Penerapan Algoritma K-Means Dalam Pengelompokan Pekerja Tetap Perusahaan Konstruksi Menurut Provinsi

Khairwa Bakhsar[1], Widodo Saputra[2], Heru Satria Tambunan[3]
[1,3]*STIKOM Tunas Bangsa, Pematangsiantar, Medan – Indonesia*
[2]*AMIK Tunas Bangsa, Pematangsiantar, Medan – Indonesia*
Jln. Sudirman Blok A No. 1-3 Pematangsiantar, Sumatera Utara
[1]*kbakhsar@gmail.com,* [2]*widodosaputra01@gmail.com,* [3]*heru@amiktunasbangsa.ac.id*

Abstract

A construction company is one of the businesses in the economic sector that deals with planning planning and supervision of construction activities to form a building. Permanent workers, namely workers who have a work agreement with the employer for a period of time not based on a decision letter. Based on data obtained from the official Government website which is located at https://www.bps.go.id/ which displays several numbers that vary from 2010 to 2018 according to provinces in Indonesia. For this reason, the data will be processed by clustering in 2 clusters, namely the highest and lowest using the Data Mining method with the K-Means Algorithm. In this research technique, the data will be grouped based on the name of the province that has the number of permanent construction company workers from 2010 to 2018. The Data Mining method is used to facilitate data grouping of the number of construction company permanent workers according to provinces in Indonesia. The results obtained from the decision based on the number of permanent workers in construction companies by province with the highest 1 province, namely DKI Jakarta and 33 other provinces including the lowest level. It is hoped that this research can be an input to the government, provinces that have become more concerned with construction companies based on the clusters that have been carried out.

Keywords: *Data Mining, K-Means, Construction Workers, Clustering, Rapidminer*

Abstrak

Perusahaan konstruksi adalah salah satu usaha dalam sector ekonomi yang berhubungan dengan suatu perencanaan atau pelaksanaan dan pengawasan suatu kegiatan konstruksi untuk membentuk suatu bangunan. Tenaga kerja tetap yaitu pekerja yang memiliki perjanjian kerja dengan pengusaha untuk jangka waktu tidak tertentu berdasarkan surat keputusan. Berdasarkan data yang diperoleh dari situs resmi Pemerintah yang beralamatkan di https://www.bps.go.id. yang menampilkan beberapa jumlah yang bervariasi mulai dari tahun 2010 sampai 2018 menurut provinsi di Indonesia. Untuk itu data akan diolah dengan melakukan clustering dalam 2 cluster yaitu cluster tertinggi dan terendah menggunakan metode Data Mining dengan Algoritma K-Means. Teknik penelitian ini, data akan dikelompokan berdasarkan nama provinsi yang memiliki jumlah pekerja tetap perusahaan konstruksi dari tahun 2010 sampai 2018. Itu sebabnya digunakan metode Data Mining untuk memudahkan pengelompokan data jumlah pekerja tetap perusahaan konstruksi menurut provinsi di Indonesia. Hasil yang diperoleh dari proses penilaian berdasarkan Banyaknya Pekerja Tetap Perusahaan Konstruki Menurut Provinsi dengan 1 provinsi tertinggi yaitu DKI Jakarta dan 33 provinsi lainnya termasuk tingkat terendah. Diharapkan penelitian ini dapat menjadi masukan kepada pemerintah, provinsi yang menjadi perhatian lebih pada perusahaan konstruksi berdasarkan cluster yang telah dilakukan.

Kata Kunci: *Data Mining, K-Means, Pekerja Konstruksi, Clustering, Rapidminer*

ISSN: 2720-992X
Copyright © KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)

1. Pendahuluan

Perusahaan konstruksi adalah perusahaan yang menjalankan usaha di bidang pekerjaan konstruksi, baik pelaksana konstruksi maupun konsultan konstruksi. Dalam menjalankan aktivitasnya, perusahaan menggunakan suatu bentuk strategi sumber daya manusia, dimana untuk memenuhi kebutuhan sumber daya manusia perusahaan. Perusahaan menggunakan dua jenis pekerja, yaitu pekerja tetap (yang berasal dari dalam perusahaan) dan *outsourcee* (pekerja yang berasal dari luar perusahaan/pekerja kontrak). Pada industri konstruksi, pekerja merupakan penggerak utama kegiatan-kegiatan konstruksi, mereka adalah komponen yang sangat penting dalam perusahaan.

