
Penerapan Metode *Clustering* Dengan Algoritma *K-Means* Untuk Rekomendasi Pemilihan Jalur Peminatan Sesuai Kemampuan Pada Progam Studi Teknik Informatika - S1 Universitas Dian Nuswantoro

Application of Clustering Method With K-Means Algorithm For Specialisation Line Selection Recommendations In accordance Capability Study Program Informatics Engineering - S1 University Dian Nuswantoro

Rizky Adrianto¹, Amiq Fahmi²

^{1,2}Program Studi Sistem Informasi, Fakultas Ilmu Komputer,
Universitas Dian Nuswantoro

Jl. Nakula I No. 5-11 Semarang, Indonesia

E-mail : 112201204593@mhs.dinus.ac.id¹, amiq.fahmi@dsn.dinus.ac.id²

Abstrak

Terdapat 2 pilihan jalur peminatan pada program studi Teknik Informatika-S1 Universitas Dian Nuswantoro, dimana mahasiswa kesulitan dalam menentukan pilihan peminatan yang sesuai dengan kemampuan akademisnya. Disamping itu belum dimanfaatkannya data mahasiswa yang sudah mengambil peminatan, yang dapat dijadikan suatu informasi sehingga dapat berguna sebagai rekomendasi dalam pemilihan peminatan. Oleh karena itu penelitian ini memiliki tujuan menerapkan salah satu metode dalam data mining, yaitu metode clustering dengan algoritma K-Means. Untuk dapat mengelompokkan mahasiswa berdasarkan kemampuan akademisnya, sehingga dapat menjadi salah satu alternatif untuk program studi dalam memberikan rekomendasi kepada mahasiswa mengenai pemilihan jalur peminatan. Pada penelitian ini menggunakan data mahasiswa Teknik Informatika-S1 Universitas Dian Nuswantoro angkatan 2011, 2012, dan 2013. Atribut yang digunakan dalam pengelompokan ialah IPS 1-4, rerata nilai matakuliah penunjang SC, dan rerata nilai matakuliah penunjang RPLD. Sedangkan kesesuaian kemampuan mahasiswa dengan jalur peminatan akan diketahui dari atribut peminatan, rerata nilai matakuliah peminatan SC, rerata nilai matakuliah peminatan RPLD. Tools yang digunakan yaitu rapidminer dan MS.Excel, metode pengembangan data mining mengikuti fase-fase dari CRISP-DM. Hasil pada penelitian ini yaitu didapatkan 5 kelompok mahasiswa berdasarkan tingkat kemampuan akademisnya. Yang mana pada setiap kelompok yang terbentuk kemudian dilakukan analisa lebih lanjut dan profilisasi guna diketahui karakteristik dan kesesuaiannya dengan jalur peminatan.

Kata Kunci: Peminatan, Rekomendasi, Data Mining, Clustering, K-Means

Abstract

There are two route of specialization in Department of Informatics Engineering Dian Nuswantoro University, where the student is difficult to determine the choice of specialization in accordance with academic skills. In addition, the student data who has taken specialization has not been utilized, which can be used as an information that can be useful as a recommendation in the choice of specialization. Therefore, this study has the purpose of implementing one of the methods in datamining is the method of clustering with k-means algorithm. To be able to classify students based on academic

ability, so it can be an alternative for the study program in providing recommendations to students on specialization path selection. In this study, using the student data from Bachelor Degree of Informatics Engineering Dian Nuswantoro University class of 2011, 2012, and 2013. The attribute used in the grouping is IPS 1-4, rerata nilai matakuliah penunjang SC, rerata nilai matakuliah penunjang RPLD. While conformity ability of students with specialization pathway will be known from the attributes peminatan, rerata nilai matakuliah peminatan SC, rerata nilai matakuliah peminatan RPLD. Tools used are rapidminer and MS.Excel, data mining development methods follow the phases of CRISP-DM. The results of this study are obtained 5 groups of students based on their level academic skills. In which each group is formed, then conducted further analysis and profiling to known characteristics and compatibility with specialization pathways.

Keywords: Recommendation, Data Mining, Clustering, K-Means

1. PENDAHULUAN

Peminatan atau konsentrasi merupakan fokus mahasiswa terhadap suatu bidang studi tertentu sesuai dengan minatnya. Tujuannya yaitu untuk lebih memfokuskan mahasiswa terhadap ilmu yang didapat dari matakuliah sebelumnya, agar dapat lebih terarah. Disamping itu juga peminatan bertujuan mengarahkan para mahasiswa, dalam memberikan beberapa pilihan mengenai topik penelitian tugas akhir. Berdasarkan kurikulum 2013, program studi Teknik Informatika-S1 Universitas Dian Nuswantoro menerapkan sistem pembelajaran dengan peminatan kepada mahasiswa semester ke-lima. Peminatan pada program studi Teknik Informatika-S1 Universitas Dian Nuswantoro dapat dikategorikan menjadi dua pilihan yaitu peminatan Sistem Cerdas (SC) dan peminatan Rekayasa Perangkat Lunak dan Data (RPLD) [1].

