

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1. Pre-Processing Data.

Pre-Processing Data adalah tahapan dari Data Mining yaitu suatu proses atau tahapan yang dilakukan untuk mengolah data mentah menjadi data yang berkualitas yang baik dilanjutkan ke proses selanjutnya. Tahapan *pre-processing data* sebagai berikut:

4.1.1. Data Reduction

Data Reduction adalah proses untuk mengurangi atau mereduksi sejumlah data yang tidak dibutuhkan. Data Reduction sangat berguna untuk mendapatkan atribut dan data yang akan digunakan dalam penelitian ini. Atribut yang digunakan pada penelitian ini adalah kecamatan asal sekolah mahasiswa, asal sekolah mahasiswa, jenis kelamin mahasiswa dan nilai UN Sekolah periode 2016 sampai 2019.

A	B	C	D	E	F	G	H
Nama	kelamin	program	kecamatan	asal_sekolah	jurusan	NilaiSekolah	tahunid
ADE IRFAN	P	R.2	Pakis aji	SMK NEGERI 1 PAKIS AJI JEPARA	Teknik Komputer dan Jaringan	80	2017-5
MUHAMMAD JANGKI DAUSAT	P	R.2	Pakis aji	SMK NEGERI 1 PAKIS AJI JEPARA	Teknik Komputer dan Jaringan	80	2017-5
MUHAMMAD NAUFAL AFKAR RIFAL	P	R.1	Jepara	SMK BHAKTI PRAJA JEPARA	Mesin	75	2017-2
ACHMAD AZIS	P	R.2	Welahlan	PKBM WIDYATAMA WELAHAN	IPS	72	2017-5
MIFTAKHUS SALMAH	W	R.1	Kudus	MA NU MUJALLIMAT KUDUS	IPA	87	2017-2
ALFRISTHA SELLA NATALIA	W	R.1	Jepara	SMA MASEHI JEPARA	IPA	80	2017-2
AKHMAD DAIL FALAH	P	R.1	Kudus	SMK N 2 KUDUS	Teknik Komputer dan Jaringan	34.74	2017-2
NOVIAWATI	W	R.1	Mayong	SMK ISLAM AL HIKMAH MAYONG	Administrasi Perkantoran/ Sekretaris	75	2017-2
ERVANDHA STIEFANI	P	R.1	Jepara	SMK NEGERI 3 JEPARA	Teknik Komputer dan Jaringan	85	2017-2
DHIMAS ANTHONY FIRMANSYAH	P	R.1	Jepara	SMK NEGERI 3 JEPARA	Teknik Komputer dan Jaringan	80	2017-2
MUHAMMAD DANIAL FITRIAN	P	R.1	Jepara	SMK NEGERI 2 JEPARA	Animasi	85	2017-2
FARID AGUS PRIYONO	P	R.1	Jepara	SMK NEGERI 2 JEPARA	Animasi	8	2017-2
HAFIDH NIZAR ARFAKHSYAL	P	R.2	Kedung	SMK BHAKTI PRAJA JEPARA	Teknik Mekanik Otomotif	8.0	
NOR WAHID	P	R.2	Kedung	MA ITTHADUL MUSLIMIN KERSO	IPS	8.41	2017-5
FAIZ ALI SHOFIN	P	R.2	Nalumsari	SMK DATUK SINGARAJA KEDUNG	Multimedia	201.36.00	2017-5
AHMAD ANDRIAN	P	R.1	Tahunan	SMK ROUDLOTUL MUBTADIN NALUMSARI	Teknik Komputer dan Jaringan	252.8	2017-2
ADI HANIF FATHONI	P	R.2	Pakis aji	SMK ISLAM TSAVIROTUL HUDA TAHUNAN	Multimedia	80	2017-5
FAJAR LISTYA GALIH KUSUMA	P	R.1	Batealit	SMK NEGERI 1 PAKIS AJI JEPARA	Administrasi Perkantoran	81.8	2017-2
MUHAMMAD ADI MAFUT	P	R.2	Jepara	SMK NEGERI 1 BATEALIT	Administrasi Perkantoran	85	2017-5
FAHMI YUSUF	P	R.1	Kalinyamatan	SMK NEGERI 3 JEPARA	Teknik Komputer dan Jaringan	85	2017-2
DANY WIDODO ARIYANTO	P	R.2	Jepara	SMK NEGERI 1 KALINYAMATAN	Produksi Grafika	85	2017-5
GANDA YOGA MIFTAKHUDIN	P	R.1	Mlonggo	SMK NEGERI 3 JEPARA	Teknik Komputer dan Jaringan	80	2017-2
MUHAMMAD BAGAS BADRUT TAM	P	R.1	Mayong	MA MATHALIBUL HUDA MLONGGO	IPS	85	2017-2
FIRMAN SYAH	P	R.2	Bonang	SMK NEGERI 1 MAYONG	IPA	80	2017-5
MUHAMMAD LULI ABSOR	P	R.1	Batealit	SMK AL AMIN	Multimedia	80	2017-2
DEVI NUR CAHAYATI	W	R.2	Jepara	MA NEGERI 01 JEPARA	IPA	8.2	2017-5

Gambar 4. 1. Sampel Dataset Calon Mahasiswa Baru sebelum dilakukan *Pre-processing Data*.

Pada data calon mahasiswa baru periode 2016 sampai 2019, peneliti mereduksi beberapa atribut yang tidak diperlukan yaitu Nama Mahasiswa, Program, Jurusan sekolah dan tahun akademik. Dalam penelitian ini data yang tidak direduksi adalah asal kecamatan sekolah, asal sekolah, jenis kelamin mahasiswa dan nilai UN. Sehingga proses reduksi mendapatkan data calon mahasiswa baru teknik informatika periode 2016 sampai 2019 berjumlah 478 data calon mahasiswa baru.

	A	B	C	D
1	Asal Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah
2	Jepara	SMK	P	7,7
3	Demak	MA	P	7,5
4	Pakis aji	SMK	P	8,7
5	Demak	SMK	P	8,8
6	Pati	SMK	P	8,7
7	Keling	MA	W	4,1
8	Pakis aji	SMK	P	6,7
9	Tahunan	SMA	W	6,6
10	Jepara	SMK	W	7,8
11	Tahunan	SMA	P	8
12	Jepara	SMK	P	7,6
13	Kedung	MA	W	8,2
14	Batealit	SMK	P	8,2
15	Kedung	MA	P	7,8
16	Batealit	SMK	P	6,9
17	Jepara	SMK	P	8,7
18	Kembang	SMA	P	6,5
19	Mlonggo	SMK	P	7
20	Pecangaan	MA	P	8,1
21	Jepara	SMK	P	8,2
22	Jepara	SMK	P	8
23	Mlonggo	MA	P	5,6
24	Mlonggo	MA	P	7,5
25	Jepara	SMK	W	6,8
26	Jepara	PKBM	P	7,3
27	Pecangaan	SMA	P	7,1

Gambar 4. 2. Sampel *Dataset* Setelah Tahapan *Reduksi*

4.1.2. Data Cleaning

Data cleaning adalah proses untuk mengisi *missing value* atau yang kurang pas, mengoreksi data yang tidak konsisten, dan mendeteksi *redudansi* data yaitu penumpukan data atau duplikasi data. Penelitian ini melakukan proses data cleaning dengan menggunakan aplikasi *Microsoft Excel* dengan cara memfilter setiap kolom dan mencari data yang kosong atau *missing*.

4.1.3. Data Transformation

Data transformation adalah proses mengubah suatu data agar mendapatkan data yang lebih atau sesuai dengan kebutuhan. Dalam proses algoritma *K-means Clustering* dan algoritma *K-Medoids* ini data yang bisa diolah adalah data yang berupa numerik sehingga data yang berupa text seperti jenis kelamin, asal

kecamatan sekolah dan asal sekolah harus diubah menjadi data yang berupa numerik agar bisa dilanjutkan ke dalam proses selanjutnya.

Sehingga pada atribut asal kecamatan sekolah didapatkan 16 data dengan frekuensi dan inisial sebagai berikut:

Tabel 4.1. Inisialisasi Data Asal Kecamatan Sekolah

Asal Kecamatan Sekolah	Frekuensi	Inisial
Bangsri	32	0
Batealit	21	1
Donorojo	5	2
Jepara	173	3
Kalinyamatan	6	4
Kedung	23	5
Keling	8	6
Kembang	8	7
Mayong	18	8
Mlonggo	27	9
Nalumsari	14	10
Pakis aji	19	11
Pecangaan	28	12
Tahunan	36	13
Welahan	5	14
Luar Kecamatan Jepara	55	15

Pada atribut jenis kelamin dengan frekuensi dan inisial sebagai berikut:

Tabel 4. 2. Inisialisasi Data Jenis Kelamin

Jenis Kelamin	Frekuensi	Inisial
Pria	396	0
Wanita	82	1

Pada atribut asal sekolah peneliti mendapatkan 4 data asal sekolah dengan frekuensi dan inisial sebagai berikut:

Tabel 4. 3. Inisialisasi Data Asal Sekolah

Asal Sekolah	Frekuensi	Inisial
SMA	119	0
SMK	251	1
MA	101	2
PKBM	7	3

4.1.4. Data Intregation

Data Integration adalah suatu proses untuk menggabungkan data dari beberapa file sumber. Data integration ini dilakukan apabila data yang diolah berasal dari beberapa file sumber.

4.2. Penerapan Algoritma Menggunakan Aplikasi *Microsoft Excel*.

4.2.1. Penerapan Algoritma *K-Means* Menggunakan Aplikasi *Microsoft Excel*.

Penerapan Algoritma *K-Means Clustering*, penelitian ini menggunakan aplikasi *Microsoft Excel* 2016. Adapun tahapan yang diperlukan adalah sebagai berikut :

1. Menentukan Jumlah *Cluster*.

Menentukan jumlah *cluster* yang digunakan pada data calon mahasiswa sebanyak 3 *cluster* yaitu diantaranya wilayah tinggi, wilayah sedang dan wilayah rendah peminatnya berdasarkan data calon mahasiswa baru periode 2016 sampai periode 2019.

2. Menentukan Nilai Titik Pusat Awal *Cluster*.

Titik pusat *cluster* atau disebut dengan *centroid* digunakan sebagai nilai pengurang untuk penghitungan jarak antara data ke setiap *cluster* atau disebut *distance*. Dalam proses ini penentuan titik pusat awal bias ditentukan secara acak sesuai keinginan peneliti dengan syarat nilai centroid masih dalam range nilai data dari setiap atribut. Selain itu penentuan titik pusat awal juga bisa menggunakan nilai rata-rata dari setiap atribut.

Pada proses ini peneliti menentukan titik pusat awal *cluster* menggunakan nilai rata-rata dari setiap atribut dataset yaitu Asal Kecamatan Sekolah, Asal Sekolah,

Jenis Kelamin, dan Nilai Ujian Sekolah. Nilai rata-rata pada setiap atribut tersebut digunakan untuk nilai *centroid* awal *cluster* 1.