Dalam penelitian ini penulis memperoleh data dari situs resmi pemerintah Badan Pusat Statistik yang beralamatkan *https://www.bps.go.id* diperoleh sejumlah data cukup banyak yaitu data pekerja tetap perusahaan konstruksi menurut provinsi di Indonesia mulai dari tahun 2010 sampai 2018. Oleh karena itu, penggunaan *Data Mining* dengan Algoritma *K-Means* dapat mempermudah pengelompokan data [1] yang berjumlah besar dan menentukan *cluster* setiap variabel mulai dari terendah sampai tertinggi sehingga dapat memberikan informasi dan masukan kepada pusat pemerintah daerah pada perusahaan konstruksi berdasarkan provinsi di Indonesia.

K-Means merupakan salah satu Algoritma *Clustering.* Tujuan Algoritma ini yaitu membagi data menjadi beberapa kelompok. Algoritma *K-Means* memiliki kemampuan dalam pengelompokan data dalam jumlah yang cukup besar. Metode *K-Means* berusaha mengelompokkan data yang ada kedalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain [2]. Algoritma *K-Means* diterapkan untuk mengetahui provinsi mana saja yang memiliki *cluster* tertinggi dan terendah dari data pekerja tetap perusahaan konstruksi dan Memudahkan dalam pengelompokan pekerja tetap perusahaan konstruksi berdasarkan provinsi di Indonesia. Hasil dari permasalahan ini diharapkan dapat bermanfaat bagi pemerintah daerah yang memiliki pekerja tetap perusahaan konstruksi di Indonesia.

2. Metodologi Penelitian

2.1. Data Mining

Data Mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Metode untuk menganalisis data dalam penerapan data mining ini adalah *Knowledge Discovery in Database* (KDD) yang terdiri dari beberapa tahapan, yaitu *Cleaning and Integration*, Selection and Transformation, Data Mining, dan *Evaluation* and *Interpretation* [3].

Istilah *Knowledge Discovery in Database* (KDD) dan Data Mining seringkali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain, dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining [4]. Pengelompokkan Data Mining dibagi menjadi beberapa kelompok antara lain Deskripsi, Estimasi, Prediksi, Klasifikasi, dan Pengklasteran [5].

2.2. Algoritma *K-means*

K-Means merupakan suatu Algoritma yang digunakan dalam pengelompokkan secara partisi yang memisahkan data ke dalam kelompok yang berbeda–beda. Algoritma ini mampu meminimalkan jarak antara data ke klasternya. Pada dasarnya penggunaan Algoritma ini dalam proses *clustering* tergantung pada data yang didapatkan dan konklusi yang ingin dicapai di akhir proses [6]. Inti dari Algoritma *K-Means* adalah klasterisasi pengelompokan data berdasarkan titik pusat klaster

(*centroid*) untuk memaksimalkan kemiripan data yang didapatkan berdasarkan jarak terpendek data terhadap titik *centroid* awal [7].

Langkah-langkah dalam algoritma *K-means* [8]:

a) Tentukan jumlah *cluster* (k) pada data set
b) Tentukan nilai pusat (*centroid*)
   Penentuan nilai *centroid* pada tahap awal dilakukan secara *random* atau dapat diambil dari nilai maksimum untuk *cluster* tinggi dan nilai minimum untuk *cluster* rendah.
c) Pada masing-masing *record*, hitung jarak terdekat dengan *centroid*. Jarak *centroid* yang digunakan adalah *Euclidean Distance*, dengan rumus seperti dibawah ini:

$$D(x_2, x_1) = \sqrt{\sum_{j=1}^{p} |x_{2j} - x_{1j}|^2} \quad (1)$$

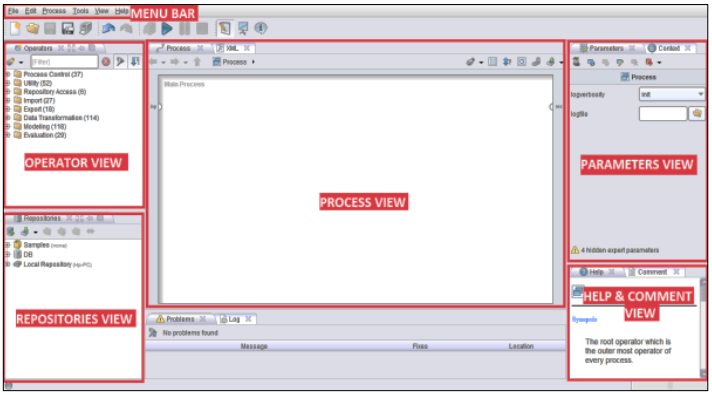
   Keterangan :
   D = *Euclidean Distance*
   x = Banyaknya objek
   Σp = Jumlah data *record*

$$De = \sqrt{(xi - si)^2 + (yi - ti)^2} \quad (2)$$

   Keterangan :
   De = *Eulidean Distance*
   i = Banyaknya objek [2]
   (x, y)= Koordinat objek
   (s, t) = Koordinat *centroid*
d) Kelompokkan objek berdasarkan jarak ke *centroid* terdekat untuk membuat *centroid* baru. *Centroid* baru diambil dari penjumlahan nilai berdasarkan jarak dari iterasi sebelumnya lalu dibagi dengan jumlah jarak dari tiap *cluster*.
e) Ulangi langkah ke-2, lakukan *iterasi* hingga *centroid* bernilai optimal.

2.3. *RapidMiner*

RapidMiner adalah salah satu aplikasi *opensource* yang dapat digunakan untuk melakukan proses *data mining*. Salah satu metode *data mining* adalah menggunakan regresi linier. Regresi linier merupakan metode statistik yang digunakan untuk melakukan estimasi atau perkiraan berdasarkan data yang ada [9]. Penelitian yang dilakukan penulis menggunakan tool *RapidMiner Studio versi 5.3*, tampilan *RapidMiner Studio versi 5.3* terdapat pada Gambar 1. :



Gambar 1. Tampilan *RapidMiner Studio* Versi 5.3.

3. Hasil dan Pembahasan

Sumber data dalam penelitian ini diperoleh dari situs resmi Pemerintah yang beralamatkan *https://www.bps.go.id*.. Data yang diambil adalah data Banyaknya

Pekerja Tetap Perusahaan Konstruksi Menurut Provinsi di Indonesia tahun 2010-2018 yang terdiri dari 34 Provinsi. Data ini akan menggunakan 2 *cluster* yaitu tinggi dan rendah dalam pengolahannya. Dalam melakukan *clustering*, data yang diperoleh akan dihitung terlebih dahulu. Pada penelitian ini digunakan aplikasi *RapidMiner 5.3.* sebagai pengujian data yang di analisa. Sampel data yang akan digunakan dalam proses *clustering* pada data banyaknya pekerja tetap perusahaan konstruksi menurut provinsi. Berikut daftar tabel penggunaan data dapat dilihat pada tabel 1 :

Tabel 1. Banyaknya Pekerja Tetap Perusahaan Konstruksi

Provinsi	Banyaknya Pekerja Tetap Perusahaan Konstruksi (Orang)								
	2010	2011	2012	2013	2014	2015	2016	2017	2018
Aceh	7696	8058	8450	8786	8983	9143	9252	9524	9614
Sumatera Utara	31321	32494	33542	34813	35827	36464	37167	38168	39051
Sumatera Barat	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
Sumatera Selatan	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 Belitu	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
Jawa Barat	79793	82047	86765	89076	91246	93526	95052	96939	99191
Jawa Tengah	32947	34447	35790	37306	38802	40054	40683	41144	42089
Di Yogyakarta	5801	6046	6265	6388	6546	6620	6676	6750	6812
Jawa Timur	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
Nusa Tenggara Bara	5693	5787	5933	6061	6217	6302	6397	6533	6588
Nusa Tenggara Tim	10490	10735	11222	11442	11655	11809	11934	12128	12189
Kalimantan Barat	12425	12945	13281	13627	13850	14076	14292	14540	14700
Kalimantan Tengah	9108	9480	9716	9922	10229	10365	10467	10738	10877
Kalimantan Selatan	6374	6641	6814	6991	7106	7221	7309	7459	7491
Kalimantan Timur	21299	22507	23708	24952	26357	27699	28285	28982	29606
Kalimantan Utara	3052	3252	3564	3749	3995	4244	4480	4722	4815
Sulawesi Utara	5419	5475	5614	5763	5822	5881	5978	6060	6152
Sulawesi Tengah	8802	9025	9231	9367	9470	9577	9653	9774	9823
Sulawesi Selatan	18279	19406	20231	20864	21461	21970	22271	22551	22884
Sulawesi Tenggara	6062	6238	6405	6561	6714	6861	6969	7083	7234
Gorontalo	2345	2411	2509	2602	2711	2782	2805	2834	2869
Sulawesi Barat	2098	2145	2190	2230	2286	2325	2325	2350	2348
Maluku	3448	3594	3704	3791	3873	3913	4193	4285	4277
Maluku Utara	2887	2959	3022	3073	3123	3223	3283	3359	3413
Papua Barat	7342	7953	8597	8721	8975	9155	9242	9381	9456
Papua	28417	29830	31582	31653	32421	32676	33234	33971	34274

