

BAB III METODE PENELITIAN

3.1 Alat Penelitian

Dalam Penelitian selain menggunakan data sebagai bahan penelitian juga diperlukan komponen atau alat pendukung seperti *software* dan *hardware* sebagai bahan pendukung berjalannya penelitian tersebut, alat yang digunakan penelitian adalah

3.1.2 Perangkat Keras

Kebutuhan perangkat keras (*hardware*) yang digunakan dalam penelitian ini adalah: laptop dengan spesifikasi Processor Intel core I3, dan menggunakan Operasi windows 10, RAM 4GB dan HDD 500 GB

3.1.3 Perangkat Lunak

Kebutuhan perangkat Lunak (*software*) merupakan faktor yang sangat penting dalam penyusunan penelitian ini, perangkat lunak yang di pake adalah: sistem oprasi yang diipakai penelitian adalah sistem oprasi windows 10 adapun tambahan aplikasi lainnya adalah Sublime text, xampp, rapidminer dan Microsoft Excel

3.2 Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data Guru yang diperoleh dari website Dapodik yang berjumlah 2.746.420 pada tahun 2019. Data guru yang digunakan adalah Guru tingkat SMP.

3.2.1 Tahap Cleaning

Data yang didapatkan adalah ada beberapa data Guru diseluruh Indonesia yang didapat sebelum dilakukan perhitungan diantaranya data jumlah Guru semua tingkat pendidikannya diantaranya pada tabel 3.1 dan tabel 3.2

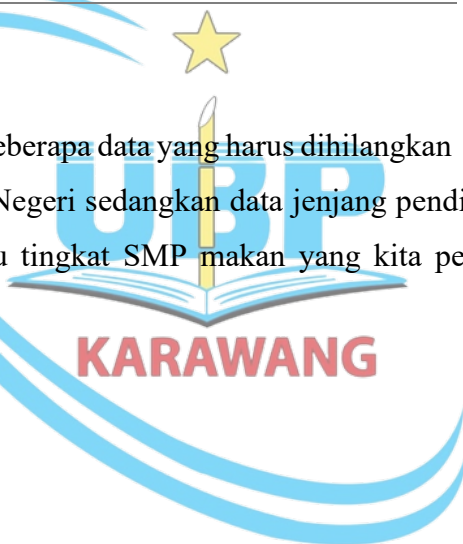
Tabel 3. 1 Tabel wilayah

WILAYAH
Prov. Jawa Barat
Prov. Jawa Timur
Prov. Jawa Tengah
Prov. Sumatera Utara
Prov. Sulawesi Selatan
Prov. Sumatera Selatan
Prov. Banten
Prov. Lampung
Prov. Nusa Tenggara Timur
Prov. Aceh
Prov. Riau
Prov. D.K.I. Jakarta
Prov. Sumatera Barat
Prov. Nusa Tenggara Barat
Prov. Kalimantan Barat
Prov. Bali
Prov. Jambi
Prov. Kalimantan Selatan
Prov. Sulawesi Tengah
Prov. Kalimantan Timur
Prov. Sulawesi Tenggara
Prov. D.I. Yogyakarta
Prov. Kalimantan Tengah
Prov. Sulawesi Utara
Prov. Maluku
Prov. Papua
Prov. Bengkulu
Prov. Kepulauan Riau
Prov. Sulawesi Barat
Prov. Maluku Utara
Prov. Gorontalo
Prov. Kepulauan Bangka Belitung
Prov. Papua Barat
Prov. Kalimantan Utara
Luar Negeri

Tabel 3. 2 Tabel jenjang pendidikan

	Jml	L	P
TK	0	0	0
KB	0	0	0
TPA	0	0	0
SPS	0	0	0
PKBM	0	0	0
SKB	0	0	0
SD	1.467.461	455.033	1.012.428
SMP	643.266	247.848	395.418
SMA	314.619	127.824	186.795
SMK	304.634	144.529	160.105
SLB	25.04	6.752	18.288