Untuk mengisi nilai *centroid* awal pada *cluster* 0 dan *cluster* 2, peneliti menentukan berdasarkan nilai *centroid* pada *cluster* 1 yaitu nilai *cluster* 0 kurang dari nilai *centroid* pada *cluster* 1 dan nilai *centroid* pada *cluster* 2 lebih besar dari nilai *centroid* pada *cluster* 1. Adapun titik pusat awal pada setiap atribut ditunjukkan pada tabel 4.4 dibawah ini.

Tabel 4. 4. Titik Pusat Awal dari setiap *Cluster*

TITIK PUSAT	Asal Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah
C0	11	1	0	6,7
C1	12	2	1	7,7
C2	1	2	0	8,7

3. Menghitung Jarak Data ke setiap *Cluster*.

Setelah titik pusat awal *cluster* telah ditentukan, proses selanjutnya adalah menghitung jarak setiap *cluster* atau *distance*. Dalam menghitung nilai jarak setiap *cluster*, peneliti menggunakan rumus *Euclidean Distance*. Adapun rumus tersebut adalah sebagai berikut.

$$Distance = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + (x_{3i} - x_{3j})^2 + (x_{4i} - x_{4j})^2}$$

Keterangan :

- x_{1i} adalah Data ke i pada atribut data ke 1.
- x_{1j} adalah centroid ke j pada atribut data ke 1.

	CLUSTER 0	CLUSTER 1	CLUSTER 2	Titik Pusat Awal	Asal Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah
2	8,062257748	9,110433579	2,449489743	Cluster 0	11	1	0	6,7
3	4,2	3,168595904	14,05133446	Cluster 1	12	2	1	7,7
4	2	2	10,04987562	Cluster 2	1	2	0	8,7
5	4,517742799	3,494281042	14,03602508					
6	4,472135955	3,464101615	14,03566885					
7	5,810335619	6,997142274	6,867313885					
8	0	2	10,24695077					
9	2,451530134	2,491987159	12,38587906					
10	8,136952747	9,058937279	2,60959767					
11	2,586503431	2,467792536	12,1856473					
12	8,050465825	9,110982384	2,491987159					
13	6,34428877	7,017834424	4,153311931					
14	10,11187421	11,10180166	1,118033989					
15	6,181423784	7,071774883	4,1					
16	10,0019998	11,1193525	2,059126028					
17	8,246211251	9,16515139	2,236067977					
18	4,127953488	5,607138308	6,696267617					
19	2,022374842	3,389690251	8,239538822					
20	1,989974874	1,077032961	11,01635148					
21	8,139410298	9,124143795	2,291287847					
22	8,104936767	9,115371633	2,343074903					
23	2,491987159	3,796050579	8,579627032					
24	2,374868417	3,168595904	8,089499366					
25	8,062877898	9,1	3,1					
26	8,268010643	9,119210492	2,638181192					

Gambar 4. 3. Sampel Menghitung Jarak Data ke setiap Cluster pada *Microsoft Excel*.

4. Mengelompokkan Data ke dalam Cluster

Setelah proses perhitungan jarak data ke setiap *cluster* berikutnya adalah proses mengelompokkan data ke setiap *cluster* yang terbentuk. Pengelompokkan data ini bertujuan untuk menentukan nilai pusat *cluster* baru pada proses selanjutnya. Pada proses ini peneliti mengelompokkan data menggunakan rumus seperti dibawah ini.

$$d(1,0)=. \sqrt{(3 - 11)^2 + (1 - 1)^2 + (0 - 0)^2 + (7,7 - 6,7)^2} = 8,66226$$

$$d(1,1)=. \sqrt{(3 - 12)^2 + (1 - 2)^2 + (0 - 1)^2 + (7,7 - 7,7)^2} = 9,11043$$

$$d(1,2)=. \sqrt{(3 - 1)^2 + (1 - 2)^2 + (0 - 0)^2 + (7,7 - 8,7)^2} = 2,4495$$

$$d(2,0)=. \sqrt{(15 - 11)^2 + (2 - 1)^2 + (0 - 0)^2 + (7,5 - 6,7)^2} = 4,2$$

$$d(2,1)=. \sqrt{(15 - 12)^2 + (2 - 2)^2 + (0 - 1)^2 + (7,5 - 7,7)^2} = 3,1686$$

$$d(2,2)=. \sqrt{(15 - 1)^2 + (2 - 2)^2 + (0 - 0)^2 + (7,5 - 8,7)^2} = 14,051$$

Dari contoh perhitungan diatas dapat diketahui bahwa pada data pertama dan data kedua dari setiap atribut di kelompokkan ke *cluster 2* dan *cluster 1*.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Asal Kecamatan	Inisial Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah	CLUSTER 0	CLUSTER 1	CLUSTER 2	Jarak Terdekat	C0	C1	C2
2	Jepara	3	1	0	7,7	8,062257748	9,110433579	2,449489743	2,449489743			OK
3	Demak	15	2	0	7,5	4,2	3,168595904	14,05133446	3,168595904		OK	
4	Pakis aji	11	1	0	8,7	2	2	10,04987562	2			
5	Demak	15	1	0	8,8	4,517742799	3,494281042	14,03602508	3,494281042		OK	
6	Pati	15	1	0	8,7	4,472135955	3,464101615	14,03566885	3,464101615		OK	
7	Keling	6	2	1	4,1	5,810335619	6,997142274	6,867313885	5,810335619	OK		
8	Pakis aji	11	1	0	6,7	0	2	10,24695077	0	OK		
9	Tahunan	13	0	1	6,6	2,451530134	2,491987159	12,38587906	2,451530134	OK		
10	Jepara	3	1	1	7,8	8,136952747	9,055937279	2,60959767	2,60959767			OK
11	Tahunan	13	0	0	8	2,586503431	2,467792536	12,1856473	2,467792536		OK	
12	Jepara	3	1	0	7,6	8,050465825	9,110982384	2,491987159	2,491987159			OK
13	Kedung	5	2	1	8,2	6,34428877	7,017834424	4,153311931	4,153311931		OK	
14	Batealit	1	1	0	8,2	10,11187421	11,10180166	1,118033989	1,118033989			OK
15	Kedung	5	2	0	7,8	6,181423784	7,071774883	4,1	4,1			OK
16	Batealit	1	1	0	6,9	10,0019998	11,1193525	2,059126028	2,059126028			OK
17	Jepara	3	1	0	8,7	8,246211251	9,16515139	2,236067977	2,236067977			OK
18	Kembang	7	0	0	6,5	4,127953488	5,607138308	6,696267617	4,127953488	OK		
19	Mlonggo	9	1	0	7	2,022374842	3,389690251	8,239538822	2,022374842	OK		
20	Pecangaan	12	2	0	8,1	1,989974874	1,077032961	11,01635148	1,077032961		OK	
21	Jepara	3	1	0	8,2	8,139410298	9,124143795	2,291287847	2,291287847			OK
22	Jepara	3	1	0	8	8,104936767	9,115371633	2,343074903	2,343074903			OK
23	Mlonggo	9	2	0	5,6	2,491987159	3,796050579	8,579627032	2,491987159	OK		
24	Mlonggo	9	2	0	7,5	2,374868417	3,168595904	8,089499366	2,374868417	OK		
25	Jepara	3	1	1	6,8	8,062877898	9,1	3,1	3,1			OK
26	Jepara	3	3	0	7,3	8,268010643	9,119210492	2,638181192	2,638181192			OK

Gambar 4. 4. Sampel Pengelompokan Data ke dalam Cluster iterasi ke -1 pada Microsoft Excel.

Dari hasil perhitungan diatas di dapatkan pusat cluster baru dan pengelompokan data iterasi ke-1 seperti tabel di bawah ini :

Tabel 4. 5. Centroid Baru Iterasi ke-1

TITIK PUSAT	Asal Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah	Jumlah Data
CLUSTER 0	9,84	0,82	0,21	7,44	111
CLUSTER 1	13,72	1,08	0,24	7,90	93
CLUSTER 2	2,76	1,03	0,12	7,74	270

Proses iterasi selanjutnya melakukan perhitungan lagi menggunakan centroid baru dari hasil iterasi ke-1

	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW
1	Asal Kecamatan	Inisial Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah	C0	C1	C2	Jarak Terdekat	KC0	KC1	KC2
2	Jepara	3	1	0	7,7	6,85699633	10,72670504	0,273554061	0,273554061			OK
3	Demak	15	2	0	7,5	5,28895011	1,642582531	12,2785155	1,642582531		OK	
4	Pakis aji	11	1	0	8,7	1,72799486	2,848464056	8,293371872	1,72799486	OK		
5	Demak	15	1	0	8,8	5,336000672	1,58440696	12,28322322	1,58440696		OK	
6	Pati	15	1	0	8,7	5,311420851	1,529886139	12,27502804	1,529886139		OK	
7	Keling	6	2	1	4,1	5,288064286	8,687918798	5,045850731	5,045850731			OK
8	Pakis aji	11	1	0	6,7	1,397044023	2,986280494	8,303904613	1,397044023	OK		
9	Tahunan	13	0	1	6,6	3,459172755	1,98785878	10,38889446	1,98785878		OK	
10	Jepara	3	1	1	7,8	6,90403476	10,74882084	0,907876221	0,907876221			OK
11	Tahunan	13	0	0	8	3,314250221	1,327983919	10,29264833	1,327983919		OK	
12	Jepara	3	1	0	7,6	6,853954111	10,72904503	0,305896331	0,305896331			OK
13	Kedung	5	2	1	8,2	5,105633924	8,807086828	2,630240513	2,630240513			OK
14	Batealit	1	1	0	8,2	8,883374619	12,72783438	1,825645084	1,825645084			OK
15	Kedung	5	2	0	7,8	5,003505305	8,773850267	2,442330848	2,442330848			OK
16	Batealit	1	1	0	6,9	8,867473843	12,76363775	1,958726526	1,958726526			OK
17	Jepara	3	1	0	8,7	6,966743557	10,7545404	0,993692315	0,993692315			OK
18	Kembang	7	0	0	6,5	3,117810291	6,95531652	4,535997825	3,117810291	OK		
19	Mlonggo	9	1	0	7	0,992096026	4,813922576	6,282552145	0,992096026	OK		
20	Pecangaan	12	2	0	8,1	2,546437632	1,976546778	9,295551424	1,976546778		OK	
21	Jepara	3	1	0	8,2	6,893979781	10,72898738	0,530215165	0,530215165			OK
22	Jepara	3	1	0	8	6,874847911	10,72527774	0,372303105	0,372303105			OK
23	Mlonggo	9	2	0	5,6	2,350028277	5,337770013	6,667351241	2,350028277	OK		
24	Mlonggo	9	2	0	7,5	1,461217104	4,832998157	6,318029637	1,461217104	OK		
25	Jepara	3	1	1	6,8	6,924491234	10,80459214	1,30829914	1,30829914			OK
26	Jepara	3	3	0	7,3	7,187227317	10,9111147	2,037468103	2,037468103			OK

Gambar 4. 5. Sampel Pengelompokan Data ke dalam Cluster iterasi ke -2 pada Microsoft Excel.

