

Implementasi Clustering K-Means Terhadap Pemilihan Konsentrasi Mahasiswa Menggunakan Bahasa R

Jimmie¹⁾, Zulhipni Reno Saputra Elsi²⁾, Dedi Haryanto³⁾

^{1,2,3}Departemen Teknologi Informasi, Universitas Muhammadiyah Palembang,
Jalan Jenderal Ahmad Yani 13 Ulu, Palembang 30263

e-mail: *jimmie@um-palembang.ac.id, Zulhipni_renosaputra@um-palembang.ac.id,
dedi_haryanto@um-palembang.ac.id

Abstrak

Beberapa mahasiswa dalam memilih konsentrasi sifatnya ikut-ikutan teman yang tidak memperhatikan kemampuan mereka pada bidang tersebut. Diharapkan dengan adanya penelitian ini akan terbentuk pola pengelompokan (clusterisasi) konsentrasi yang sesuai dengan minat dan kemampuan dari mahasiswa. Informasi pengelompokan ini dalam pemilihan konsentrasi mahasiswa menjadi salah satu informasi yang sangat penting, untuk mendapatkan informasi ini adalah dengan menggunakan konsep data mining. Data mining dikelompokkan berdasarkan tugas yang dilakukan, salah satunya adalah clustering. Clustering merupakan proses mempartisi sekumpulan objek data menjadi subset. Algoritma K-Means merupakan metode untuk melakukan pengelompokan objek ke sejumlah K klaster.

Kata kunci—Bahasa R, Konsentrasi, Algoritma K-Means, Clustering

Abstract

Some students in choosing concentration are like friends who don't pay attention to their abilities in that field. It is hoped that this research will form a concentration clustering pattern that is in accordance with the interests and abilities of the students. This grouping information in the selection of student concentration becomes one of the most important information, to get this information is to use the concept of data mining. Data mining is grouped based on the tasks performed, one of which is clustering. Clustering is the process of partitioning a set of data objects into subsets. The K-Means algorithm is a method for grouping objects into a number of K clusters.

Keywords— R Language, Concentration, K-Means Algorithm, Clustering

1. PENDAHULUAN

Pemanfaatan teknologi informasi merupakan suatu keharusan di era globalisasi saat ini. Dimana perkembangan teknologi informasi yang begitu pesat membawa perubahan di segala aspek kehidupan masyarakat. Perkembangan teknologi informasi mencakup kombinasi perangkat keras dan perangkat lunak yang digunakan untuk melakukan tugas-tugas penting yang dibutuhkan dan digunakan orang sehari-hari. Tanpa kita sadari kita sudah berada di era digital dimana teknologi diterapkan di berbagai sektor seperti aktivitas transaksi, pelayanan masyarakat, pendidikan, perbankan transportasi dan masih banyak lagi. Dibalik penggunaan teknologi informasi, terjadi pertukaran data yang sangat cepat. Dimana seluruh aktivitas terekam dan disimpan di dalam database. Data ini dapat diproses dan digunakan sebagai alat pengambilan keputusan.[1]

Seiring dengan perkembangan teknologi, banyak sekali data yang dapat dimanfaatkan dan diolah menjadi sebuah informasi, termasuk teknologi informasi yang dapat membantu mahasiswa dalam menentukan konsentrasi atau peminatan pada perguruan tinggi. Saat ini

sebagian besar mahasiswa masih belum tahu konsentrasi apa yang akan dipilih dikarenakan mereka belum mengetahui kemampuan ataupun minat yang akan mereka pilih. Pemilihan konsentrasi sangat penting dan perlu dipertimbangkan secara matang untuk menentukan kompetensi apa yang akan mereka miliki pada saat mereka telah menyelesaikan perkuliahan. Beberapa kasus yang sering terjadi mahasiswa memilih konsentrasi yang kurang tepat, sehingga menyebabkan mahasiswa kesulitan dalam mengikuti dan memahami materi-materi yang diberikan, dan juga hal ini mempengaruhi proses pembuatan tugas akhir dimana hasil yang didapat kurang maksimal dikarenakan kurang tepatnya di dalam pemilihan konsentrasi.[1][2]

Program Studi Teknologi Informasi di bawah Fakultas Teknik Universitas Muhammadiyah Palembang yang berdiri pada tahun 2018. Program Studi Teknologi Informasi berada di Kampus B beralamat di Jalan Banten Plaju Palembang. Program Studi Teknologi Informasi Fakultas Teknik Universitas Muhammadiyah Palembang memiliki beberapa konsentrasi, untuk saat ini konsentrasi yang dipilih oleh mahasiswa yaitu Rekayasa Perangkat Lunak atau *Software Engineering* dan Tata Kelola Teknologi Informasi. Kedua konsentrasi ini dari angkatan pertama selalu dipilih oleh mahasiswa.