Setelah data semua didapatkan ada beberapa data yang harus dihilangkan diantaranya data wilayah yang dihilangkan adalah data Luar Negeri sedangkan data jenjang pendidikan yang dihilangkan karena yang di butuhkan data Guru tingkat SMP maka yang kita perlukan hanya data SMP selebihnya itu dihilangkan



Gambar 3. 1 Tampilan Dataset

3.3 Lokasi Penelitian dan Waktu Penelitian

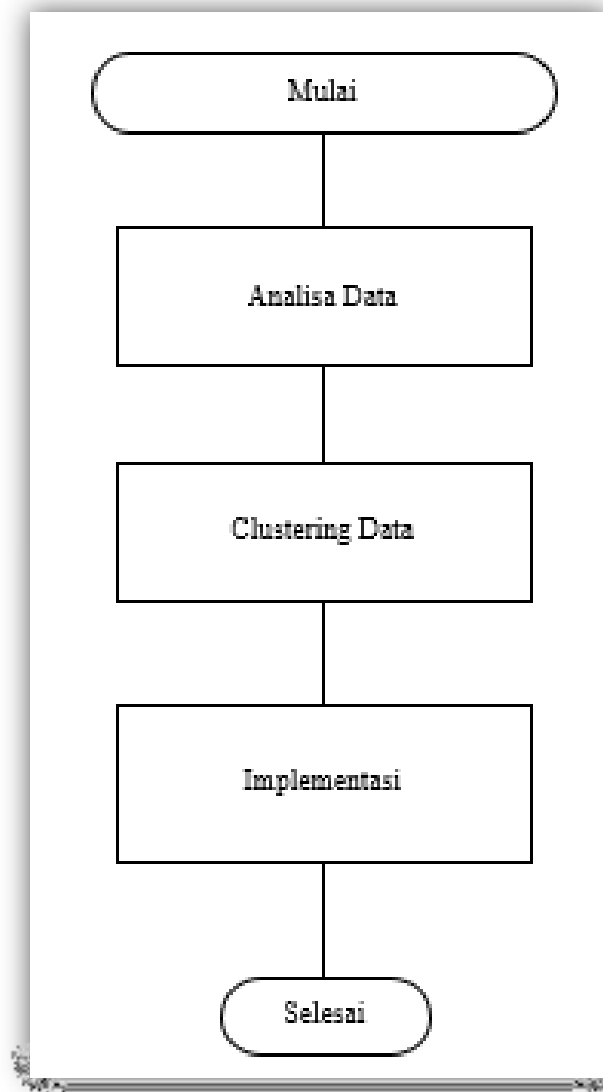
Lokasi Penelitian dilakukan di laboratorium riset kampus Universitas Buana Perjuangan Karawang, waktu penelitian pada bulan April sampai dengan Juli 2020

Tabel 3. 3 Tabel Jadwal Penelitian

Kegiatan	Bulan				
	1	2	3	4	5
Pengumpulan Data					
Analisis Data					
Pengajuan Proposal					
Pengolahan Data					
Pengujian Data					
Tahap Evaluasi					

3.4 Prosedur percobaan

Prosedur percobaan terdiri dari beberapa tahapan, sebelum melakukan penelitian ada beberapa tahapan untuk membuat rancangan penelitian. Prosedur percobaan di gambar sebagai berikut