a) Menentukan Nilai k Jumlah *Cluster*
   Jumlah *cluster* sebanyak 2 *cluster*. *Cluster* yang dibentuk yaitu *cluster* tinggi (C1) dan *cluster* rendah (C2).
b) Menentukan Nilai *Centroid* (Pusat *Cluster*)
   Penentuan pusat *cluster* awal ditentukan secara *random* yang diambil dari data yang ada dalam *range*. Adapun nilai untuk *cluster* tinggi (*cluster* 1) diambil dari nilai tertinggi yang terdapat pada tabel 1 dan nilai untuk *cluster* rendah (*cluster* 2) diambil dari nilai terendah yang terdapat pada tabel 1. Berikut daftar tabel *centroid* data dapat dilihat pada tabel 2 :

Tabel 2. *Centroid* Data Awal

	2010	2011	2012	2013	2014	2015	2016	2017	2018
***Cluster* 1**	354131	367163	376207	382088	391225	396664	405632	415421	425445
***Cluster* 2**	1744	1772	1825	1847	1874	1895	1916	1941	1964

c) Menghitung Jarak Setiap Data Terhadap *Centroid* (Pusat *Cluster*)
   Setelah data nilai pusat *cluster* awal ditentukan, maka langkah selanjutnya adalah menghitung jarak masing-masing data terhadap pusat *cluster*. Proses

pencarian jarak terpendek pada iterasi 1 dapat dilihat pada perhitungan dan tabel dibawah ini :

$$D_{Aceh,c1} = \sqrt{\begin{array}{l}(7696-354131)^2+(8058-367163)^2+\\(8450-376207)^2+(8786-382088)^2+\\(8983-391225)^2+(9143-396664)^2+\\(9252-405632)^2+(9524-415421)^2+\\(9614-425445)^2\end{array}} = 1081009$$

$$D_{Aceh,c2} = \sqrt{\begin{array}{l}(7696-1744)^2+(8058-1772)^2+\\(8450-1825)^2+(8786-1847)^2+\\(8983-1874)^2+(9143-1895)^2+\\(9252-1916)^2+(9524-1941)^2+\\(9614-1964)^2\end{array}} = 20974$$

Hasil dari keseluruhan perhitungan dapat dilihat pada tabel 3 sebagai berikut:

Tabel 3. Hasil Perhitungan Jarak Pusat *Cluster* Iterasi 1

No	Provinsi	C1	C2	Jarak Terpendek	C1	C2
1	Aceh	1081009	20974	20974		1
2	Sumatera Utara	1006237	100944	100944		1
3	Sumatera Barat	1081259	20678	20678		1
4	Riau	1043535	61031	61031		1
5	Jambi	1085357	16305	16305		1
6	Sumatera Selatan	1059896	43540	43540		1
7	Bengkulu	1089384	12052	12052		1
8	Lampung	1072290	30223	30223		1
9	Kep. Bangka Belitung	1100710	0	0		1
10	Kep. Riau	1059644	43798	43798		1
11	Dki Jakarta	389351	1167543	389351	1	
12	Jawa Barat	854633	266261	266261		1
13	Jawa Tengah	998691	109187	109187		1
14	Di Yogyakarta	1087812	13730	13730		1
15	Jawa Timur	749952	383837	383837		1
16	Banten	953747	157587	157587		1
17	Bali	1072976	29516	29516		1
18	Nusa Tenggara Barat	1088541	12930	12930		1
19	Nusa Tenggara Timur	1073478	28980	28980		1
20	Kalimantan Barat	1067143	35705	35705		1
21	Kalimantan Tengah	1077445	24751	24751		1
22	Kalimantan Selatan	1086063	15566	15566		1
23	Kalimantan Timur	1032722	72670	72670		1
24	Kalimantan Utara	1094630	6557	6557		1
25	Sulawesi Utara	1089592	11806	11806		1
26	Sulawesi Tengah	1079394	22661	22661		1
27	Sulawesi Selatan	1046461	57864	57864		1
28	Sulawesi Tenggara	1087076	14478	14478		1
29	Gorontalo	1098485	2388	2388		1
30	Sulawesi Barat	1099611	1175	1175		1
31	Maluku	1094922	6134	6134		1
32	Maluku Utara	1097053	3867	3867		1
33	Papua Barat	1081244	20759	20759		1
34	Papua	1015910	90575	90575		1

Dalam menentukan posisi *cluster* masing-masing data hasil Banyaknya Pekerja Tetap Perusahaan Konstruksi Menurut Provinsi berdasarkan jarak minimum data terhadap pusat *cluster*.

d) Menghitung *centroid* baru menggunakan hasil dari setiap anggota pada masing-masing *cluster*. Setelah didapatkan hasil jarak dari setiap objek pada iterasi ke-1 maka lanjut ke iterasi ke-2 pada perhitungan dan tabel dibawah :

$$D_{2010,c1} = \frac{354131}{1} = 354131$$

$$D_{2011,c1} = \frac{367163}{1} = 367163$$

$$D_{2010,c2} = \frac{\begin{matrix}7696+31321+8028+18996+6615+14120+5324+10982+1744+\\ 13851+79793+32947+5801+115525+50232+10370+5693+\\ 10490+12425+9108+6374+21299+3052+5419+8802+18279+\\ 6062+2345+2098+3448+2887+7342+28417\end{matrix}}{33} = 17178.33$$

$$D_{2011,c2} = \frac{\begin{matrix}8058+32494+8247+19794+6774+14725+5510+11365+1772+\\ 14414+82047+34447+6064+120476+51997+10758+5787+\\ 10735+12945+9480+6641+22507+3252+5475+9025+19406+\\ 6238+2411+2145+3594+2959+7953+29830\end{matrix}}{33} = 17857.79$$

Tabel 4. *Centroid* Baru Iterasi 1

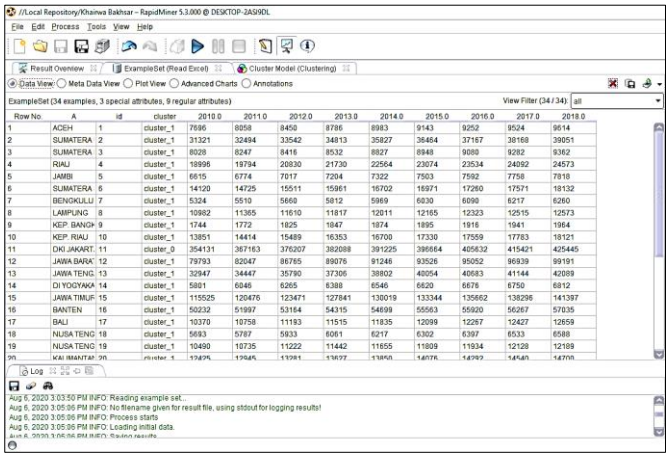
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
***Cluster* 1**	354131	367153	376207	382088	391225	396664	405632	415421	425445
***Cluster* 2**	17178,33	17857,79	18555,18	19111,06	19581,55	20024,48	20337,88	20709,82	21080,21

e) Selanjutnya dilakukan kembali langkah ke 4 sampai 6. Langkah selanjutnya adalah melakukan perhitungan jarak terhadap data banyaknya pekerja tetap perusahaan konstruksi dengan menggunakan data centroid baru. Berikut hasil dari perhitungan iterasi 2 dapat dilihat pada tabel 5 :