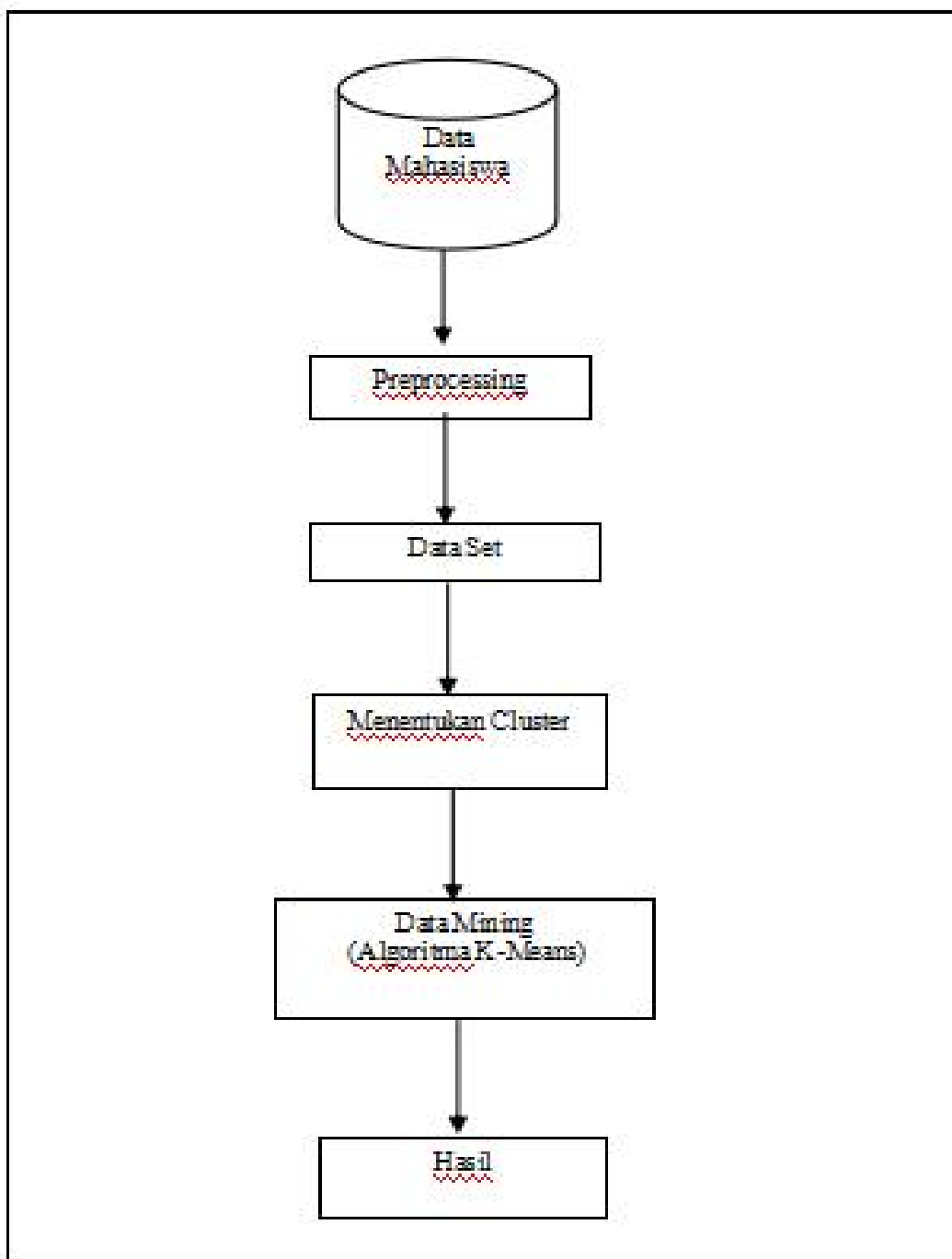
Observasi yang dilakukan di Program Studi Teknologi Informasi untuk menentukan pemilihan konsentrasi mahasiswa masih berfokus pada mahasiswa itu sendiri dalam pemilihan konsentrasi dengan berkonsultasi dengan pembimbing akademik. Beberapa mahasiswa dalam memilih konsentrasi sifatnya ikut-ikutan teman yang tidak memperhatikan kemampuan mereka pada bidang tersebut. Diharapkan dengan adanya penelitian ini akan terbentuk pola pengelompokan (*clusterisasi*) konsentrasi yang sesuai dengan minat dan kemampuan dari mahasiswa. Informasi pengelompokan ini dalam pemilihan konsentrasi mahasiswa menjadi salah satu informasi yang sangat penting, untuk mendapatkan informasi ini adalah dengan menggunakan konsep data mining.