Dari *centroid* baru iterasi ke-2 dilakukan perhitungan kembali, sehingga didapatkan hasil perhitungan jarak dan pengelompokkan data iterasi ke -2 dimana *cluster* 0 memiliki 90 data, *cluster* 1 memiliki 114 data dan *cluster* 2 memiliki 274 data. Berdasarkan dari perhitungan diatas bahwa pusat *cluster* baru (*centroid*) iterasi ke -2 seperti tabel dibawah ini :

Tabel 4. 6. Centroid Baru Iterasi ke-2

TITIK PUSAT	Asal Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah
CLUSTER 0	9,33]	1,01	0,24	7,69
CLUSTER 1	13,68	0,89	0,20	7,68
CLUSTER 2	2,81	1,03	0,13	7,72

Dari hasil *centroid* yang didapatkan pada iterasi ke-2, kemudian lakukan lagi perhitungan yang sama sampai data tiap *cluster* tidak ada lagi yang berubah.

	A	B	C	D	E	F	G	H	I	J	K	L
	Asal Kecamatan	Inisial Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah	C0	C1	C2	Jarak Terdekat	KC0	KC1	KC2
1	Jepara	3	1	0	7,7	6,277847157	10,65775817	0,233725147	0,233725147			OK
2	Demak	15	2	0	7,5	5,820184425	1,762115358	12,23110309	1,762115358		OK	
3	Pakis aji	11	1	0	8,7	1,994310412	2,866852567	8,249036201	1,994310412	OK		
4	Demak	15	1	0	8,8	5,830815555	1,78853114	12,23807149	1,78853114		OK	
5	Pati	15	1	0	8,7	5,813385918	1,725533056	12,22967074	1,725533056		OK	
6	Keling	6	2	1	4,1	5,044123622	8,546034502	4,999243709	4,999243709			OK
7	Pakis aji	11	1	0	6,7	2,029330705	2,82688122	8,254432103	2,029330705	OK		
8	Tahunan	13	0	1	6,6	4,097532674	1,717119335	10,33947999	1,717119335		OK	
9	Jepara	3	1	1	7,8	6,317778049	10,68703124	0,892973217	0,892973217			OK
10	Tahunan	13	0	0	8	3,882156315	1,176588662	10,24623269	1,176588662		OK	
11	Jepara	3	1	0	7,6	6,279204598	10,65769346	0,262830742	0,262830742			OK
12	Kedung	5	2	1	8,2	4,470992509	8,780751805	2,592358169	2,592358169			OK
13	Batealit	1	1	0	8,2	8,289574582	12,66946705	1,877029723	1,877029723			OK
14	Kedung	5	2	0	7,8	4,390666901	8,729984922	2,400188496	2,400188496			OK
15	Batealit	1	1	0	6,9	8,318571852	12,67901669	1,992767873	1,992767873			OK
16	Jepara	3	1	0	8,7	6,351449471	10,70988339	1,005038256	1,005038256			OK
17	Kembang	7	0	0	6,5	2,792766862	6,813643679	4,486056223	2,792766862	OK		
18	Mlonggo	9	1	0	7	0,823380736	4,704900107	6,233230911	0,823380736	OK		
19	Pecangaan	12	2	0	8,1	2,930637756	2,055325006	9,249566655	2,055325006		OK	
20	Jepara	3	1	0	8,2	6,294929519	10,67214626	0,531379968	0,531379968			OK
21	Jepara	3	1	0	8	6,283329417	10,66358041	0,362311717	0,362311717			OK
22	Mlonggo	9	2	0	5,6	2,37721216	5,20774247	6,616427223	2,37721216	OK		
23	Mlonggo	9	2	0	7,5	1,07110878	4,792402268	6,270766353	1,07110878	OK		
24	Jepara	3	1	1	6,8	6,386295798	10,71908782	1,281376882	1,281376882			OK
25	Jepara	3	3	0	7,3	6,596107766	10,86970873	2,028706322	2,028706322			OK
26	Pecangaan	12	0	0	7,1	2,991655276	1,965271461	9,269077372	1,965271461	OK		
27												

Gambar 4. 6. Sampel Pengelompokkan Data ke dalam Cluster iterasi ke -3 pada Microsoft Excel.

Karena pada iterasi ke-2 dan ke -3 posisi *cluster* tidak mengalami perubahan, maka iterasi dihentikan dan hasil dari perhitungan di atas bahwa pusat *cluster* baru (*centroid*) iterasi ke -3 seperti berikut :

Tabel 4.7. Centroid Baru Iterasi ke-3 (terakhir)

TITIK PUSAT	Asal Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah
CLUSTER 0	9,27	1,02	0,25	7,73
CLUSTER 1	13,6	0,89	0,20	7,64
CLUSTER 2	2,81	1,03	0,13	7,72

Dan hasil dari hasil akhir perhitungan diperoleh bahwa :

- 1) *Cluster 0* memiliki 88 data yang diartikan bahwa kelompok pertama adalah kategori wilayah yang rendah peminat.

Tabel 4. 8. Hasil *Cluster 0* Penerapan Algoritma *K-Means* (Micosoft Excel).

ASAL KECAMATAN SEKOLAH		ASAL SEKOLAH		JENIS KELAMIN	
PAKIS AJI	20	SMA	19	LAKI-LAKI	66
KEMBANG	7	SMK	48	PEREMPUAN	22
MLONGGO	27	MA	21		
MAYONG	18	PKBM	0		
TAHUNAN	1				
NALUMSARI	14				
BANGSRI	1				
TOTAL	88	TOTAL	88	TOTAL	88

Berdasarkan hasil *cluster 0* dapat diketahui terdapat 88 data mahasiswa. Dalam tabel tersebut terdapat informasi tentang persebaran wilayah kecamatan asal, asal sekolah mahasiswa maupun jenis kelamin beserta dengan jumlah dari masing-masing. Jumlah data mahasiswa pada *cluster 0* adalah jumlah data yang paling sedikit dibandingkan dengan *cluster* yang lain. Informasi yang menonjol dari *cluster 0* adalah informasi persebaran asal kecamatan sekolah mahasiswa yaitu banyaknya jumlah persebaran asal kecamatan sekolah yang berbeda beda dengan frekuensi yang berbeda pula. Berdasarkan hal tersebut dapat diketahui bahwa

persebaran asal kecamatan sekolah dengan frekuensi paling sedikit adalah kecamatan yang kurang diminati untuk masuk di Universitas Islam Nahdlatul Ulama Jepara

- 2) *Cluster 1* memiliki 116 data yang diartikan bahwa kelompok kedua adalah kategori wilayah yang sedang peminat.

Tabel 4. 9. Hasil *Cluster 1* Penerapan Algoritma *K-Means* (Microsoft Excel).

ASAL KECAMATAN SEKOLAH		ASAL SEKOLAH		JENIS KELAMIN	
PECANGAAN	27	SMA	46	LAKI-LAKI	93
TAHUNAN	35	SMK	39	PEREMPUAN	23
WELAHAN	5	MA	29		
LUAR KECAMATAN JEPARA	49	PKBM	2		
TOTAL	116	TOTAL	116	TOTAL	116

Berdasarkan hasil *cluster 1* dapat diketahui terdapat 116 data mahasiswa. Jumlah *cluster 1* lebih kurang dari jumlah *cluster 2* tetapi lebih besar dari jumlah *cluster 0*. Dalam tabel tersebut terdapat informasi tentang persebaran wilayah kecamatan asal, asal sekolah mahasiswa maupun jenis kelamin beserta dengan jumlah dari masing-masing. Informasi yang menonjol dari *cluster 1* adalah informasi asal sekolah dan jenis kelamin mahasiswa. Berdasarkan hal tersebut dapat diketahui bahwa asal sekolah dan jenis kelamin mahasiswa dengan frekuensi paling sedikit adalah yang kurang diminati untuk masuk di Universitas Islam Nahdlatul Ulama Jepara

- 3) *Cluster 2* memiliki 274 data yang diartikan bahwa kelompok ketiga adalah kategori wilayah yang tinggi peminat.

Tabel 4. 10. Hasil Cluster 2 Penerapan Algoritma *K-Means* (*Microsoft Excel*).

ASAL KECAMATAN SEKOLAH		ASAL SEKOLAH		JENIS KELAMIN	
BATEALIT	21	SMA	54	LAKI-LAKI	238
DONOROJO	6	SMK	163	PEREMPUAN	36
JEPARA	173	MA	52		
KALINYAMATAN	6	PKBM	5		
KEDUNG	28				
KELING	9				
BANGSRI	31				
TOTAL	274	TOTAL	274	TOTAL	274

Berdasarkan hasil *cluster 2* dapat diketahui terdapat 274 data mahasiswa. Dalam tabel tersebut terdapat informasi tentang persebaran wilayah kecamatan asal mahasiswa maupun asal sekolah mahasiswa beserta dengan jumlah dari masing-masing kecamatan asal dan asal sekolah tersebut. Jumlah data mahasiswa pada *cluster 2* adalah jumlah data yang paling banyak diantara cluster yang lain sehingga dapat dikatakan bahwa mayoritas mahasiswa UNISNU paling banyak masuk ke dalam kelompok *cluster 2*. Berdasarkan hal tersebut pihak admisi dapat mempertimbangkan strategi promosi di tahun depan menggunakan hasil persebaran wilayah dan asal sekolah pada *cluster 2*.

4.2.2. Penerapan Algoritma *K-Medoids* Menggunakan Aplikasi *Microsoft Excel*.

Beberapa tahapan dalam proses perhitungan manual algoritma *K-Medoids* menggunakan aplikasi *Microsoft Excel* hampir sama dengan perhitungan algoritma *K-Means*. Adapun tahapan-tahapannya adalah sebagai berikut :

1. Menentukan Jumlah Cluster.

Dalam hal ini menentukan jumlah *cluster* yang digunakan pada data calon mahasiswa sama dengan algoritma *K-Means* sebanyak 3 *cluster* yaitu diantaranya wilayah tinggi, wilayah sedang dan wilayah rendah peminatnya berdasarkan data calon mahasiswa baru periode 2016 sampai periode 2019.