Gambar 3. 2 Prosedur Percobaan

Penjelasan Flowchart

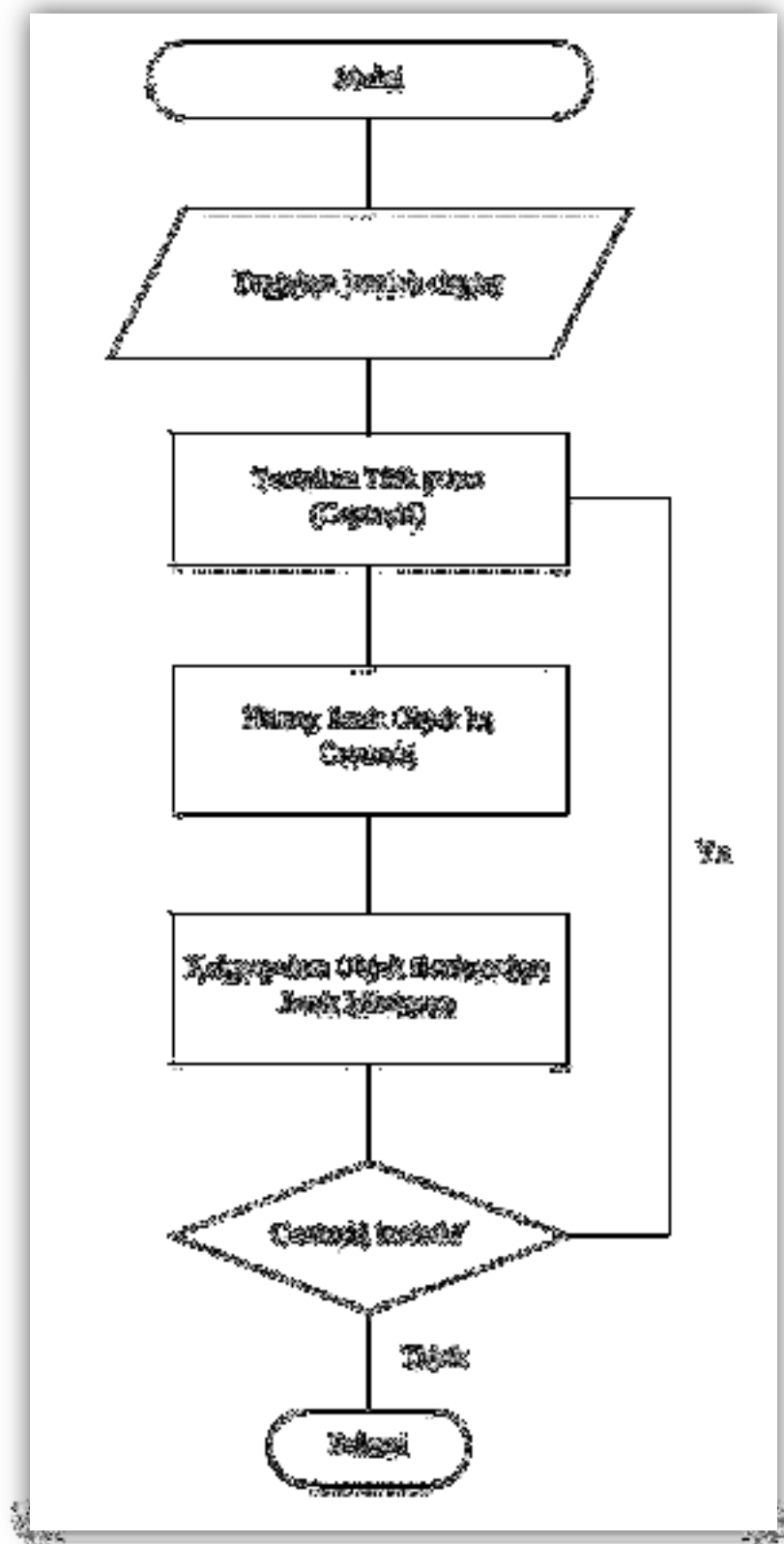
1. Menyiapkan data yang akan diolah
2. Data diambil dari data obyek untuk dilakukan analisa
3. Clustering ialah mencari suatu nilai atau data yang sudah masuk kedalam suatu cluster atau kelompok data
4. Implementasi ialah Menghitung setiap objek ke cluster terdekat dengan centroid awal menggunakan rumus Euclidean distance
5. Proses tahapan data mining telah selesai

