0.07401	0.22119	0.07401	1
24	0.24271	0.38961	0.24271	1
25	0.22858	0.37537	0.22858	1
26	0.20062	0.34737	0.20062	1
27	0.11053	0.25746	0.11053	1
28	0.22177	0.36859	0.22177	1
29	0.25288	0.39970	0.25288	1
30	0.25592	0.40272	0.25592	1
31	0.24328	0.39011	0.24328	1
32	0.24903	0.39584	0.24903	1
33	0.20578	0.35264	0.20578	1
34	0.02667	0.17314	0.02667	1
Amount	**9,4639**	**13,2413**	**9,877**	
Total Cost	**22,7052**			

Calculate Total Deviation (S).

After obtaining the distance value between the 1st and 2nd iterations, calculate the total deviation (S) by finding the difference between the new total cost and the old total cost. With the condition that if S < 0, then exchange the object value by specifying a new medoid.

S = new total cost - old total cost =9,877- 6,121335 = 2,7557

Because the value of S> 0 then the iteration is stopped.

2.2. Testing With RapidMiner

To simplify the data in the application, the year is initialized with a variable x1-x9. Display data as shown below:

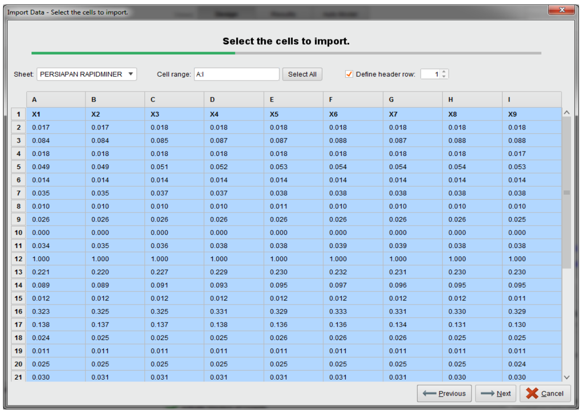

Figure 1. Display of Import Data on RapidMiner

The picture above shows the data retrieval process to be processed, after that select next and next again and select the data type where the attribute text is changed to "id" then select finish on the data so that a display like the image below will appear:

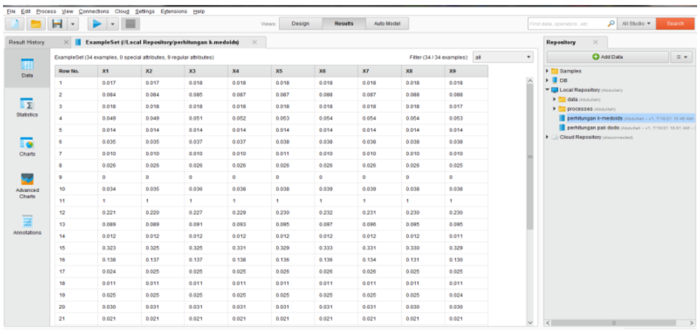

Figure 2. RapidMiner Import Data Display

To run data processing so as to get the final result in a system output. The following is a display of the final results of the above data processing:

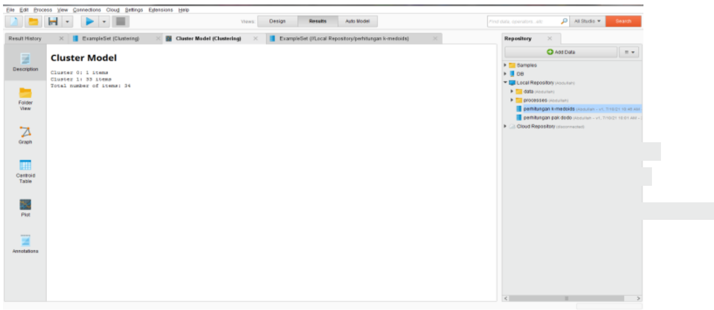

Figure 3. RapidMiner Model Cluster Value

Information :
1. The number of Cluster 0 (Low) is 1items
2. The number of Cluster 1 (High) is 33items
3. The total number of items is 34

So that it can be seen the results of the grouping of Rapidminer 8.1 can be seen from the image below:

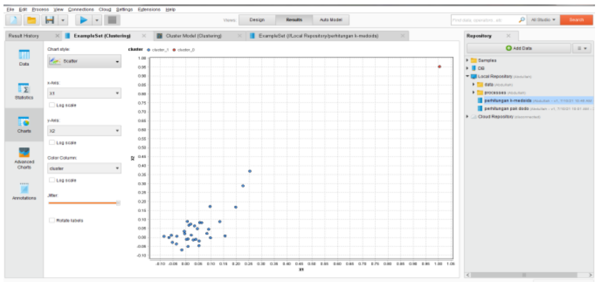

Figure 4. Graph of the Final Result of RapidMiner Grouping

The results of cluster calculations using the k-Medoids algorithm with calculations using the RapidMiner application.