2. Menentukan Nilai Titik Pusat Awal Cluster.

Penentuan titik pusat awal *cluster (centroid)* dilakukan secara *random* atau secara acak yang diambil dari data yang ada, namun peneliti menggunakan *centroid* awal sama dengan algoritma *K-Means* yaitu seperti dibawah ini :

Tabel 4. 11. Titik Pusat Awal dari setiap *Cluster K-Medoids*

TITIK PUSAT	Asal Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah
C0	11	1	0	6,7
C1	12	2	1	7,7
C2	1	2	0	8,7

3. Menghitung Jarak Data ke setiap *Cluster* dan mengelompokkan data ke dalam *Cluster*.

Dalam menempatkan objek-objek non medoids ke dalam *cluster* yang paling dekat dengan medoids berdasarkan jarak *Euclidian Distance*. Menghitung jarak data pada algoritma *K-Medoids* sama dengan menghitung jarak data pada algoritma *K-Means*. Adapun perhitungan jarak secara manual adalah sebagai berikut :..

$$Distance = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + (x_{3i} - x_{3j})^2 + (x_{4i} - x_{4j})^2}$$

Kemudian proses selanjutnya adalah pengelompokkan data ke dalam *cluster* dan juga menghitung jumlah jarak terdekat antar *cluster*. Adapun perhitungannya adalah sebagai berikut :

$$d(1,0) = \sqrt{(3 - 11)^2 + (1 - 1)^2 + (0 - 0)^2 + (7,7 - 6,7)^2} = 8,66226$$

$$d(1,1) = \sqrt{(3 - 12)^2 + (1 - 2)^2 + (0 - 1)^2 + (7,7 - 7,7)^2} = 9,11043$$

$$d(1,2) = \sqrt{(3 - 1)^2 + (1 - 2)^2 + (0 - 0)^2 + (7,7 - 8,7)^2} = 2,4495$$

$$d(2,0) = \sqrt{(15 - 11)^2 + (2 - 1)^2 + (0 - 0)^2 + (7,5 - 6,7)^2} = 4,2$$

$$d(2,1) = \sqrt{(15 - 12)^2 + (2 - 2)^2 + (0 - 1)^2 + (7,5 - 7,7)^2} = 3,1686$$

$$d(2,2) = \sqrt{(15 - 1)^2 + (2 - 2)^2 + (0 - 0)^2 + (7,5 - 8,7)^2} = 14,051$$

Dari contoh perhitungan diatas dapat diketahui bahwa pada data pertama dan data kedua dari setiap atribut di kelompokkan ke *cluster 2* dan *cluster 1*. Untuk melihat hasil perhitungan jarak dan pengelompokkan data ke setiap medoids, dapat dilihat pada Tabel 4.12 berikut ini :

Tabel 4. 12. Sampel Hasil Perhitungan Jarak ke Setiap Medoids.

No	Data Ke i	Jarak ke medoids			Terdekat	Cluster
		C0	C1	C2		
1.	Jepara	8,06225774 8	9,11043357 9	2,4494897 43	2,449489743	C2
2.	Demak	4,2	3,16859590 4	14,051334 46	3,168595904	C1
...
476.	Mlonggo	2,15406592 3	3,32264954 5	8,1510735 49	2,154065923	C0
477.	Pecangan	1,97230829 2	3,04795013 1	11,648605 07	1,972308292	C0
478.	Jepara	8,03990049 7	9,1126286	2,5377155 08	2,537715508	C2
479.	Jepara	8,08949936 6	9,1126286	2,3748684 17	2,374868417	C2
JUMLAH COST					1260,96946	

Dalam perhitungan jarak dan pengelompokkan data diatas diperoleh bahwa *cluster 0* memiliki 111 data, *cluster 1* memiliki 93 data dan *cluster 2* memiliki 270 data.

- Menentukan *Centroid* Baru Sebagai Non Medoids secara Random atau Secara Acak.

Tabel 4. 13. Titik Pusat Baru (*Centroid*) Non Medoids Iterasi ke -1.

TITIK PUSAT	Asal Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah
C0	3	1	0	7,7
C1	15	2	0	7,5
C2	11	1	0	8,7

5. Menempatkan kembali objek-objek non *medoids* ke dalam cluster yang paling dekat dengan *medoids* untuk mencari jarak non *medoids* baru. Adapun cara perhitungannya seperti no 3. Adapun hasil hasil perhitungan jarak ke setiap non *medoids* baru dapat dilihat pada tabel 4.14 dibawah ini.

Tabel 4. 14. Sampel Hasil Perhitungan Jarak ke Setiap *Medoids* Iterasi ke -1.

No	Data Ke i	Jarak ke medoids			Terdekat	Cluster
		C0	C1	C2		
1.	Jepara	0	12,0432553 7	8,0622577 48	0	C0
2.	Demak	12,0432553 7	0	4,2941821 11	0	C1
...
476.	Mlonggo	6,00333240 8	6,08276253	2,3323807 58	2,332380758	C2
477.	Pecangan	9,39627585 8	4,03112887 4	3,8327535 79	3,832753579	C2
478.	Jepara	0,2	12,0415945 8	8,0894993 66	0,2	C0
479.	Jepara	0,2	12,0482363 9	8,0399004 97	0,2	C0
JUMLAH COST					825,3021379	

Hasil perhitungan diatas menghasilkan pengelompokkan data bahwa *cluster 0* memiliki 281 data, *cluster 1* memiliki 67 data dan *cluster 2* memiliki 130 data

6. Menghitung nilai Simpangan (S) dengan persamaan sebagai berikut :

$$S = \text{jumlah cost baru} - \text{jumlah cost lama}$$

Keterangan :

S = Simpangan (selisih).

jumlah cost baru = jumlah cost non medoids.

jumlah cost lama = jumlah cost medoids.

Sehingga diperoleh :

$$S = 825,3021379 - 1260,96946 = -435,6673224$$

7. Menukar nilai non *Medoids* jika nilai $S < 0$ sehingga *centroid non Medoids* menjadi *centroid Medoids*.
8. Membuat *centroid* baru non *Medoids* sebagai berikut.

Tabel 4. 15. Titik Pusat Baru (*Centroid*) Non *Medoids* iterasi ke -2.

TITIK PUSAT	Asal Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah
C0	12	1	0	5
C1	3	1	0	7,5
C2	3	1	0	7,9

9. Menempatkan kembali objek-objek non medoids ke dalam *cluster* yang paling dekat dengan *medoids* untuk mencari jarak non *medoids* baru.

Tabel 4. 16. Sampel Hasil Perhitungan Jarak ke Setiap *Medoids* Iterasi ke -2.

No	Data Ke i	Jarak ke medoids			Terdekat	Cluster
		C0	C1	C2		
1.	Jepara	9,3962758 58	0,2	0,2	0,2	C0
2.	Demak	4,0311288 74	12,0415945 8	12,048236 39	4,031128874	C0
...
47 6.	Mlonggo	3,9051248 38	6	6,0133185 51	3,905124838	C0
47 7.	Pecang aan	0	9,34077084 6	9,4556861 2	0	C0
47 8.	Jepara	9,3407708 46	0	0,4	0	C1
47 9.	Jepara	9,4556861 2	0,4	0	0	C2
JUMLAH COST					1129,397747	

10. Hitung kembali simpangan (S) dari *centroid* baru non *medoids*.

$$S = 1129,397747 - 825,3021379 = 304,0956092$$

11. Jika nilai $S > 0$ maka proses pengklusteran dihentikan. Sehingga diperoleh anggota setiap *cluster* sebagai berikut.

AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	A
Asal Kecamatan	Inisial Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah	CLUSTER 0	CLUSTER 1	CLUSTER 2	Jarak Terdekat	C0	C1	C2	
Jejara	3	1	0	7,7	9,396275858	0,2	0,2	0,2				
Demak	15	2	0	7,5	4,031128874	12,04159458	12,04823639	4,031128874	OK			
Pakis aji	11	1	0	8,7	3,832753579	8,089499366	8,039900497	3,832753579	OK			
Demak	15	1	0	8,8	4,841487375	12,07021127	12,03370267	4,841487375	OK			
Pati	15	1	0	8,7	4,763402146	12,05985075	12,0266371	4,763402146	OK			
Keling	6	2	1	4,1	6,229767251	4,749736835	5,043808085	4,749736835		OK		
Pakis aji	11	1	0	6,7	1,972308292	8,039900497	8,089499366	1,972308292	OK			
Tahunan	13	0	1	6,6	2,357965225	10,13952662	10,18282868	2,357965225	OK			
Jejara	3	1	1	7,8	9,478396489	1,044030651	1,004987562	1,004987562			OK	
Tahunan	13	0	0	8	3,31662479	10,0623059	10,05037313	3,31662479	OK			
Jejara	3	1	0	7,6	9,368030743		0,1	0,1			OK	
Kedung	5	2	1	8,2	7,825599019	2,547547841	2,467792536	2,467792536			OK	
Batealit	1	1	0	8,2	11,45600279	2,11896201	2,022374842	2,022374842			OK	
Kedung	5	2	0	7,8	7,605261337	2,256102835	2,238302929	2,238302929			OK	
Batealit	1	1	0	6,9	11,16288493	2,088061302	2,236067977	2,088061302			OK	
Jejara	3	1	0	8,7	9,730878686		1,2	0,8			OK	
Kembang	7	0	0	6,5	5,315072906	4,242640687	4,354308211	4,242640687			OK	
Mlonggo	9	1	0	7	3,605551275	6,020797289	6,067124525	3,605551275	OK			
Pecangaan	12	2	0	8,1	3,257299495	9,075241044	9,057593499	3,257299495	OK			
Jejara	3	1	0	8,2	9,551963149		0,7	0,3			OK	
Jejara	3	1	0	8	9,486832981		0,5	0,1			OK	
Mlonggo	9	2	0	5,6	3,218695388	6,372597587	6,503076195	3,218695388	OK			
Mlonggo	9	2	0	7,5	4,031128874	6,08276253	6,095900262	4,031128874	OK			
Jenara	3	1	1	6,8	9,23255111	1,220655562	1,486606875	1,220655562			OK	

Gambar 4. 7. Sampel Hasil Pengklusteran Data dengan *K-Medoids Clustering*

Berdasarkan perhitungan diatas diperoleh hasil sebagai berikut :

- 1) *Cluster 0* memiliki 189 data yang diartikan bahwa kelompok pertama adalah kategori wilayah tinggi peminat.

Tabel 4. 17. Hasil *Cluster 0* Penerapan Algoritma *K-Medoids* (*Microsoft Excel*).

ASAL KECAMATAN SEKOLAH		ASAL SEKOLAH		JENIS KELAMIN	
MAYONG	10	SMA	60	LAKI-LAKI	149
MLONGGO	28	SMK	81	PEREMPUAN	40
NALUMSARI	14	MA	46		
PAKIS AJI	21	PKBM	2		
PECANGAAN	27				
TAHUNAN	35				
WELAHAN	5				
LUAR KECAMATAN JEPARA	49				
TOTAL	189	TOTAL	189	TOTAL	189

Berdasarkan hasil *cluster 0* dapat diketahui terdapat 189 data mahasiswa. Dalam tabel tersebut terdapat informasi tentang persebaran wilayah kecamatan asal mahasiswa maupun asal sekolah mahasiswa beserta dengan jumlah dari masing-masing kecamatan asal dan asal sekolah tersebut. Jumlah data mahasiswa pada *cluster 0* adalah jumlah data yang paling banyak diantara *cluster* yang lain sehingga dapat dikatakan bahwa mayoritas mahasiswa UNISNU paling banyak masuk ke dalam kelompok *cluster 0*. Berdasarkan hal tersebut pihak admisi dapat mempertimbangkan strategi promosi di tahun depan menggunakan hasil persebaran wilayah dan asal sekolah pada *cluster 0*.

2) *Cluster 1* memiliki 137 data yang diartikan bahwa kelompok kedua adalah kategori wilayah yang rendah peminat.

Tabel 4. 18. Hasil *Cluster 1* Penerapan Algoritma *K-Medoids* (*Micosoft Excel*).