Data mining adalah analisis kumpulan data observasional untuk menemukan hubungan yang tidak terduga dan untuk meringkas data dengan cara baru yang dapat dimengerti dan berguna bagi pemiliki data. Data mining dikelompokkan berdasarkan tugas yang dilakukan, salah satunya adalah *clustering*. *Clustering* merupakan proses mempartisi sekumpulan objek data menjadi subset. Subset adalah *cluster*, sehingga objek dalam *cluster* sama satu sama lain, namun berbeda dengan objek di *cluster* lain.[3]. Algoritma *K-Means* merupakan metode untuk melakukan pengelompokan objek ke sejumlah K klaster. Metode ini adalah metode pengelompokan data yang sederhana dan banyak digunakan dalam aplikasi data mining. Input dari *K-Means clustering* adalah data atau objek dan jumlah klaster (k) yang diinginkan. Setiap klaster direpresentasikan oleh sebuah titik pusat (centroid). Setiap data akan dikelompokkan pada klaster dengan titik pusat yang terdekat dari data tersebut.[4][5] Algoritma *K-Means* juga digunakan untuk pengelompokan data (*cluster*), hasil dipenelitian Puji Subekti diperoleh hasil bahwa *cluster* 1 terdapat 3 mahasiswa yang masuk di konsentrasi sistem cerdas. *Cluster* ke 2 terdapat 20 mahasiswa yang masuk di konsentrasi Multimedia dan Games, *cluster* ke 3 terdapat 56 mahasiswa yang masuk dikonsentrasi jaringan.[6]

2. METODE PENELITIAN

Tahapan yang akan dilakukan dalam menentukan pemilihan konsentrasi oleh mahasiswa pada program studi Teknologi Informasi sehingga pilihan konsentrasi sesuai dengan kemampuan dari mahasiswa tersebut. Persiapan pertama adalah mempersiapkan data mahasiswa, hal ini penting dilakukan karena dataset yang akan digunakan memiliki tipe data yang sama yaitu berupa tipe data numerik. Pada tahapan berikutnya adalah melakukan *preprocessing*, dimana proses ini dilakukan untuk merubah data mentah kedalam format yang dapat dimengerti. Data mentah ini biasanya tidak lengkap (*missing value*) tidak konsisten dan memiliki tipe data yang tidak sama. Selanjutnya adalah membaca dataset, kemudian dilakukan pengklasteran data mahasiswa, setelah itu dilakukan eksplorasi data dengan menggunakan fungsi algoritma *K-Means* dan tahap terakhir adalah melakukan analisa hasil pengelompokan. Hasil analisa

pengelompokan nantinya dapat diketahui seperti apa kedekatan tiap titik data dari tiap kelompok yang terbentuk sehingga bisa ditentukan jumlah kluster yang optimal. Gambar pemodelan sistem seperti di bawah ini.



Gambar 1. Pemodelan Sistem[7]

Dataset Mahasiswa

Dataset mahasiswa yang digunakan pada penelitian ini adalah data mahasiswa angkatan 2020 program studi Teknologi Informasi yang berjumlah 25 mahasiswa, variable matakuliah yang diambil yaitu matakuliah semester 1 sampai dengan semester 4.

Tabel 1. Dataset Mahasiswa

No.	NIM	Nama Mahasiswa	Basis Data	PBO	TKTI	PWD	JARINGAN	RPL	STD
1	16202001	Verent A. Wijaya	82	80	82	80	82	82	80
2	16202002	Salsabilah	80	80	83	82	83	82	81
3	16202003	Yogi Candra Pratama	79	76	78	79	80	79	79
4	16202004	Dwiki Ilham	78	76	80	80	81	79	77
5	16202005	Triono	80	79	80	82	80	81	80
6	16202006	Nova Aprianto	80	78	81	81	81	80	79
7	16202007	Zulfa Mukholadun	77	76	80	79	80	78	77
8	16202008	M. Ikhsan	80	78	81	82	80	80	79
9	16202009	Acok	65	60	79	78	76	68	65
10	16202010	Novan Maulana	80	80	82	80	80	81	80
11	16202011	Pirman Saputra	80	81	81	83	83	81	82
12	16202012	Piska Eka Oktariani	80	80	83	84	81	80	81
13	16202013	Muhammad Yoga W N	80	79	80	80	80	81	80
14	16202014	Janu Pratama	79	76	78	80	80	79	78
15	16202015	Anggun Carmenta	79	77	79	81	78	78	79
16	16202016	Rahmat Sodikin	80	79	80	80	80	82	80
17	16202017	Muhammad Nur F R	60	60	65	79	78	77	67
18	16202018	Meidy Saputra	80	80	82	84	82	83	80
19	16202019	Riski Ananda	80	81	83	81	83	81	81
20	16202020	Moehammad I	81	80	80	81	81	82	81
21	16202021	Tomi Setiawan	81	80	81	80	80	81	82
22	16202022	Naufal Zacky Anwar	80	76	80	81	81	81	77
23	16202023	Stefen Alvayen	80	80	82	82	82	81	80
24	16202024	Ragita Cahyantika	81	81	83	83	82	82	81
25	16202025	Rihhadatul Nabilah	81	79	80	82	80	80	80

Dari tabel 1 di atas adalah persiapan awal untuk pengolahan data pemilihan konsentrasi mahasiswa yaitu data nilai mahasiswa. Pembersihan data dengan menghapus data yang tidak dibutuhkan dalam penelitian ini, disebabkan karena tabel mahasiswa terdiri dari informasi secara keseluruhan.