Table 8. Calculation Results of K-Medoids with RapidMiner Application

No	Province	K-Medoids Method Data Validation	
		Manual	Rapidminner
1	Aceh	Cluster_2	Cluster_2
2	North Sumatra	Cluster_2	Cluster_2
3	West Sumatra	Cluster_2	Cluster_2
4	Riau	Cluster_2	Cluster_2
5	Jambi	Cluster_2	Cluster_2
6	South Sumatra	Cluster_2	Cluster_2
7	Bengkulu	Cluster_2	Cluster_2
8	Lampung	Cluster_2	Cluster_2
9	Kep. Bangka Belitung	Cluster_2	Cluster_2
10	Kep. Riau	Cluster_2	Cluster_2
11	Dki Jakarta	Cluster_1	Cluster_1
12	West Java	Cluster_2	Cluster_2
13	Central Java	Cluster_2	Cluster_2
14	Di Yogyakarta	Cluster_2	Cluster_2
15	East Java	Cluster_2	Cluster_2
16	Banten	Cluster_2	Cluster_2
17	Bali	Cluster_2	Cluster_2
18	West Nusa Tenggara	Cluster_2	Cluster_2
19	East Nusa Tenggara	Cluster_2	Cluster_2
20	West Kalimantan	Cluster_2	Cluster_2
21	Central Kalimantan	Cluster_2	Cluster_2
22	South Kalimantan	Cluster_2	Cluster_2
23	East Kalimantan	Cluster_2	Cluster_2
24	North Kalimantan	Cluster_2	Cluster_2
25	North Sulawesi	Cluster_2	Cluster_2
26	Central Sulawesi	Cluster_2	Cluster_2
27	South Sulawesi	Cluster_2	Cluster_2
28	Southeast Sulawesi	Cluster_2	Cluster_2

29	Gorontalo	Cluster_2	Cluster_2
30	West Sulawesi	Cluster_2	Cluster_2
31	Maluku	Cluster_2	Cluster_2
32	North Maluku	Cluster_2	Cluster_2
33	West Papua	Cluster_2	Cluster_2
34	Papua	Cluster_2	Cluster_2

3. Conclusion

Based on the previous discussion, it can be concluded that the application of data mining using the k-medoids algorithm in the grouping of permanent workers in construction companies in Indonesia can be applied. Sources of data used in this study is data obtained from BPS (Central Bureau of Statistics). The amount of data used is 34 provinces consisting of 2010 2018. From the grouping results, two clusters are obtained, namely high and low. The high cluster consists of 33 items and the low cluster consists of 1 item. Data testing on Rapidminer 8.1 using the k-means algorithm can display the accuracy of the data between manual and system calculations.

References

[1] Yuli Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika," *J. Edik Inform.*, vol. 2, 2019.

[2] D. Firdaus, "Penggunaan Data Mining dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer," *J. Format*, vol. 6, no. 2, pp. 91–97, 2017.

[3] M. Y. Helmy, Kushartantya, and N. Bahtiar, "Implementasi Data Mining Untuk Memprediksi Kelayakan Permintaan Pinjaman Nasabah Di Lembaga Keuangan," *J. Informatics Technol.*, vol. 2, no. 1, pp. 33–42, 2013.

[4] N. Putu, E. Merliana, and A. J. Santoso, "Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means," pp. 978–979.

[5] W. M. S. Karsito, "Karsito, Winda Monika Sari," vol. 9, no. September, pp. 67–78, 2018.

[6] A. Maddeppungeng, "Alasan Utama Perusahaan Jasa Kontruksi Melakukan Investasi Teknologi Informasi," *Poli-Teknologi*, vol. 9, no. 1, p. 160282, 2010.

[7] E. Widajanti, "PERENCANAAN SUMBERDAYA MANUSIA YANG EFEKTIF: STRATEGI MENCAPAI KEUNGGULAN KOMPETITIF Erni Widajanti Fakultas Ekonomi Universitas Slamet Riyadi Surakarta," *J. Ekon. dan Kewirausahaan*, vol. 7, no. 2, pp. 105–114, 2007.

[8] Augustinus & Eric, "Pengelolaan SDM Pada PT. Aneka Sejahtera Engineering," *Pengelolaan SDM Pada PT. Aneka Sejah. Eng.*, vol. 1, no. 2, 2013.

[9] S. Sugriyono and M. U. Siregar, "Preprocessing kNN algorithm classification using K-means and distance matrix with students' academic performance dataset," *J. Teknol. dan Sist. Komput.*, vol. 8, no. 4, pp. 311–316, 2020, doi: 10.14710/jtsiskom.2020.13874.

[10] R. Franita, "Implementasi Sinergitas Lembaga Pemerintah Untuk Mendukung Budaya Sadar Bencana di Kota Balikpapan," *Nusant. J. Ilmu Pengetah. Sos.*, vol. 7, no. 2, pp. 408–420, 2020.

[11] D. Rudianto, "Analisis Perbandingan Kinerja Keuangan PT. Telkom, Tbk dengan PT. Indosat, Tbk Periode 2005-2010," *J. Nas. Univ. Bakrie Jakarta, ISSN*, pp. 2089–3590, 2012.

[12] A. U. Fitriyadi and A. Kurniawati, "Analisis Algoritma K-Means dan K-Medoids Untuk Clustering Data Kinerja Karyawan Pada Perusahaan Perumahan Nasional," *Kilat*, vol. 10, no. 1, pp. 157–168, 2021.

[13] Z. Mustofa and I. S. Suasana, "Algoritma Clustering K-Medoids Pada E-Government Bidang Information And Communication," *J. Teknol. dan Komun.*, vol. 9, pp. 1–10, 2018.

[14] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. Ilmi R.H.Zer, and D. Hartama, "Analisis algoritma K-Medoids clustering dalam pengelompokan penyebaran Covid-19 di Indonesia," *Jti (Jurnal Teknol. Informasi)*, vol. 4, no. 1, pp. 166–173, 2020.

[15] P. Kumar and D. Sirohi, "Comparative analysis of FCM and HCM algorithm on Iris data set," *Int. J. Comput. Appl.*, vol. 5, no. 2, pp. 33–37, 2010, doi: 10.5120/888-1261.