3.4.1 Analisa Data

Dalam penulisan tugas akhir ini dicari nilai yang mempunyai data terendah dan tertinggi dari data Guru tingkat SMP hanya atribut jumlah, laki-laki dan perempuan saja yang dibutuhkan. Data yang didapat digabungkan untuk mencari nilai rata-rata dari jumlah data Guru tingkat SMP dan selanjutnya dilakukan penentuan centroid awal atau dapat diinputkan secara manual oleh user kemudian dilakukan perhitungan jarak antara centroid awal pada setiap cluster. Setelah dilakukan proses perhitungan maka akan dapat centroid baru, lakukan proses iterasi sampai titik centroid benar-benar sama dengan centroid sebelumnya maka proses telah selesai

3.4.2 K-Means Clustering

Setelah mendapatkan data yang kita cari dari website dapodik kemudian dilakukan pemisahan data atau *pre-processing* dan sudah mendapatkan datasetnya kemudian dilakukan pengolahan data menggunakan Algoritma K-Means Clustering. Dalam penelitian ada beberapa tahapan clustering yang akan dilakukan beberapa tahapan dapat dilihat seperti pada Gambar 3.4.



Gambar 3. 3 Tahapan Clustering

1. Menentukan Jumlah Cluster

Menentukan banyak cluster sesuai dengan masalah yang diteliti untuk mempermudah mengelompokkan data yang memiliki karakter sejenis.

2. Menentukan Titik Pusat Cluster

Titik pusat cluster atau bias disebut centroid digunakan sebagai nilai pengurang untuk perhitungan jarak antara data ke setiap cluster atau disebut *distance*. Dalam proses ini penentuan titik pusat cluster bisa ditentukan sesuai keinginan peneliti dengan syarat nilai centroid masih termasuk kedalam nilai data pada setiap atribut selain itu penentuan pusat cluster bias ditentukan dengan menggunakan nilai rata-rata disetiap atribut.

3. Menghitung Jarak Data Ke Setiap Cluster

Setelah menentukan titik pusat selanjutnya menghitung jarak data ke setiap cluster yang sudah dibentuk untuk mencari jarak dari satu cluster

4. Mengelompokkan Data Kedalam Cluster

Setelah mendapatkan jarak antara data ke setiap cluster yang sudah dibentuk proses selanjutnya adalah clustering atau pengelompokan data kedalam cluster. Untuk tahapan ini cukup melihat dari nilai yang mempunyai jarak terdekat dari setiap cluster

5. Menentukan Titik Pusat Cluster Baru

Setelah proses pengelompokan data sudah selesai maka tahapan selanjutnya adalah menentukan cluster baru. Penentuan centroid baru ini menggunakan metode yang hampir sama dengan menentukan cluster pada tahapan sebelumnya, yang membedakan dari kedua tahapan ini adalah dari jumlah data yang digunakan. Dalam tahapan ini centroid dapat ditentukan dengan menghitung nilai rata-rata sesuai dengan data yang sudah dikelompokkan kedalam cluster masing-masing. Dalam proses ini peneliti mencari nilai rata-rata berdasarkan yang sudah dikelompokkan untuk mendapatkan hasil.

6. Menverifikasi Titik Pusat Cluster

Proses tahapan ini adalah memverifikasi pusat Cluster. Proses tersebut adalah proses untuk memverifikasi antara data cluster baru dengan cluster yang lama, apabila kedua pusat cluster mempunyai nilai yang centroidnya berbeda maka proses clustering masih berlanjut ke tahap proses selanjutnya yang akan dimulai proses ke-2 yaitu (menghitung jarak data ke setiap cluster) dengan menggunakan nilai cluster yang baru, sedangkan jika nilai kedua

centroid tersebut bernilai sama maka proses clustering berhenti. Proses ini dilakukan bertujuan untuk mengetahui apakah sudah selesai atau masih perlu dilakukan prulangan proses kembali.

3.5 Implementasi K-Means Menggunakan *Microsoft Excel*

Setelah melewati tahapan – tahapan proses K-Means Clustering maka proses berikutnya adalah perhitungan secara manual menggunakan aplikasi *Microsoft Excel*. Untuk mendapatkan hasil yang diolah menggunakan algortima K-Means Clustering peneliti melakukan perulangan tahapan beberapa kali proses. Proses secara manual adalah sebagai berikut:

Iterasi ke 1

Tabel 3. 4 Titik Pusat

Centorid	Provinsi	Jumlah	Laki-laki	Perempuan
c1	Prov. Jawa Barat	86.943	37.729	49.214
c2	Prov. Kepulauan Riau	5.14	1.699	3.441

Pada Tabel 3.4 adalah penentuan titik pusat yang diambil dari dataset data Guru tingkat SMP yang diambil secara acak untuk dilakukan proses pengelompokan data pada tahapa berikutnya.