Persiapan Data

Untuk proses awal adalah membaca dataset mahasiswa dengan menggunakan bahasa R. Dataset mahasiswa yang digunakan pada penelitian ini dengan format excel. Hasil dari import file excel dengan bahasa R seperti pada tabel 2.

Tabel 2. Hasil Proses Import Data Bahasa R

No.	NIM	Nama.Mahasiswa	Basis.Data	PBO	TKTI	PWD	JARINGAN	RPL	STD
1	16202001	Verent Agnes wijaya	82	80	82	80	82	82	80
2	16202002	Salsabilah	80	80	83	82	83	82	81
3	16202003	Yogi Candra Pratama	79	76	78	79	80	79	79
4	16202004	Dwiki Ilham	78	76	80	80	81	79	77
5	16202005	Triono	80	79	80	82	80	81	80
6	16202006	Nova Aprianto	80	78	81	81	81	80	79
7	16202007	zulfa Mukholadun	77	76	80	79	80	78	77
8	16202008	M. Ikhsan	80	78	81	82	80	80	79
9	16202009	Acok	65	60	79	78	76	68	65
10	16202010	Novan Maulana	80	80	82	80	80	81	80
11	16202011	Pirman Saputra	80	81	81	83	83	81	82
12	16202012	Piska Eka Oktariani	80	80	83	84	81	80	81
13	16202013	Muhammad Yoga w N	80	79	80	80	80	81	80
14	16202014	Janu Pratama	79	76	78	80	80	79	78
15	16202015	Anggun Carmenta	79	77	79	81	78	78	79
16	16202016	Rahmat Sodikin	80	79	80	80	80	82	80
17	16202017	Muhammad Nur F R	60	60	65	79	78	77	67
18	16202018	Meidy Saputra	80	80	82	84	82	83	80
19	16202019	Riski Ananda	80	81	83	81	83	81	81
20	16202020	Moehammad Irfansyah	81	80	80	81	81	82	81
21	16202021	Tomi Setiawan	81	80	81	80	80	81	82
22	16202022	Naufal zacky Anwar	80	76	80	81	81	81	77
23	16202023	Stefen Alvayen	80	80	82	82	82	81	80
24	16202024	Ragita Cahyantika	81	81	83	83	82	82	81
25	16202025	Rihhadatul Nabilah	81	79	80	82	80	80	80

Pada tabel 2 di atas merupakan hasil *import* data excel dengan menggunakan bahasa R, dengan perintah `data=read.delim("clipboard")`. Tahap selanjutnya adalah melakukan *summary* data, yang berfungsi untuk melihat deskripsi data apakah data yang diimportkan sudah lengkap apa belum. Hasil *summary* bisa dilihat pada gambar 2. di bawah ini.

No.	NIM	Nama.Mahasiswa	Basis.Data
Min. : 1	Min. :16202001	Length:25	Min. :60.00
1st Qu. : 7	1st Qu. :16202007	Class :character	1st Qu. :79.00
Median :13	Median :16202013	Mode :character	Median :80.00
Mean :13	Mean :16202013		Mean :78.52
3rd Qu. :19	3rd Qu. :16202019		3rd Qu. :80.00
Max. :25	Max. :16202025		Max. :82.00
		PBO	TKTI
Min. :60.00	Min. :65.00	Min. :78.00	Min. :76.00
1st Qu. :76.00	1st Qu. :80.00	1st Qu. :80.00	1st Qu. :80.00
Median :79.00	Median :80.00	Median :81.00	Median :80.00
Mean :77.28	Mean :80.12	Mean :80.96	Mean :80.56
3rd Qu. :80.00	3rd Qu. :82.00	3rd Qu. :82.00	3rd Qu. :82.00
Max. :81.00	Max. :83.00	Max. :84.00	Max. :83.00
		PWD	JARINGAN
Min. :68.00	Min. :65.00		
1st Qu. :79.00	1st Qu. :79.00		
Median :81.00	Median :80.00		
Mean :79.96	Mean :78.64		
3rd Qu. :81.00	3rd Qu. :81.00		
Max. :83.00	Max. :82.00		