ASAL KECAMATAN SEKOLAH		ASAL SEKOLAH		JENIS KELAMIN	
BATEALIT	14	SMA	29	LAKI-LAKI	113
DONOROJO	4	SMK	73	PEREMPUAN	24
JEPARA	81	MA	32		
KALINYAMATAN	4	PKBM	3		
KEDUNG	9				
KELING	4				
BANGSRI	18				
KEMBANG	3				
TOTAL	137	TOTAL	137	TOTAL	137

Berdasarkan hasil *cluster 1* dapat diketahui terdapat 137 data mahasiswa. Dalam tabel tersebut terdapat informasi tentang persebaran wilayah kecamatan asal, asal sekolah mahasiswa maupun jenis kelamin beserta dengan jumlah dari masing-masing. Jumlah data mahasiswa pada *cluster 1* adalah jumlah data yang paling sedikit dibandingkan dengan *cluster* yang lain. Informasi yang menonjol dari *cluster 1* adalah informasi persebaran asal kecamatan sekolah dan jenis kelamin mahasiswa. Berdasarkan hal tersebut dapat diketahui bahwa persebaran asal

kecamatan sekolah dan jenis kelamin dengan frekuensi paling sedikit adalah kurang minat untuk masuk di Universitas Islam Nahdlatul Ulama Jepara.

- 3) *Cluster 2* memiliki 152 data yang diartikan bahwa kelompok ketiga adalah kategori wilayah sedang peminat.

Tabel 4. 19. Hasil *Cluster 2* Penerapan Algoritma *K-Medoidss* (*Microsoft Excel*).

ASAL KECAMATAN SEKOLAH		ASAL SEKOLAH		JENIS KELAMIN	
BATEALIT	7	SMA	30	LAKI-LAKI	135
DONOROJO	2	SMK	96	PEREMPUAN	17
JEPARA	92	MA	24		
KALINYAMATAN	2	PKBM	2		
KEDUNG	19				
KELING	5				
BANGSRI	13				
KEMBANG	4				
TOTAL	152	TOTAL	152	TOTAL	152

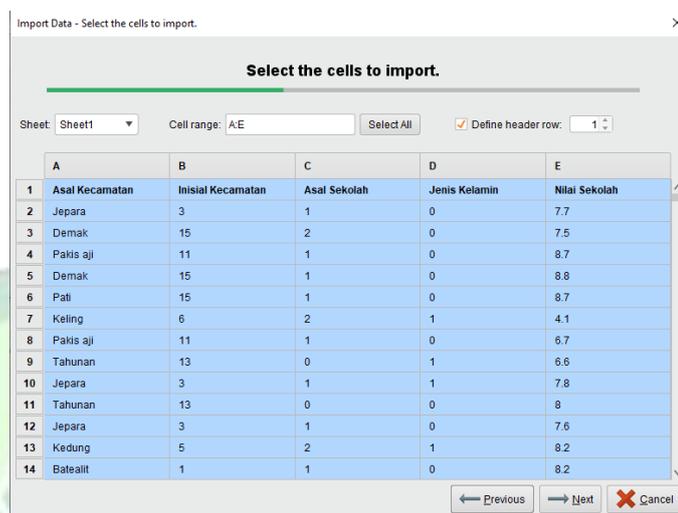
Berdasarkan hasil *cluster 2* dapat diketahui terdapat 152 data mahasiswa. Dalam tabel tersebut terdapat informasi tentang persebaran wilayah kecamatan asal, asal sekolah mahasiswa maupun jenis kelamin beserta dengan jumlah dari masing-masing. Jumlah data mahasiswa pada *cluster 2* adalah jumlah data yang lebih kurang dibandingkan dengan *cluster 0* dan lebih besar dari jumlah data pada *cluster 1*. Informasi yang menonjol dari *cluster 2* adalah informasi persebaran asal sekolah mahasiswa yaitu banyaknya jumlah jurusan asal sekolah yang berbeda beda dengan frekuensi yang berbeda pula. Berdasarkan hal tersebut dapat diketahui bahwa persebaran asal sekolah dengan frekuensi paling sedikit adalah sekolah yang kurang diminati untuk masuk di Universitas Islam Nahdlatul Ulama Jepara.

4.3. Penerapan Algoritma menggunakan aplikasi *Rapidminer Studio*.

RapidMiner adalah perangkat lunak untuk pengolahan data dengan menggunakan prinsip dan algoritma data mining. *RapidMiner* mengekstrak pola-

pola dataset yang besar dengan mengkombinasi metode statistika, kecerdasan buatan dan database.

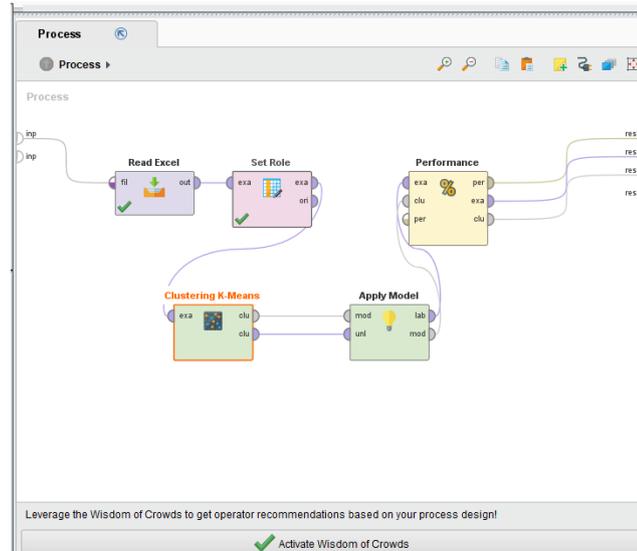
Data yang sudah melalui proses *pre-processing* data selanjutnya akan di olah menggunakan aplikasi Rapidminer Studio. Pada Gambar 4.8 dibawah ini adalah proses melakukan import data ke *RapidMiner Studio*.



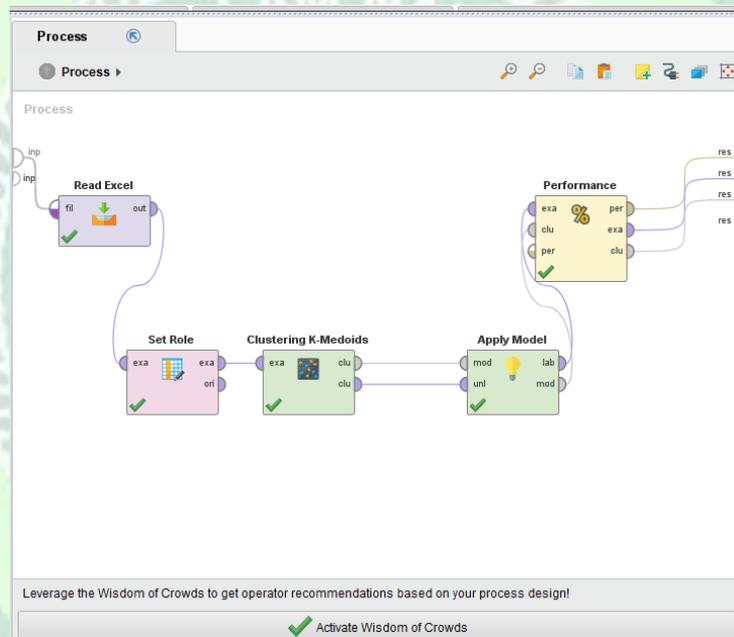
Gambar 4. 8. Proses impor dataset ke *RapidMiner Studio* algoritma *K-means*.

Pada Gambar 4.8 adalah proses import data ke *RapidMiner*, peneliti menggunakan operator *Read Excel* dengan fitur *import configuration wizard*. Fitur tersebut digunakan untuk memasukkan file data serta memilih tipe data atribut dan *Role*.

Selanjutnya pemodelan desain *clustering* menggunakan algoritma *K-Means Clustering* dan *K-Medoids* seperti yang ditunjukkan pada Gambar di bawah ini.



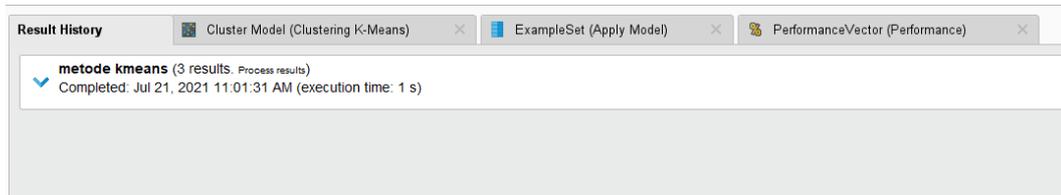
Gambar 4. 9. Desain *Clustering* Menggunakan Algoritma *K-Means Clustering*.



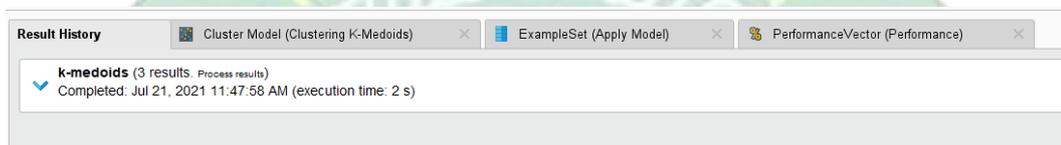
Gambar 4. 10. Desain *Clustering* Menggunakan Algoritma *K-Medoids Clustering*.

Pada Gambar 4.9 dan 4.10 diatas dapat dilihat bahwa proses pengujian *clustering* dimulai dengan operator *Read Excel* data PMB 2016 – 2019 dengan fitur *Import Configuration Wizard*, fitur ini digunakan unuk memasukkan data file ke dalam Rapidminer. Dataset itu kemudian dihubungkan terlebih dahulu ke *set Role* sebelum dilakukan pengujian menggunakan algoritma *Clusteing K-Means* dan *K-Medoids*. Setelah dihubungkan dengan model *clustering* algoritma *K-Means* dan *K-Medoids*, selanjutnya dihubungkan lagi dengan operator *Apply Model*. Kemudian

dihubungkan dengan operator *Performance* untuk menghasilkan output dari algoritma *Clustering*. Dan terakhir dihubungkan dengan ke *Result*. Setelah pengujian berjalan maka didapatkan informasi lama waktu pemrosesan *clustering* seperti yang ditunjukkan pada Gambar 4.11 dan 4.12.