Tabel 3. 5 Mencari Nilai Centroid 1

	c1	c2	Jarak terdekat ke c
1	0		
2	10.374478		
3	22.013954		
4	55.586068		
5	76.724039		
.			
.			
.			
31	102.32251		
32	102.86483		
33	103.85512		
34	102.15114		

Pada Tabel 3.5 ini adalah dimana proses mencari jarak nilai c1 atau centroid 1 pada tahap awal dengan menggunakan rumus yang sudah ada pada aplikasi *Microsoft Excel*

Tabel 3. 6 Mencari Nilai Centroid 2

	c1	c2	jarak terdekat ke c
1	0	100.4243956	
2	10.37447801	90.09467355	
3	22.0139538	78.41744484	
4	55.58606761	45.63374153	
5	76.72403885	23.70517665	
.			
.			
.			
31	102.322509	1.962309354	
32	102.8648281	2.496368963	
33	103.8551193	3.520661586	
34	102.1511385	1.728552574	

Pada Tabel 3.6 adalah sama dengan proses sebelumnya dimana proses ini mencari jarak nilai C2 atau centroid 2 menggunakan rumus pada aplikasi *Microsoft Excel*.

Tabel 3. 7 Mengelompokkan Kedalam Cluster

Jarak			
	c1	c2	terdekat ke c
1	0	100.4243956	0
2	10.37447801	90.09467355	10.37447801
3	22.0139538	78.41744484	22.0139538
4	55.58606761	45.63374153	45.63374153
5	76.72403885	23.70517665	23.70517665
.			
.			
.			
31	102.322509	1.962309354	1.962309354
32	102.8648281	2.496368963	2.496368963
33	103.8551193	3.520661586	3.520661586
34	102.1511385	1.728552574	1.728552574

Pada proses ini dimana centroid 1 dan centroid 2 sudah didapatkan maka tahapan selanjutnya ialah mengelompokkan ke dalam jumlah cluster seperti pada tabel 3.7

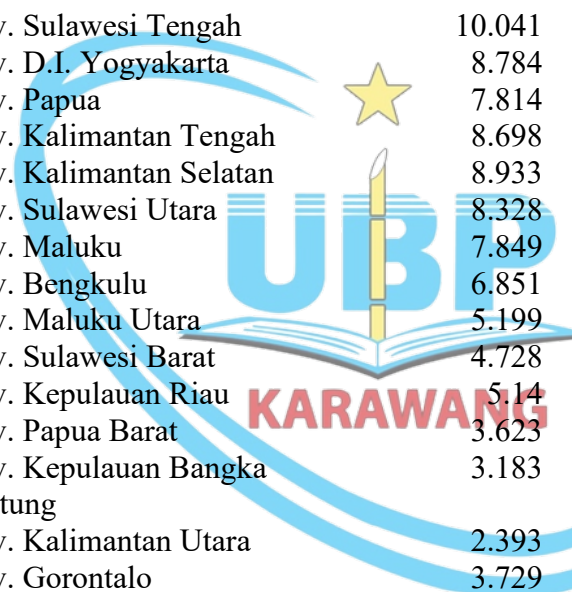
Tabel 3. 8 Mencari Nilai Rata-rata

Cluster 1	Prov. Jawa Barat	86.943	37.729	49.214
	Prov. Jawa Timur	78.576	34.69	43.886
	Prov. Jawa Tengah	68.988	29.473	39.515
Centroid1		58.62675	25.473	33.15375