Gambar 2. Summary Data

Setelah proses *summary* data, selanjutnya dilakukan perintah struktur data, perintah struktur data berfungsi untuk mengecek type data dari dataset yang dibuat seperti pada gambar 5. di bawah ini.

```
'data.frame': 25 obs. of 10 variables:
 $ No. : int 1 2 3 4 5 6 7 8 9 10 ...
 $ NIM : int 16202001 16202002 16202003 16202004 16202005 16202006 16202007 16202008 16202009 16202010 ...
 $ Nama.Mahasiswa: chr "Verent Agnes wijaya" "Salsabilah" "Yogi Candra Pratama" "Dwiki Ilham" ...
 $ Basis.Data : int 82 80 79 78 80 80 77 80 65 80 ...
 $ PBO : int 80 80 76 76 79 78 76 78 60 80 ...
 $ TKTI : int 82 83 78 80 80 81 80 81 79 82 ...
 $ PWD : int 80 82 79 80 82 81 79 82 78 80 ...
 $ JARINGAN : int 82 83 80 81 80 81 80 80 76 80 ...
 $ RPL : int 82 82 79 79 81 80 78 80 68 81 ...
 $ STD : int 80 81 79 77 80 79 77 79 65 80 ...
```

Gambar 3. Hasil pengecekan struktur data

3. HASIL DAN PEMBAHASAN

3.1 Analisis Sistem

Proses standarisasi dan normalisasi data berfungsi untuk mengambil data numerik yang akan digunakan untuk pengolahan data, sedangkan data yang ada masing mengandung charater. Setelah data menjadi numerik langkah selanjut adalah membuat skala nilai pada data yang akan diolah. Seperti pada tabel 3 di bawah ini.

Tabel 3. Proses standarisasi dan normalisasi

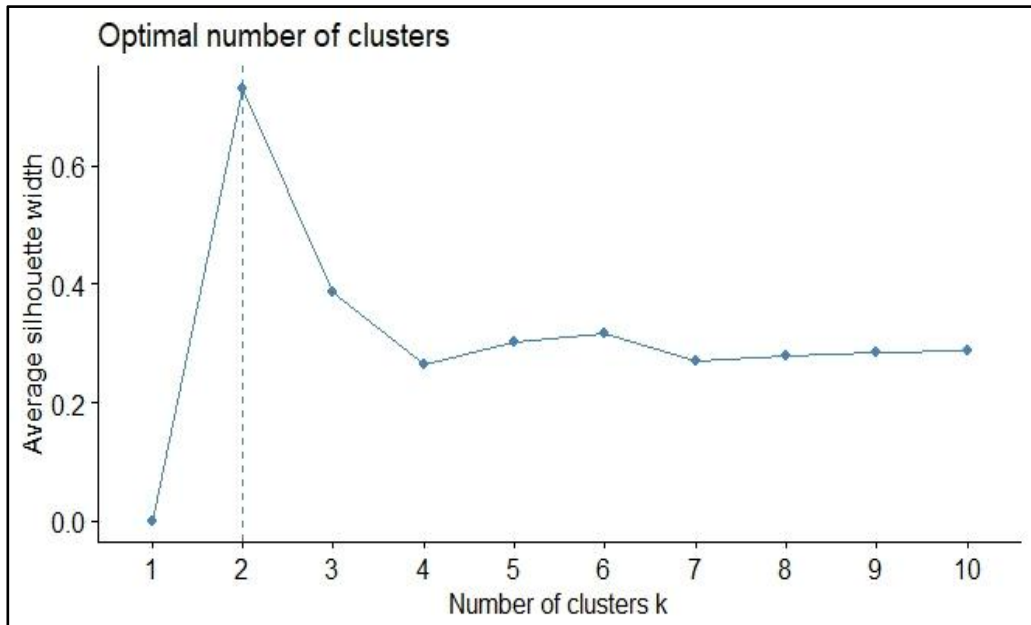
	Basis. Data	PBO	TKTI	PWD	JARINGAN	RPL	STD
1	82	80	82	80	82	82	80
2	80	80	83	82	83	82	81
3	79	76	78	79	80	79	79
4	78	76	80	80	81	79	77
5	80	79	80	82	80	81	80
6	80	78	81	81	81	80	79
7	77	76	80	79	80	78	77
8	80	78	81	82	80	80	79
9	65	60	79	78	76	68	65
10	80	80	82	80	80	81	80
11	80	81	81	83	83	81	82
12	80	80	83	84	81	80	81
13	80	79	80	80	80	81	80
14	79	76	78	80	80	79	78
15	79	77	79	81	78	78	79
16	80	79	80	80	80	82	80
17	60	60	65	79	78	77	67
18	80	80	82	84	82	83	80
19	80	81	83	81	83	81	81
20	81	80	80	81	81	82	81
21	81	80	81	80	80	81	82
22	80	76	80	81	81	81	77
23	80	80	82	82	82	81	80
24	81	81	83	83	82	82	81
25	81	79	80	82	80	80	80

Setelah dilakukan proses standarisasi dan normalisasi data yang mana berfungsi untuk mengambil data numerik yang akan digunakan untuk pengolahan data, sehingga dihasilkan penskalaan nilai, seperti tabel 4 di bawah ini:

Tabel.4 Penskalaan data

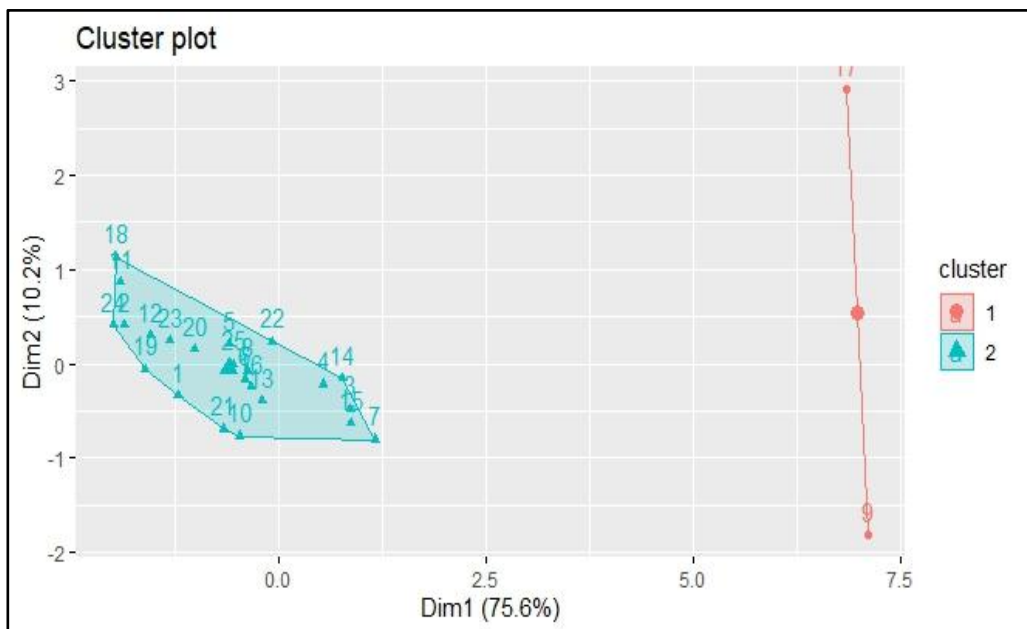
	Basis.Data	PBO	TKTI	PWD	JARINGAN	RPL	STD
1	0.69936505	0.49693318	0.54023881	-0.61248875	0.8807161	0.70498679	0.33491036
2	0.29743111	0.49693318	0.82759987	0.66352948	1.4923245	0.70498679	0.58116798
3	0.09646414	-0.23385091	-0.60920546	-1.25049787	-0.3425007	-0.33175849	0.08865274
4	-0.10450282	-0.23385091	-0.03448333	-0.61248875	0.2691077	-0.33175849	-0.40386249
5	0.29743111	0.31423716	-0.03448333	0.66352948	-0.3425007	0.35940503	0.33491036
6	0.29743111	0.13154114	0.25267774	0.02552036	0.2691077	0.01382327	0.08865274
7	-0.30546979	-0.23385091	-0.03448333	-1.25049787	-0.3425007	-0.67734025	-0.40386249
8	0.29743111	0.13154114	0.25267774	0.66352948	-0.3425007	0.01382327	0.08865274
9	-2.71707340	-3.15698729	-0.32184440	-1.88850698	-2.7889343	-4.13315783	-3.35895389
10	0.29743111	0.49693318	0.54023881	-0.61248875	-0.3425007	0.35940503	0.33491036
11	0.29743111	0.67962921	0.25267774	1.30153859	1.4923245	0.35940503	0.82742559
12	0.29743111	0.49693318	0.82759987	1.93954771	0.2691077	0.01382327	0.58116798
13	0.29743111	0.31423716	-0.03448333	-0.61248875	-0.3425007	0.35940503	0.33491036
14	0.09646414	-0.23385091	-0.60920546	-0.61248875	-0.3425007	-0.33175849	-0.15760487
15	0.09646414	-0.05115489	-0.32184440	0.02552036	-1.5657175	-0.67734025	0.08865274
16	0.29743111	0.31423716	-0.03448333	-0.61248875	-0.3425007	0.70498679	0.33491036
17	-3.72190824	-3.15698729	-4.34489934	-1.25049787	-1.5657175	-1.02292200	-2.86643866
18	0.29743111	0.49693318	0.54023881	1.93954771	0.8807161	1.05056854	0.33491036
19	0.29743111	0.67962921	0.82759987	0.02552036	1.4923245	0.35940503	0.58116798
20	0.49639808	0.49693318	-0.03448333	0.02552036	0.2691077	0.70498679	0.58116798
21	0.49639808	0.49693318	0.25267774	-0.61248875	-0.3425007	0.35940503	0.82742559
22	0.29743111	-0.23385091	-0.03448333	0.02552036	0.2691077	0.35940503	-0.40386249
23	0.29743111	0.49693318	0.54023881	0.66352948	0.8807161	0.35940503	0.33491036
24	0.49639808	0.67962921	0.82759987	1.30153859	0.8807161	0.70498679	0.58116798
25	0.49639808	0.31423716	-0.03448333	0.66352948	-0.3425007	0.01382327	0.33491036

Tahap selanjutnya apabila sudah dilakukan standarisasi dan normalisasi serta membuat skala nilai, yaitu melakukan clustering atau klasifikasi dengan menggunakan metode silhouette untuk menentukan nilai K dimana pada grafik ditunjukkan untuk nilai tertinggi yaitu 2.