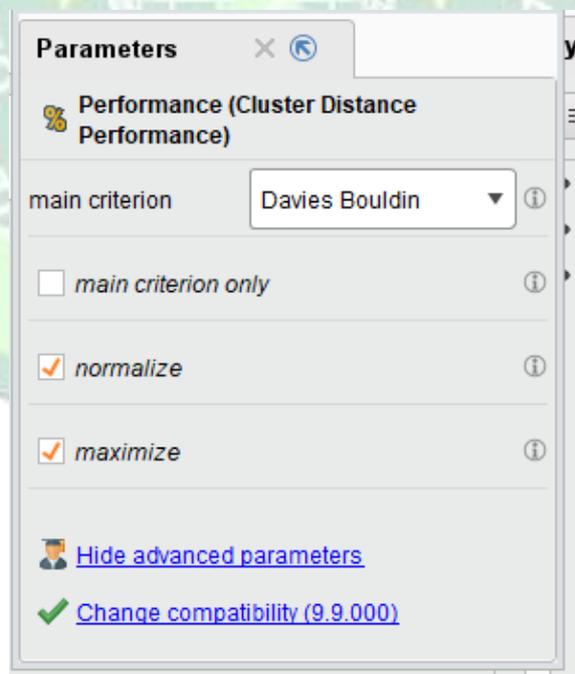


Gambar 4. 11. Result History menggunakan Algoritma *K-Means Clustering*



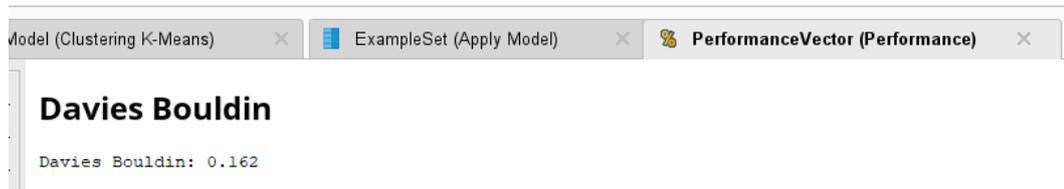
Gambar 4. 12. Result History menggunakan Algoritma *K-Medoids Clustering*.

Dalam memvalidasi penelitian ini, peneliti menggunakan *Davies Bouldin Index* (DBI) dapat dilihat pada gambar 4.13 dibawah ini.

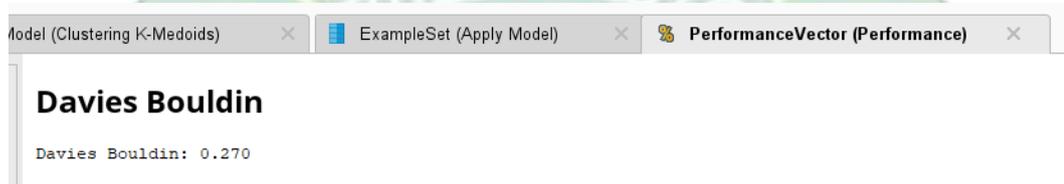


Gambar 4. 13. Parameter Metode *Davies Bouldin Index* Algoritma *Clustering*

Hasil dari nilai *Davies Bouldin* dan deskripsi *Performance* yang menghasilkan *Centroid Distance* pada setiap *cluster* yang ada didalam *Performance Vector K-Means* dapat dilihat pada gambar di bawah ini. Semakin rendah nilai DBI, maka *cluster* tersebut semakin baik.



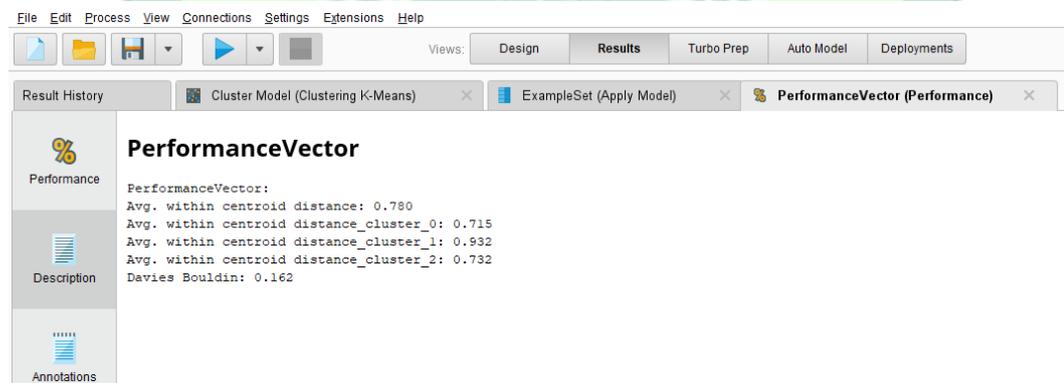
Gambar 4. 14. *Davies Bouldin Index (DBI) Algoritma K-Means Clustering*



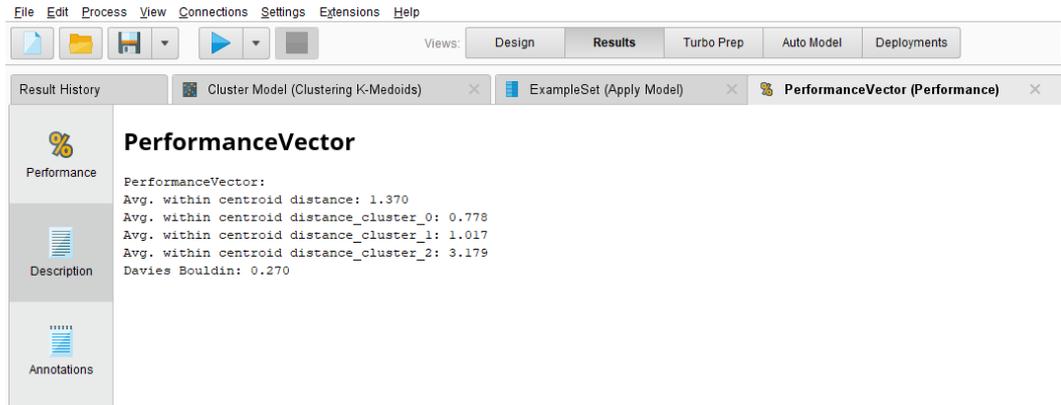
Gambar 4. 15. *Davies Bouldin Index (DBI) untuk Performance Vector Algoritma K-Medoids Clustering.*

Dari gambar 4.14 dan 4.15 diatas diperoleh nilai DBI *K-Means* dan *K-Medoids* secara berurutan adalah 0,162 dan 0,270. Dari hasil DBI tersebut berarti valid dan sangat baik karena nilai tersebut < 1 dan mendekati nilai 0.

Untuk menunjukkan deskripsi *Performance* yang menghasilkan *Centroid Distance* pada setiap *cluster* dapat dilihat pada Gambar 4.16 dan 4.17.

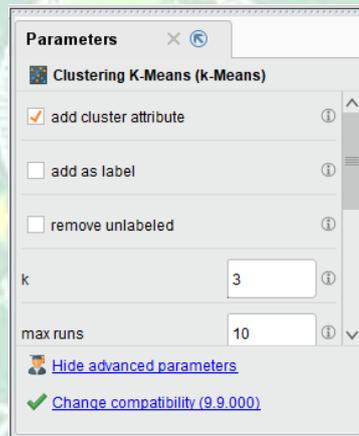


Gambar 4. 16. Hasil DBI dan *Cenroid Distance Performance Vector* Algoritma *K-Means Clustering*.

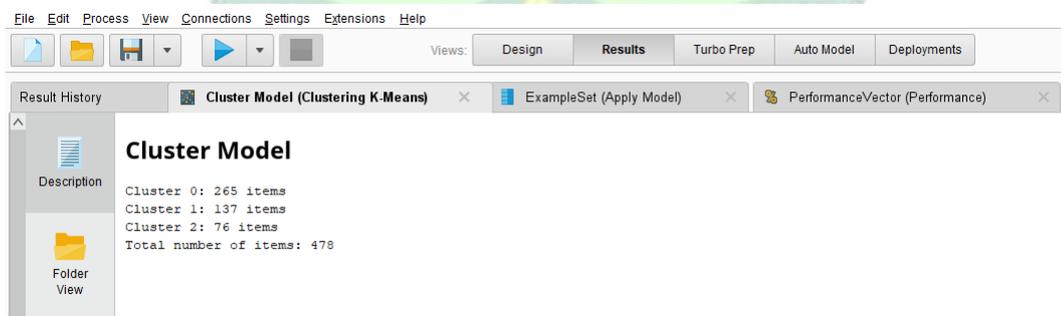


Gambar 4. 17. Hasil DBI dan *Cenroid Distance* Algoritma *K-Medoids Clustering*.

Pada Gambar 4.19 dan 4.20 menunjukkan bahwa penelitian ini *cluster* yang dihasilkan berjumlah 3 *cluster*, dimulai dari *cluster* 0 dan diakhiri dengan *cluster* 2. Total *dataset* yang di-*cluster* adalah 478 dengan jumlah items pada tiap-tiap *cluster*. Untuk pengaturan jumlah *cluster* yang diinginkan dapat dilihat pada Gambar 4.18.

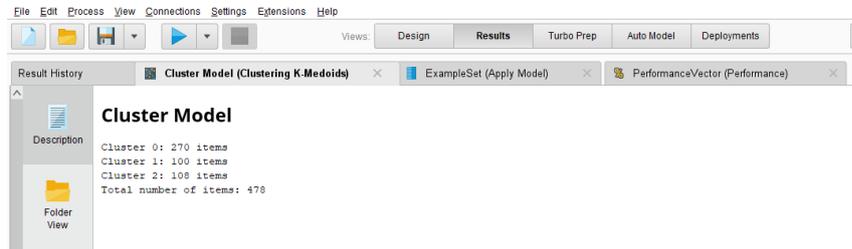


Gambar 4. 18. Penentuan Jumlah *Cluster*.



Gambar 4. 19. Hasil *Cluster* Algoritma *K-Means Clustering*.

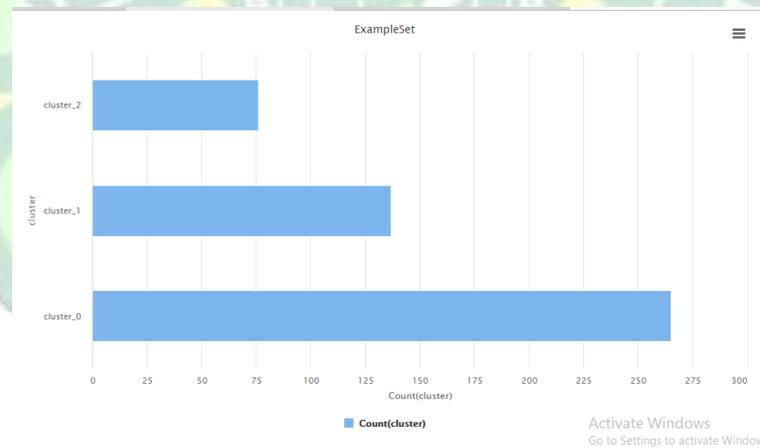
Pada Gambar 4.19 menunjukkan bahwa penelitian menggunakan algoritma *K-Means* ini menghasilkan *cluster 0* berjumlah 265 items, jumlah *cluster 1* adalah 137 items, *cluster 2* berjumlah 76 items pada tiap-tiap *cluster*.



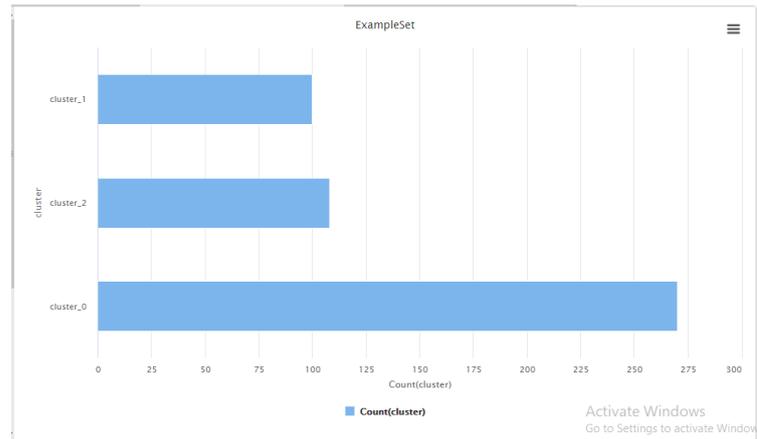
Gambar 4. 20. Hasil Cluster Algoritma *K-Medoids Clustering*.