Pada tahapan ini setelah mencari beberapa nilai centroid data pada setiap objek maka peneliti mengelompokkan kedalam cluster yang memiliki nilai yang termasuk kedalam cluster 1 setelah nilai didapatkan kemudian dilakukan perhitungan mencari nilai rata-rata dari setiap data yang ada seperti pada Tabel 3.8

Tabel 3. 9 Mencari Nilai Rata-rata

Cluster2	Prov. Sumatera Utara	41.603	13.293	28.31
----------	----------------------	--------	--------	-------



Prov. Nusa Tenggara Timur	24.47	10.509	13.961
Prov. Sulawesi Selatan	28.481	9.63	18.851
Prov. Banten	20.67	9.063	11.607
Prov. Lampung	21.787	7.999	13.788
Prov. Nusa Tenggara Barat	17.363	7.868	9.495
Prov. D.K.I. Jakarta	19.114	7.544	11.57
Prov. Sumatera Selatan	24.524	7.369	17.155
Prov. Kalimantan Barat	14.422	5.935	8.487
Prov. Bali	11.754	5.51	6.244
Prov. Aceh	21.819	5.437	16.382
Prov. Riau	18.6	5.222	13.378
Prov. Sumatera Barat	17.668	4.268	13.4
Prov. Sulawesi Tenggara	10.74	4.16	6.58
Prov. Kalimantan Timur	9.763	3.894	5.869
Prov. Jambi	10.365	3.691	6.674
Prov. Sulawesi Tengah	10.041	3.689	6.352
Prov. D.I. Yogyakarta	8.784	3.584	5.2
Prov. Papua	7.814	3.385	4.429
Prov. Kalimantan Tengah	8.698	3.168	5.53
Prov. Kalimantan Selatan	8.933	3.154	5.779
Prov. Sulawesi Utara	8.328	2.469	5.859
Prov. Maluku	7.849	2.338	5.511
Prov. Bengkulu	6.851	2.285	4.566
Prov. Maluku Utara	5.199	1.998	3.201
Prov. Sulawesi Barat	4.728	1.94	2.788
Prov. Kepulauan Riau	5.14	1.699	3.441
Prov. Papua Barat	3.623	1.387	2.236
Prov. Kepulauan Bangka Belitung	3.183	1.214	1.969
Prov. Kalimantan Utara	2.393	1.059	1.334
Prov. Gorontalo	3.729	1.021	2.708
Centroid1	13.17535	4.702645	8.47271

Pada tahapan ini setelah mencari beberapa nilai centroid data pada setiap objek maka peneliti mengelompokkan kedalam cluster yang memiliki nilai yang termasuk kedalam cluster 2 setelah nilai didapatkan kemudian dilakukan perhitungan mencari nilai rata-rata dari setiap data yang ada seperti pada tabel 3.9. Setelah mendapatkan nilai centroid baru kemudian dilakukan perhitungan seperti pada proses awal sampai nilai centroid benar-benar tidak berubah maka proses perhitungan atau iterasi telah selesai dilakukan.

Iterasi ke 2

Tabel 3. 10 Pusat Cluster Baru Pada Iterasi Ke 2

C1	58.62675	25.473	33.15375
C2	13.17535	4.702645	8.47271

Proses ini sama seperti pada tahap iterasi ke 1 yang membedakan proses ini adalah centroid yang diambil ialah hasil dari perhitungan rata-rata dari proses iterasi ke 1, proses ini dapat dilihat pada Tabel 3.10.

Tabel 3. 11 Mencari Nilai C1 Pada Iterasi Ke 2

	C1	C2	terdekat ke c
1	34.78435252		
2	24.45622323		
3	12.79925791		
4	21.48539918		
5	41.9400348		
.			
.			
.			
31	67.53854585		
32	68.08081321		
33	69.07079358		
34	67.36918399		

Proses ini sama pada proses sebelumnya pada iterasi ke 1 yang membedakan ialah pencaarian nilai centroid 1 berbeda datanya, dikarenakan data centroid yang diambil dari proses iterasi ke 1 berbeda pada proses iterasi ke 2, proses ini dapat dilihat pada Tabel 3.11.