Tabel 5. Hasil Perhitungan Jarak Pusat *Cluster* Iterasi 2

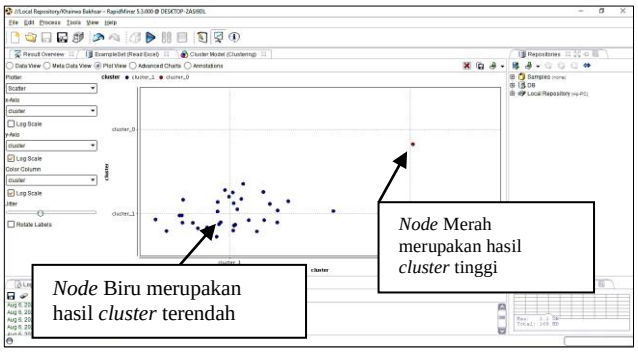
No	Provinsi	C1	C2	Jarak Terpendek	C1	C2
1	Aceh	1146574,20	31700,18	31700,18		1
2	Sumatera Utara	1066611,72	48275,44	48275,44		1
3	Sumatera Barat	1146862,88	31997,12	31997,12		1
4	Riau	1106544,39	8434,16	8434,16		1
5	Jambi	1151236,18	36366,46	36366,46		1
6	Sumatera Selatan	1124019,18	9164,64	9164,64		1
7	Bengkulu	1155489,33	40619,86	40619,86		1
8	Lampung	1137321,05	22468,65	22468,65		1
9	Kep. Bangka Belitung	1167540,24	52670,84	52670,84		1
10	Kep. Riau	1123782,71	8974,86	8974,86		1
11	Dki Jakarta	10,00	1114878,40	10,00	1	
12	Jawa Barat	901327,86	213591,14	213591,14		1
13	Jawa Tengah	1058402,73	56527,90	56527,90		1
14	Di Yogyakarta	1153812,53	38943,52	38943,52		1
15	Jawa Timur	783728,73	331167,83	331167,83		1
16	Banten	1009992,81	104949,93	104949,93		1
17	Bali	1138028,15	23154,81	23154,81		1
18	Nusa Tenggara Barat	1154610,92	39743,08	39743,08		1
19	Nusa Tenggara Timur	1138563,16	23700,39	23700,39		1
20	Kalimantan Barat	1131836,82	16975,95	16975,95		1
21	Kalimantan Tengah	1142790,11	27922,04	27922,04		1
22	Kalimantan Selatan	1151974,73	37106,23	37106,23		1
23	Kalimantan Timur	1095005,88	20207,50	20207,50		1
24	Kalimantan Utara	1161098,76	46230,02	46230,02		1
25	Sulawesi Utara	1155735,22	40869,28	40869,28		1
26	Sulawesi Tengah	1144885,95	30032,40	30032,40		1
27	Sulawesi Selatan	1109691,44	5214,17	5214,17		1
28	Sulawesi Tenggara	1153062,60	38193,05	38193,05		1
29	Gorontalo	1165162,73	50290,89	50290,89		1
30	Sulawesi Barat	1166366,56	51497,06	51497,06		1
31	Maluku	1161414,79	46544,15	46544,15		1
32	Maluku Utara	1163674,71	48805,11	48805,11		1
33	Papua Barat	1146798,12	31924,98	31924,98		1
34	Papua	1076974,56	37918,30	37918,30		1

Perhitungan manual pada data diatas didapatkan hasil akhir yang dimana pada iterasi 1 dan iterasi 2 pengelompokan data yang dilakukan terhadap 2 *cluster* didapatkan hasil yang sama. Hasil dari kedua iterasi tersebut benilai C1 =1 dan C2 =33 pada posisi data tiap *cluster*. Sehingga posisi *cluster* pada data tersebut tidak mengalami perubahan lagi maka proses iterasi berhenti sampai iterasi 2.