Pada Gambar 4.20 menunjukkan bahwa penelitian menggunakan algoritma *K-Medoids* menghasilkan *cluster 0* berjumlah 270 items, jumlah *cluster 1* adalah 100 items, *cluster 2* berjumlah 108 items pada tiap-tiap *cluster*.

Chart merupakan tampilan grafik hasil pengelompokan atau *cluster* sampel data PMB periode 2016 - 2019 dengan 3 *cluster*. Berikut ini adalah hasil dari *cluster* model *K-Means* dan *K-Medoids* yang digambarkan dalam bentuk Chart Bar yang dapat dilihat pada gambar 4.21 dan 4.22.



Gambar 4. 21. Tampilan Chart Bar Algoritma *K-Means Clustering*.



Gambar 4.22. Tampilan Chart Bar Algoritma *K-Medoids Clustering*.

Berdasarkan perhitungan hasil penerapan algoritma *K-Means* menggunakan aplikasi *RapidMiner studio* diatas diperoleh hasil sebagai berikut :

1. *Cluster 0* memiliki 265 data yang diartikan bahwa kelompok pertama adalah kategori wilayah tinggi peminat.

Tabel 4.20. Hasil *Cluster 0* Penerapan Algoritma *K-Means (Rapid Miner)*.

ASAL KECAMATAN SEKOLAH		ASAL SEKOLAH		JENIS KELAMIN	
BATEALIT	21	SMA	53	LAKI-LAKI	236
DONOROJO	6	SMK	156	PEREMPUAN	34
JEPARA	173	MA	51		
KALINYAMATAN	6	PKBM	5		
KEDUNG	28				
KELING	5				
BANGSRI	31				
TOTAL	265	TOTAL	265	TOTAL	265

2. *Cluster 1* memiliki 137 data yang diartikan bahwa kelompok kedua adalah kategori wilayah yang rendah peminat.

Tabel 4. 21. Hasil Cluster 1 Penerapan Algoritma K-Means (Rapid Miner).

ASAL KECAMATAN SEKOLAH		ASAL SEKOLAH		JENIS KELAMIN	
PAKIS AJI	19	SMA	46	LAKI-LAKI	75
PECANGAAN	11	SMK	59	PEREMPUAN	25
NALUMSARI	13	MA	30		
MLONGGO	28	PKBM	2		
MAYONG	18				
KEMBANG	7				
KELING	4				
TOTAL	100	TOTAL	137	TOTAL	137

3. Cluster 2 memiliki 76 data yang diartikan bahwa kelompok ketiga adalah kategori wilayah sedang peminat.

Tabel 4. 22. Hasil Cluster 2 Penerapan Algoritma K-Means (Rapid Miner).

ASAL KECAMATAN SEKOLAH		ASAL SEKOLAH		JENIS KELAMIN	
PAKIS AJI	2	SMA	20	LAKI-LAKI	86
NALUMSARI	1	SMK	35	PEREMPUAN	22
PECANGAAN	12	MA	21		
TAHUNAN	35	PKBM	0		
WELAHAN	5				
LUAR KECAMATAN JEPARA	49				
TOTAL	76	TOTAL	76	TOTAL	76

Sedangkan hasil perhitungan penerapan algoritma *K-Medoids* menggunakan aplikasi RapidMiner studio diperoleh hasil sebagai berikut :

1. Cluster 0 memiliki 270 data yang diartikan bahwa kelompok pertama adalah kategori wilayah tinggi peminat.

Tabel 4. 23. Hasil *Cluster 0* Penerapan Algoritma *K-Medoids (Rapid Miner)*.

ASAL KECAMATAN SEKOLAH		ASAL SEKOLAH		JENIS KELAMIN	
BATEALIT	21	SMA	53	LAKI-LAKI	236
DONOROJO	6	SMK	161	PEREMPUAN	34
JEPARA	173	MA	51		
KALINYAMATAN	6	PKBM	5		
KEDUNG	28				
KELING	5				
BANGSRI	31				
TOTAL	270	TOTAL	274	TOTAL	274

2. *Cluster 1* memiliki 100 data yang diartikan bahwa kelompok kedua adalah kategori wilayah yang rendah peminat.

Tabel 4. 24. Hasil *Cluster 1* Penerapan Algoritma *K-Medoids (Rapid Miner)*.

ASAL KECAMATAN SEKOLAH		ASAL SEKOLAH		JENIS KELAMIN	
PAKIS AJI	19	SMA	25	LAKI-LAKI	75
PECANGAAN	11	SMK	51	PEREMPUAN	25
NALUMSARI	13	MA	24		
MLONGGO	28	PKBM	0		
MAYONG	18				
KEMBANG	7				
KELING	4				
TOTAL	100	TOTAL	100	TOTAL	100

3. *Cluster 2* memiliki 108 data yang diartikan bahwa kelompok ketiga adalah kategori wilayah sedang peminat.

Tabel 4. 25. Hasil Cluster 2 Penerapan Algoritma *K-Medoids* (*RapidMiner*).

ASAL KECAMATAN SEKOLAH		ASAL SEKOLAH		JENIS KELAMIN	
PAKIS AJI	2	SMA	41	LAKI-LAKI	86
NALUMSARI	1	SMK	38	PEREMPUAN	22
PECANGAAN	12	MA	27		
TAHUNAN	35	PKBM	2		
WELAHAN	5				
LUAR KECAMATAN JEPARA	49				
TOTAL	108	TOTAL	108	TOTAL	108

4.4. Evaluasi

4.4.1. Pengujian Metode *Davies Bouldin Index*

A. Pengujian Metode *Davies Bouldin Index* (DBI) pada Algoritma *K-Means*.

Ada beberapa tahapan dalam pengujian menggunakan metode *Davies Bouldin Index* antara seperti berikut ini :

a. Menghitung SSW (*Sum of Square Within cluster*).

Dalam menghitung nilai SSW, Langkah pertama adalah menentukan iterasi yang akan di hitung. Iterasi yang diambil adalah iterasi terakhir karena memiliki kualitas centroid yang lebih baik.

Tabel 4. 26. Nilai *Centroid* Iterasi Terakhir *K-Means*.

Titik Pusat	Asal Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah
Cluster 0	9,27	1,02	0,25	7,73
Cluster 1	13,65	0,89	0,20	7,64
Cluster 2	2,81	1,03	0,13	7,72

Kemudian menghitung jarak antar *cluster* pada anggota *cluster* dengan persamaan seperti dibawah ini :

$$\text{Jarak } d_{i,c} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + (x_{3i} - x_{3j})^2 + (x_{4i} - x_{4j})^2}$$

$$\begin{aligned} Dc0 &= \sqrt{(11 - 9,27)^2} + ((1 - 1,02)^2) + ((0 - 0,25)^2) + ((8,7 - 7,73)^2)) \\ &= 1,99431 \end{aligned}$$

$$\begin{aligned} Dc1 &= \sqrt{(15 - 9,27)^2} + ((2 - 1,02)^2) + ((0 - 0,25)^2) + ((7,5 - 7,73)^2)) \\ &= 5,820184 \end{aligned}$$

$$\begin{aligned} Dc2 &= \sqrt{(15 - 9,27)^2} + ((2 - 1,02)^2) + ((0 - 0,25)^2) + ((7,5 - 7,73)^2)) \\ &= 6,557788 \end{aligned}$$

Kemudian langkah selanjutnya menghitung *Sum of Square Within cluster* (SSW) dengan persamaan dibawah ini :

$$\begin{aligned} SSW 0 &= \frac{\text{Jumlah hasil jarak anggota cluster}}{\text{jumlah anggota cluster}} \\ &= \frac{1,99431 + 2,0293 + \dots + 1,3184 + 0,43900}{88} = 1,56561 \end{aligned}$$

$$SSW 1 = \frac{5,8201841 + 5,830816 + \dots + 5,738883 + 3,870724}{116} = 4,58171$$

$$SSW 2 = \frac{5,82018 + 5,044124 + \dots + 6,282154 + 3,870724}{274} = 6,55778$$

Hasil dari menghitung jarak antar anggota *cluster* dan SSW dapat dilihat pada tabel dibawah ini.

Tabel 4. 27. Sampel hitung jarak anggota *Cluster 0* hasil *Clustering k-means*.

NO	Inisial Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah	jarak d-c
1	11	1	0	8,7	1,99431
2	11	1	0	6,7	2,0293
.....
87	9	0	1	7,5	1,3184
88	9	1	0	7,5	0,43900
SSW					1,56561

Tabel 4. 28. Sampel hitung jarak anggota *Cluster 1* hasil *Clustering k-means*.

NO	Inisial Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah	jarak d-c
1	15	2	0	7,5	5,820184
2	15	1	0	8,8	5,830816
.....
273	15	1	0	8	5,738883
274	12	1	0	5	3,870724
SSW					4,58171

Tabel 4. 29. Sampel hitung jarak anggota *Cluster 2* hasil *Clustering k-means*.

NO	Inisial Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah	jarak d-c
1	3	1	0	7,7	6,277847
2	6	2	1	4,1	5,044124
.....
273	3	1	0	7,5	6,282154
274	3	1	0	7,9	6,27991
				SSW	6,55778

b. Menghitung SSB (*Sum of Square Between cluster*).

$$\begin{aligned} \text{SSB}_{ij} &= d(c_i, c_j) \\ &= \sqrt{(c_i - c_j)^2} \end{aligned}$$

Keterangan :

c_i = Centroid tiap atribut *cluster i*

c_j = Centroid tiap atribut *cluster j*

$$\text{SSB}_{01} = \sqrt{(c_0 - c_1)^2}$$

$$\begin{aligned} &= \sqrt{(9,27 - 13,65)^2 + ((1,02 - 0,89)^2) + ((0,25 - 0,20)^2) + ((7,733 - 7,64)^2)} \\ &= 4,38 \end{aligned}$$

$$\begin{aligned} \text{SSB}_{02} &= \sqrt{(9,27 - 2,81)^2 + ((1,02 - 1,03)^2) + ((0,25 - 0,13)^2) + ((7,73 - 7,72)^2)} \\ &= 6,48 \end{aligned}$$

$$\begin{aligned} \text{SSB}_{12} &= \sqrt{(13,65 - 2,81)^2 + ((0,89 - 1,03)^2) + ((0,20 - 0,13)^2) + ((7,64 - 7,72)^2)} = 10,85 \end{aligned}$$

Tabel 4. 30. Hasil Perhitungan SSB Algoritma *K-means*

Menghitung SSB				
		CENTROID KE		
CENTROID KE	SSB	0	1	2
	0	0	4,39	6,48
	1	4,39	0	10,85
	2	6,48	10,85	0

c. Menghitung Nilai Rasio

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}}$$

$$R_{01} = \frac{SSW_0 + SSW_1}{SSB_{01}} = \frac{1,56561 + 4,58171}{4,39} = 1,401$$

$$R_{02} = \frac{SSW_0 + SSW_2}{SSB_{02}} = \frac{1,56561 + 6,55778}{6,48} = 1,254$$

$$R_{12} = \frac{SSW_1 + SSW_2}{SSB_{12}} = \frac{4,58171 + 6,55778}{10,84} = 1,027$$

d. Menghitung Nilai DBI.

$$DBI = \frac{1}{K} \sum_i^k \max_{1 \neq j} (R_{ij}) = \frac{\text{hasil rasio}}{\text{jumlah cluster}}$$

$$DBI_{01} = \frac{1,401}{3} = 0,467$$

$$DBI_{02} = \frac{1,254}{3} = 0,418$$

$$DBI_{12} = \frac{1,027}{3} = 0,342$$

Dari hasil perhitungan DBI diatas, peneliti mengambil hasil DBI yang bernilai 0,342 karena nilainya yang paling mendekati angka 0. Semakin kecil nilai DBI maka semakin baik .