Tabel 3. 12 Mencari Nilai C2 Pada Iterasi Ke2

	C1	C2	terdekat ke c
1	34.78435252	90.51109889	
2	24.45622323	80.19094456	
3	12.79925791	68.49997194	
4	21.48539918	35.71335266	
5	41.9400348	13.83489082	
.			
.			
.			
31	67.53854585	11.88012811	
32	68.08081321	12.42239696	
33	69.07079358	13.4348949	
34	67.36918399	11.6627578	

Proses ini sama pada proses sebelumnya mencari nilai C2 pada iterasi ke 1 yang membedakan ialah pencaarian nilai centroid 2 berbeda datanya, dikarenakan data pusat cluster centroid yang diambil dari proses iterasi ke 1 berbeda pada proses iterasi ke 2, proses ini dapat dilihat pada Tabel 3.12

Tabel 3. 13 Mengelompokkan Kedalam Cluster Pada Iterasi Ke2

	C1	C2	terdekat ke c
1	34.78435252	90.51109889	34.78435252
2	24.45622323	80.19094456	24.45622323
3	12.79925791	68.49997194	12.79925791
4	21.48539918	35.71335266	21.48539918
5	41.9400348	13.83489082	13.83489082
.			
.			
.			
31	67.53854585	11.88012811	11.88012811
32	68.08081321	12.42239696	12.42239696

33	69.07079358	13.4348949	13.4348949
34	67.36918399	11.6627578	11.6627578

Proses ini sama pada proses sebelumnya mengelompokkan data pada iterasi ke 1 yang membedakan ialah data centroid 1 dan centroid 2 berbeda datanya, dikarenakan data pusat cluster centroid yang diambil dari proses iterasi ke 1 berbeda pada proses iterasi ke 2, proses ini dapat dilihat pada Tabel 3.13

Tabel 3. 14 Mencari Nilai Rata-rata Pada Iterasi Ke2

Cluster1	Prov. Jawa Barat	86.943	37.729	49.214
	Prov. Jawa Timur	78.576	34.69	43.886
	Prov. Jawa Tengah	68.988	29.473	39.515
	Prov. Sumatera Utara	41.603	13.293	28.31
Centroid2		69.0275	28.79625	40.23125

Pada proses ini hasil perhitungan mencari nilai rata-rata cluster 1 pada iterasi ke 2 proses ini mencari perasamaan nilai centroid dari proses iterasi ke 1 dapat dilihat pada Tabel 3.14

Tabel 3. 15 Mencari Nilai Rata-rata Pada Iterasi Ke2

Cluster2	Prov. Nusa Tenggara Timur	24.47	10.509	13.961
	Prov. Sulawesi Selatan	28.481	9.63	18.851
	Prov. Banten	20.67	9.063	11.607
	Prov. Lampung	21.787	7.999	13.788
	Prov. Nusa Tenggara Barat	17.363	7.868	9.495
	Prov. D.K.I. Jakarta	19.114	7.544	11.57
	.			
	.			
	.			
	Prov. Kalimantan Utara	2.393	1.059	1.334
	Prov. Gorontalo	3.729	1.021	2.708
Centroid2		12.22776667	4.4163	7.811467

Pada proses ini hasil perhitungan mencari nilai rata-rata cluster 2 pada iterasi ke 2 proses ini mencari perasamaan nilai centroid dari proses iterasi ke 1 dapat dilihat pada Tabel 3.15

Iterasi ke 3

Tabel 3. 16 Pusat Cluster Baru Pada Iterasi Ke 3

c1	69.027	28.79625	40.23125
c2	12.22776667	4.4163	7.811467

Proses ini sama seperti pada tahap iterasi ke 2 yang membedakan proses ini adalah centroid yang diambil ialah hasil dari perhitungan rata-rata dari proses iterasi ke 2, proses ini dapat dilihat pada tabel 3.16