Gambar 2. Tampilan Data Perhitungan *RapidMiner*

Berdasarkan pada gambar 2 diatas mengenai tahap-tahap penggunaan serta hasil yang telah ditampilkan maka akan dibahas mengenai keterkaitan dari hasil yang didapat dari *RapidMiner* dan *Microsoft Excel* menjelaskan bahwa hasil dari perhitungan manual algoritma *K-Means dan Microsoft Excel* data memiliki nilai yang sama yaitu antara beberapa *cluster* yaitu *cluster* tinggi 1 dan rendah 33, serta memasukan perhitungan *Microsoft Excel* ke dalam *RapidMiner* memiliki nilai yang sama juga.



Gambar 3. Hasil Pengelompokan *RapidMiner*

Berdasarkan pada gambar 3. dapat diketahui bahwa pada kelompok rendah memiliki *node* yaitu 33 dengan warna biru, sedangkan pada kelompok tinggi memiliki 1 *node* dengan warna merah.

4. Kesimpulan

Berdasarkan pembahasan dan hasil *implementasi software Rapidminer* pada pengelompokan banyaknya pekerja tetap perusahaan konstruksi menurut provinsi, maka dapat kesimpulan sebagai berikut :

a) Data yang diolah untuk memperolah nilai dari pekerja tetap perusahaan konstruksi menerapkan metode *clustering K-Means.* Data tersebut diolah menggunakan *Microsoft Excel* untuk ditentukan nilai *centroid* dalam 2 *cluster* yaitu *cluster* tinggi dan rendah.

b) Hasil yang diperoleh dari metode *k-means clustering* yang di *implementasikan* ke dalam *rapidminer* memiliki nilai yang sama yaitu menghasilkan 2 *cluster* yaitu *cluster* tinggi sebanyak 1 yaitu DKI Jakarta. Sedangkan untuk *cluster* rendah sebanyak 33 yaitu Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep.Bangka Belitung, Kep.Riau, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua. Hasil yang didapat dari penelitian diharapkan dapat menjadi masukan Pemerintah Indonesia agar memberi perhatian lebih pada Provinsi yang memiliki pekerja tetap perusahaan konstruksi terendah berdasarkan *Cluster* yang telah dilakukan.

Daftar Pustaka

[1] A. P. Windarto, U. Indriani, M. R. Raharjo, and L. S. Dewi, "Bagian 1: Kombinasi Metode Klastering dan Klasifikasi (Kasus Pandemi Covid-19 di Indonesia)," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 855, 2020, doi: 10.30865/mib.v4i3.2312.

[2] M. G. Sadewo, A. P. Windarto, and A. Wanto, "Penerapan Algoritma Clustering Dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/ Mitigasi Bencana Alam Menurut Provinsi Dengan K-Means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 311–319, 2018, doi: 10.30865/komik.v2i1.943.

[3] A. Muzakir and R. A. Wulandari, "Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree," *Sci. J. Informatics*, vol. 3, no. 1, pp. 19–26, 2016, doi: 10.15294/sji.v3i1.4610.

[4] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.

[5] A. Maulana and A. A. Fajrin, "Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fp-Growth Pada Data Transaksi Penjualan Spare Part Motor," *Klik - Kumpul. J. Ilmu Komput.*, vol. 5, no. 1, p. 27, 2018, doi: 10.20527/klik.v5i1.100.

[6] W. Nengsih, "Descriptive Modelling Menggunakan K-Means Untuk Pengclusteran Tingkat Kemiskinan Di Propinsi Riau," no. January, 2017.

[7] R. W. Sari and D. Hartama, "Data Mining : Algoritma K-Means Pada Pengelompokkan Wisata Asing ke Indonesia Menurut Provinsi," *Semin. Nas. Sains Teknol. Inf.*, pp. 322–326, 2018.

[8] M. G. Sadewo, A. P. Windarto, and D. Hartama, "Penerapan Datamining Pada Populasi Daging Ayam Ras Pedaging Di Indonesia Berdasarkan Provinsi Menggunakan K-Means Clustering," *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 2, no. 1, pp. 60–67, 2017.

[9] M. . Imelda A.Muis & Muhammad Affandes, "Penerapan Metode Support Vector Machine (SVM) Menggunakan Kernel Radial Basis Function (RBF) Pada Klasifikasi Tweet," *Sains, Teknol. dan Ind. Sultan Syarif Kasim Riau*, vol. 12, no. 2, pp. 189–197, 2015.