B. Pengujian Metode *Davies Bouldin Index (DBI)* pada Algoritma *K-Medoids*.

Ada beberapa tahapan dalam pengujian menggunakan metode *Davies Bouldin Index* antara seperti berikut ini :

a. Menghitung SSW (*Sum of Square Within cluster*).

Tabel 4. 31. Nilai Centroid Iterasi Terakhir K-Medoids.

Titik Pusat	Asal Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah
Cluster 0	12	1	0	5
Cluster 1	3	1	0	7,5
Cluster 2	3	1	0	7,9

Kemudian menghitung jarak antar *cluster* pada anggota *cluster* dengan persamaan seperti dibawah ini :

$$\text{Jarak } d_{i,c} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + (x_{3i} - x_{3j})^2 + (x_{4i} - x_{4j})^2}$$

$$\begin{aligned} D_{c0} &= \sqrt{(15 - 12)^2 + ((2 - 1)^2) + ((0 - 0)^2) + ((7,5 - 5)^2)} \\ &= 4,031129 \end{aligned}$$

$$\begin{aligned} D_{c1} &= \sqrt{(6 - 3)^2 + ((2 - 1)^2) + ((1 - 0)^2) + ((4,1 - 7,5)^2)} \\ &= 4,749737 \end{aligned}$$

$$\begin{aligned} D_{c2} &= \sqrt{(3 - 3)^2 + ((1 - 1)^2) + ((1 - 0)^2) + ((7,8 - 7,9)^2)} \\ &= 1,0049 \end{aligned}$$

Kemudian langkah selanjutnya menghitung *Sum of Square Within cluster* (SSW) dengan persamaan dibawah ini :

$$\begin{aligned} \text{SSW } 0 &= \frac{\text{jumlah hasil jarak anggota cluster}}{\text{jumlah anggota cluster}} \\ &= \frac{4,031129 + 3,832754 + \dots + 0 + 3,257299}{190} = 3,604095 \end{aligned}$$

$$\text{SSW } 1 = \frac{4,749737 + 0,1 + \dots + 0 + 0,2}{138} = 1,510695$$

$$\text{SSW } 2 = \frac{1,00498 + 2,46779 + \dots + 0,2 + 0}{153} = 1,58722721$$

Hasil dari menghitung jarak antar anggota *cluster* dan SSW dapat dilihat pada tabel dibawah ini.

Tabel 4. 32. Sampel hitung jarak anggota *Cluster 0* hasil *Clustering k-medoids*.

NO	Inisial Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah	jarak d-c
1	15	2	0	7,5	4,031129
2	11	1	0	8,7	3,832754
.....
189	12	1	0	5	0
190	12	2	0	8,1	3,257299
SSW					3,604095

Tabel 4. 33. Sampel hitung jarak anggota *Cluster 1* hasil *Clustering k-medoids*.

NO	Inisial Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah	jarak d-c
1	6	2	1	4,1	4,749737
2	3	1	0	7,6	0,1
.....
137	3	1	0	7,5	0
138	3	1	0	7,7	0,2
SSW					1,5106

Tabel 4. 34. Sampel hitung jarak anggota *Cluster 2* hasil *Clustering k-medoids*.

NO	Inisial Kecamatan	Asal Sekolah	Jenis Kelamin	Nilai Sekolah	jarak d-c
1	3	1	1	7,8	1,0049
2	5	2	1	8,2	2,4677
.....
152	3	1	0	8,1	0,2
153	3	1	0	7,9	0
SSW					1,587227

b. Menghitung SSB (*Sum of Square Between cluster*).

$$SSB_{ij} = d(c_i, c_j)$$

$$= \sqrt{(c_i - c_j)^2}$$

Keterangan :

c_i = Centroid tiap atribut *cluster i*

c_j = Centroid tiap atribut *cluster j*

$$SSB_{01} = \sqrt{(c_0 - c_1)^2}$$

$$= \sqrt{\begin{matrix} (12 - 3)^2 + \\ (1 - 1)^2 \\ + ((0 - 0)^2) \\ + ((5 - 7,5)^2 \end{matrix}}$$

$$= 9,34077084$$

$$SSB_{02} = \sqrt{\begin{matrix} (12 - 3)^2 + \\ (1 - 1)^2 \\ + ((0 - 0)^2) \\ + ((5 - 7,9)^2 \end{matrix}}$$

$$= 9,45568612$$

$$SSB_{12} = \sqrt{\begin{matrix} (3 - 3)^2 + \\ (1 - 1)^2 \\ + ((0 - 0)^2) \\ + ((7,5 - 7,9)^2 \end{matrix}}$$

$$= 0,4$$

Tabel 4. 35. Hasil Perhitungan SSB Algoritma *K-medoids*

Menghitung SSB					
		CENTROID KE			
		SSB	0	1	2
CENTROID KE	0	0	9,340770846	9,45568612	
	1	9,340770846	0	0,4	
	2	9,45568612	0,4	0	

c. Menghitung Nilai Rasio

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}}$$

$$R_{01} = \frac{SSW_0 + SSW_1}{SSB_{01}} = \frac{3,604095 + 1,510695}{9,340770846} = 0,547576891$$

$$R_{02} = \frac{SSW_0 + SSW_2}{SSB_{02}} = \frac{3,6040951 + 1,58722721}{9,45568612} = 0,549015932$$

$$R_{12} = \frac{SSW1+SSW2}{SSB12} = \frac{1,510695+1,58722721}{0,4} = 7,74480588$$

d. Menghitung Nilai DBI.

$$DBI = \frac{1}{K} \sum_{i=1}^k \max_{i \neq j} (R_{ij}) = \frac{\text{hasil rasio}}{\text{jumlah cluster}}$$

$$DBI_{01} = \frac{0,547576891}{3} = 0,182$$

$$DBI_{02} = \frac{0,549015932}{3} = 0,183$$

$$DBI_{12} = \frac{7,74480588}{3} = 2,581$$

Dari hasil perhitungan DBI diatas, peneliti mengambil hasil DBI yang bernilai 0,182 karena nilainya yang paling mendekati angka 0. Semakin kecil nilai DBI maka semakin baik .

4.4.2. Pengujian menggunakan Metode ROC

Metode ROC digunakan untuk menghitung nilai akurasi hasil *clustering* yang telah diproses pada sistem. Selain nilai akurasi , nilai *sensifitas* dan nilai *Spesifitas* Pada penelitian ini adalah data dari hasil *clustering* data PMB 2016 sampai 2019 yang berupa data nilai *centroid* awal dan *centroid* dari iterasi terakhir berdasarkan nilai ujian sekolah yang ada pada data tersebut.

Tabel 4. 36. Nilai Centroid Data PMB Periode 2016 – 2019 pada Algoritma *K-Means Clustering*.

Titik Pusat	Centroid Awal	Centroid Iterasi Terakhir
<i>Cluster 0</i>	6,7	7,73
<i>Cluster 1</i>	7,7	7,64
<i>Cluster 2</i>	8,7	7,72

$$\text{Akurasi} = \frac{6,7+7,73}{6,7+7,73+8,7+7,72} = 0,47$$

$$\text{Sensifitas} = \frac{6,7}{6,7+7,72} = 0,46$$

$$\text{Spesifitas} = \frac{6,7}{6,7+8,7} = 0,43$$

Tabel 4. 37. Nilai Centroid Data PMB Periode 2016 – 2019 pada Algoritma *K-Medoids Clustering*.

Titik Pusat	Centroid Awal	Centroid Iterasi Terakhir
<i>Cluster 0</i>	6,7	5
<i>Cluster 1</i>	7,7	7,5
<i>Cluster 2</i>	8,7	7,9

$$\text{Akurasi} = \frac{6,7+5}{6,7+5+8,7+7,9} = 0,413$$

$$\text{Sensifitas} = \frac{6,7}{6,7+7,9} = 0,457$$

$$\text{Spesifitas} = \frac{6,7}{6,7+8,7} = 0,43$$

Dari hasil perhitungan nilai akurasi dari algoritma *K-Means Clustering* adalah 0,47, sedangkan nilai akurasi dari algoritma *K-Medoids Clustering* adalah 0,413. Nilai akurasi dari kedua algoritma ini berada pada kategori cukup baik sesuai tabel dibawah ini.

Tabel 4. 38. Nilai Rasio [20].

Nilai Rasio	Kategori
0,80 – 1,00	Sangat baik
0,60 – 0,80	Baik
0,40 – 0,60	Cukup baik
0,20 – 0,40	Kurang baik
0,00 – 0,20	Tidak baik

4.4.2. Perbandingan Hasil Perhitungan.

Berdasarkan dari hasil penelitian yang telah dilakukan pada data mahasiswa baru UNISNU Jepara periode 2016 - 2019. Selain menggunakan perhitungan manual dengan aplikasi *Microsoft Excel*, penelitian ini juga menggunakan aplikasi data mining *RapidMiner Studio*. Adapaun perbandingan hasil perhitungannya adalah sebagai berikut.

Tabel 4. 39. Hasil Perbandingan Perhitungan *Microsoft Excel* dan *RapidMiner*

Algoritma K-Means				
Perbandingan	C0	C1	C2	Total
<i>Microsoft Excel</i>	88	116	274	478
<i>RapidMiner</i>	265	137	76	478
selisih	177	21	198	396
Algoritma K-Medoids				
Perbandingan	C0	C1	C2	Total
Microsoft Excel	189	137	152	478
RapidMiner	270	100	108	478
selisih	81	37	44	162

Pada tabel 4.39 dapat dilihat terjadi adanya selisih hasil perhitungan dalam menggunakan *Microsoft Excel* dan *RapidMiner* di karenakan dalam menentukan *centroid* awal yang tidak sama. Menentukan *centroid* awal pada *Microsoft Excel*, peneliti menentukan secara acak. Sedangkan pada aplikasi *RapidMiner* dalam menentukan *centroid* awal sudah secara otomatis di lakukan oleh aplikasi.