Gambar 4. Penentuan Nilai K dengan Metode Silhouette

Untuk menentukan jumlah cluster terbaik dengan metode silhouette dilihat nilai rata-rata yang tertinggi yaitu jumlah cluster yang optimal, berdasarkan metode silhouette cluster terbaik berjumlah 2.



Gambar 5. Grafik Hasil Clustering

Gambar 5 di atas merupakan grafik hasil *clustering* pemilihan konsentrasi pada mahasiswa Program Studi Teknologi Informasi, dari grafik di atas terdapat 2 cluster.

Tabel.5 Hasil Clustering berbentuk tabel

	Basis.Data	PBO	TKTI	PWD	JARINGAN	RPL	STD	final.cluster
1	0.69936505	0.49693318	0.54023661	-0.61248675	0.8807161	0.70498679	0.33491036	2
2	0.29743111	0.49693318	0.82759967	0.66352948	1.4923245	0.70498679	0.58116798	2
3	0.09646414	-0.23385091	-0.60920546	-1.25049787	-0.3425007	-0.33175849	0.08865274	2
4	-0.10450282	-0.23385091	-0.03446333	-0.61248675	0.2691077	-0.33175849	-0.40386249	2
5	0.29743111	0.31423716	-0.03446333	0.66352948	-0.3425007	0.35940503	0.33491036	2
6	0.29743111	0.13154114	0.25287774	0.02552036	0.2691077	0.01382327	0.08865274	2
7	-0.30546979	-0.23385091	-0.03446333	-1.25049787	-0.3425007	-0.67734025	-0.40386249	2
8	0.29743111	0.13154114	0.25287774	0.66352948	-0.3425007	0.01382327	0.08865274	2
9	-2.71707340	-3.15698729	-0.32184440	-1.88850698	-2.7889343	-4.13315783	-3.35895389	1
10	0.29743111	0.49693318	0.54023661	-0.61248675	-0.3425007	0.35940503	0.33491036	2
11	0.29743111	0.67962921	0.25287774	1.30153859	1.4923245	0.35940503	0.82742559	2
12	0.29743111	0.49693318	0.82759967	1.93954771	0.2691077	0.01382327	0.58116798	2
13	0.29743111	0.31423716	-0.03446333	-0.61248675	-0.3425007	0.35940503	0.33491036	2
14	0.09646414	-0.23385091	-0.60920546	-0.61248675	-0.3425007	-0.33175849	-0.15760467	2
15	0.09646414	-0.05115489	-0.32184440	0.02552036	-1.5657175	-0.67734025	0.08865274	2
16	0.29743111	0.31423716	-0.03446333	-0.61248675	-0.3425007	0.70498679	0.33491036	2
17	-3.72190824	-3.15698729	-4.34489934	-1.25049787	-1.5657175	-1.02292200	-2.86643866	1
18	0.29743111	0.49693318	0.54023661	1.93954771	0.8807161	1.05056854	0.33491036	2
19	0.29743111	0.67962921	0.82759967	0.02552036	1.4923245	0.35940503	0.58116798	2
20	0.49839808	0.49693318	-0.03446333	0.02552036	0.2691077	0.70498679	0.58116798	2
21	0.49839808	0.49693318	0.25287774	-0.61248675	-0.3425007	0.35940503	0.82742559	2
22	0.29743111	-0.23385091	-0.03446333	0.02552036	0.2691077	0.35940503	-0.40386249	2
23	0.29743111	0.49693318	0.54023661	0.66352948	0.8807161	0.35940503	0.33491036	2
24	0.49839808	0.67962921	0.82759967	1.30153859	0.8807161	0.70498679	0.58116798	2
25	0.49839808	0.31423716	-0.03446333	0.66352948	-0.3425007	0.01382327	0.33491036	2

Hasil clustering dalam bentuk tabel dapat dijelaskan bahwa mahasiswa dengan nomor urut 1 sampai dengan 8 berada di cluster 2, mahasiswa dengan nomor urut 9 berada di cluster 1. Dari nomor urut 10 sampai 16 berada di cluster 2, nomor urut 17 berada di cluster 1. Selanjutnya dari nomor urut 18 sampai dengan 25 berada di cluster 2. Dari hasil clustering di atas bahwa dari 25 mahasiswa paling banyak berada di cluster 2, dan hanya 2 orang mahasiswa yang berada di cluster 1.

```

K-means clustering with 2 clusters of sizes 2, 23

Cluster means:
 Basis.Data      PBO      TKTI      PWD      JARINGAN      RPL
 1 -3.2194908 -3.1569873 -2.3333719 -1.5695024 -2.1773259 -2.5780399
 2  0.2799557  0.2745206  0.2029019  0.1364785  0.1893327  0.2241774
      STD
 1 -3.1126963
 2  0.2706692

Clustering vector:
 [1] 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2

within cluster sum of squares by cluster:
 [1] 14.50706 46.28518
 (between_SS / total_SS = 63.8 %)

Available components:
 [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
 [6] "betweenss"    "size"         "iter"         "ifault"
    
```