Tabel 3. 17 Mencari Nilai C1 Pada Iterasi Ke 3

	c1	c2	terdekat ke c
1	21.94235348		
2	11.80156293		
3	0.986167138		
4	33.68297401		
5	54.86233495		
.			
.			
.			
31	80.45240347		
32	80.99513952		
33	81.99049366		
34	80.27009161		

Proses ini sama pada proses sebelumnya pada iterasi ke 2 yang membedakan ialah pencaarian nilai centroid 1 berbeda datanya, dikarenakan data centroid yang diambil dari proses iterasi ke 2 berbeda pada proses iterasi ke 3, proses ini dapat dilihat pada Tabel 3.17

Tabel 3. 18 Mencari Nilai C2 Pada Iterasi Ke 3

	c1	c2	terdekat ke c
1	21.94235348	91.68572305	
2	11.80156293	81.36311762	
3	0.986167138	69.67550723	
4	33.68297401	36.90379372	
5	54.86233495	14.99366621	
.			
.			
.			
31	80.45240347	10.69132827	
32	80.99513952	11.23374159	
33	81.99049366	12.2454758	
34	80.27009161	10.47866749	

Proses ini sama pada proses sebelumnya mencari nilai C2 pada iterasi ke 2 yang membedakan ialah pencaarian niali centroid 2 berbeda datanya, dikarenakan data pusat cluster centroid yang diambil dari proses iterasi ke 2 berbeda pada proses iterasi ke 3, proses ini dapat dilihat pada tabel 3.18

Tabel 3. 19 Mengelompokkan Kedalam Cluster Pada Iterasi Ke 3

	c1	c2	terdekat ke c
1	21.94235348	91.68572305	21.94235348
2	11.80156293	81.36311762	11.80156293
3	0.986167138	69.67550723	0.986167138
4	33.68297401	36.90379372	33.68297401
5	54.86233495	14.99366621	14.99366621
.			
.			
.			
31	80.45240347	10.69132827	10.69132827
32	80.99513952	11.23374159	11.23374159
33	81.99049366	12.2454758	12.2454758
34	80.27009161	10.47866749	10.47866749

Proses ini sama pada proses sebelumnya mengelompokan data pada iterasi ke 2 yang membedakan ialah data centroid 1 dan centroid 2 berbeda datanya, dikarenakan data pusat cluster centroid yang diambil dari proses iterasi ke 2 berbeda pada proses iterasi ke 3, proses ini dapat dilihat pada tabel 3.19

Tabel 3. 20 Mencari Nilai Rata-rata Pada Iterasi Ke 3

Cluster1	Prov. Jawa Barat	86.943	37.729	49.214
	Prov. Jawa Timur	78.576	34.69	43.886
	Prov. Jawa Tengah	68.988	29.473	39.515
	Prov. Sumatera Utara	41.603	13.293	28.31
centroid		69.0275	28.79625	40.23125

Pada proses ini hasil perhitungan mencari nilai rata-rata cluster 1 pada iterasi ke 3 proses ini mencari perasamaan nilai centroid dari proses iterasi ke 2 dapat dilihat pada tabel 3.20

Tabel 3. 21 Mencari Nilai Rata-rata Pada Iterasi Ke 3

cluster2	Prov. Nusa Tenggara Timur	24.47	10.509	13.961
	Prov. Sulawesi Selatan	28.481	9.63	18.851
	Prov. Banten	20.67	9.063	11.607
	Prov. Lampung	21.787	7.999	13.788
	Prov. Nusa Tenggara Barat	17.363	7.868	9.495
	.			
	.			
	.			
	Prov. Kepulauan Bangka Belitung	3.183	1.214	1.969
	Prov. Kalimantan Utara	2.393	1.059	1.334
centroid	Prov. Gorontalo	3.729	1.021	2.708

Pada Proses perhitungan manual pada iterasi ke 3 nilai centroid adalah 69.0275 dan 12.22776667 dimana proses ini telah selesai nilai dari rata-rata cluster yang didapatkan dari iterasi ke 3 sama dengan nilai centroid pada iterasi ke 2 maka proses perhitungan yang dilakukan telah selesai proses ini dapat dilihat pada tabel 3.21