Gambar 6. K-Means Clustering dengan 2 Cluster

Pada gambar di atas dapat kita lihat hasil k-means clustering dengan 2 cluster dengan ukuran 2, 23, dengan rata-rata atributnya masih menggunakan data hasil transformasi dimana nilai rata-rata 7 atribut yaitu Matakuliah Basis data, PBO, TKTI, PWD, Jaringan, RPL dan STD.

Cluster	Basis.Data	PBO	TKTI	PWD	JARINGAN	RPL	STD	
<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	
1	1	62.5	60	72	78.5	77	72.5	66
2	2	79.9	78.8	80.8	81.2	80.9	80.6	79.7

Gambar 7. K-Means Clustering

Dari gambar di atas dimana ada 2 cluster, pada cluster 1 nilai rata-rata untuk matakuliah basis data nilai rata-ratanya 62.5, pbo 60, TKTI 72, PWD dengan nilai rata-rata 78.5, jaringan 77, RPL 72.5 dan STD 66. Untuk cluster 2 nilai rata-rata untuk matakuliah Basis Data 79.9, PBO 78.8, TKTI 80.8, nilai rata PWD 81.2, matakuliah jaringan 80.9, RPL 80.6, matakuliah STD 79.7. dari 2 cluster tersebut, nilai rata-rata tertinggi terdapat pada cluster 2 dibandingkan dengan cluster 1.

4. KESIMPULAN

Berdasarkan dari hasil penelitian yang telah dibuat, dapat disimpulkan bahwa dari hasil analisis clustering didapat nilai K yaitu 2 dengan menggunakan metode silhouette untuk mendapatkan nilai rata-rata yang tertinggi. Dari 2 cluster tersebut, nilai rata-rata tertinggi terdapat pada cluster 2 dibandingkan dengan cluster 1. Dari 25 mahasiswa yang diteliti nilai rata-rata matakuliah Pemrograman Web Dinamis (PWD) dengan nilai rata-rata tertinggi yaitu 81.2.

5. SARAN

Pada penelitian ini saran yang dapat disampaikan bahwa di dalam menentukan pemilihan konsentrasi harus berdasarkan nilai pengelompokkan pada matakuliah yang diambil berdasarkan konsentrasi yang akan dipilih, sehingga pemilihan konsentrasi oleh mahasiswa sesuai dengan kemampuan yang mereka punya.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Ketua Program Studi Teknologi Informasi yang telah banyak membantu di dalam memberikan informasi yang dibutuhkan oleh peneliti.

DAFTAR PUSTAKA

- [1] P. Algoritma, K. D. A. N. Algoritma, M. G. Ghazali, N. M. Saraswati, R. Cipta, and S. Hariyono, "Bayes Dalam Pemilihan Konsentrasi Jurusan Siswa Di Sma Islam Ta ' Alumul Huda Bumiayu," vol. 2, no. 2, pp. 19–25, 2021.
- [2] J. Aranda and W. A. G. Natasya, "Penerapan Metode K-Means Cluster Analysis Pada Sistem Pendukung Keputusan Pemilihan Konsentrasi Untuk Mahasiswa International Class STMIK Amikom Yogyakarta," *Semnasteknomedia Online*, vol. 4, no. 1, pp. 4.2-1-4.2-6, 2016.
- [3] W. Sudrajat *et al.*, "Penerapan Algoritma K-Means Clustering untuk Pengelompokan UMKM Menggunakan Rapidminer," pp. 27–36.
- [4] siti mariyah setia pramana, budi yuniarto, *Data mining dengan R Konsep serta implementasi*. jakarta: in media, 2018.

- [5] A. Asroni and R. Adrian, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *Semesta Tek.*, vol. 18, no. 1, pp. 76–82, 2016.
- [6] P. Subekti, T. D. Andini, and M. Islamiyah, "Sistem Penentuan Konsentrasi Jurusan Bagi Mahasiswa Informatika Menggunakan Metode K-Means Di Institut Asia Malang Determination System for Department Concentration for Informatics Students Using the K-Means Method at the Institute of Asia Malang," vol. 12, no. April, pp. 25–39, 2022.
- [7] R. P. Primanda, A. Alwi, and D. Mustikasari, "DATA MINING SELEKSI SISWA BERPRESTASI UNTUK MENENTUKAN KELAS UNGGULAN MENGGUNAKAN METODE K-MEANS CLUSTERING (Studi Kasus di MTS Darul Fikri)," *Komputek*, vol. 5, no. 1, p. 88, 2021.