Kemampuan akademis ialah kemampuan seorang individu dalam bidang akademik yang ukurannya dapat berupa nilai ataupun yang sering disebut prestasi akademik. Dalam hal ini kemampuan akademis sebelum pemilihan jalur peminatan dapat diketahui dari IPS 1-4 dan nilai yang diperoleh dari beberapa matakuliah penunjang / penguatan yaitu beberapa matakuliah yang terdapat pada semester 1-4 yang memiliki keterkaitan dengan matakuliah yang ditawarkan pada masing-masing jalur peminatan. Permasalahan yang dihadapi mahasiswa yang akan memilih peminatan adalah kebingungan dan kesulitannya dalam menentukan pilihan yang sesuai dengan kemampuan akademisnya. Oleh karena itu pemilihan jalur peminatan terkadang merupakan ajang untuk mengekor teman [14], atau dengan kata lain memilih hanya berdasarkan rekomendasi dari teman dan tidak berdasarkan kemampuan yang dimilikinya. Pemilihan peminatan tanpa adanya pertimbangan matang terlebih dahulu yang sesuai dengan kemampuan akademis, akan memiliki dampak yaitu nilai-nilai yang diperoleh akan kurang memuaskan pada matakuliah peminatan yang dipilihnya. Dan juga mahasiswa tersebut akan kesulitan dalam penyerapan materi perkuliahan pada peminatan yang dipilihnya [15]. Untuk itu perlu adanya suatu rekomendasi dalam memilih jalur peminatan yang tepat dan sesuai dengan tingkat kemampuan akademis dari para mahasiswa.

Data akademik pada Universitas Dian Nuswantoro khususnya untuk data mahasiswa dan transkrip nilai mahasiswa, selama ini hanya disimpan dan dikumpulkan dalam *database*. Data-data tersebut akan selalu bertambah dan menumpuk dalam jumlah yang sangat besar. Namun sangat disayangkan karena data tersebut tidak dimanfaatkan secara maksimal. Padahal data tersebut dapat diolah lebih lanjut sehingga dapat menjadi suatu informasi dan pengetahuan baru yang berguna dan menarik. Salah satu teknik yang dapat dilakukan untuk mengolah data berukuran besar yaitu dengan menggunakan *data mining*. *Data mining* ialah proses dengan menggunakan teknik kecerdasan buatan, *machine learning*, matematika, dan statistik dalam mengidentifikasi dan mengekstraksi informasi dan pengetahuan yang bermanfaat dalam berbagai *database* besar [12]. Diharapkan dengan menggunakan *data mining* dapat memberikan pengetahuan dan informasi yang sebelumnya tersembunyi di dalam gudang data sehingga

dapat menjadi informasi atau pengetahuan menarik dan berguna. Menurut [11] menyatakan bahwa “*clustering* membagi dataset ke dalam kelompok yang berbeda, proses pengelompokan objek ke dalam kelompok tersebut bahwa objek dalam kelompok yang sama memiliki kemiripan dan objek dalam kelompok yang berbeda tidak memiliki kemiripan”. Salah satu algoritma yang dapat digunakan dalam melakukan *clustering* ialah algoritma *K-Means*

Karena semakin bertambahnya mahasiswa pada program studi Teknik Informatika-S1 Universitas Dian Nuswantoro yang akan menentukan pemilihan jalur peminatan, dengan memanfaatkan teknik *data mining* diharapkan dapat memberikan informasi yang menarik dan berguna mengenai pemilihan peminatan untuk mahasiswa yang tepat dan sesuai kemampuannya. Adapun tujuan dari penelitian ini ialah menerapkan *data mining* dengan metode *clustering* menggunakan algoritma *K-Means* dalam melakukan pengelompokan terhadap data mahasiswa program studi Teknik Informatika-S1 Universitas Dian Nuswantoro berdasarkan dari kemampuan akademisnya, sehingga dapat memberikan informasi yang baru dan berguna sebagai rekomendasi dalam pemilihan jalur peminatan. Data yang digunakan ialah data mahasiswa, KHS, dan transkrip nilai mahasiswa Teknik Informatika-S1 Universitas Dian Nuswantoro angkatan 2011, 2012, dan 2013. Atribut yang digunakan yaitu NIM, IPK, Indeks Prestasi (IP) Semester 1-4, rerata nilai matakuliah penunjang, rerata nilai matakuliah peminatan, dan peminatan yang dipilih. Nantinya hasil pengelompokan akan dilakukan profilisasi untuk didapatkan informasi yang baru mengenai pemilihan peminatan. Informasi tersebut dapat digunakan oleh program studi sebagai salah satu alternatif dalam memberikan rekomendasi kepada mahasiswa, mengenai pemilihan jalur peminatan yang sesuai dengan kemampuan.

Pemilihan algoritma *K-Means* dalam penelitian ini berdasarkan karena algoritma *K-Means* sederhana dalam dijalankan dan diimplementasikan, juga relatif cepat, mudah dalam beradaptasi, dan umum penggunaannya [3]. Pada dasarnya algoritma *K-Means* dapat diterapkan untuk permasalahan dalam meringkas objek yang berjumlah besar, memahami perilaku konsumen, dan identifikasi peluang pada produk baru dipasaran. Sehingga lebih memudahkan dalam mendeskripsikan sifat ataupun karakteristik pada masing-masing kelompok [8].

2. METODE PENELITIAN

2.1 Metode Pengumpulan Data

Metode dalam pengumpulan data pada penelitian ini yaitu :

1. Wawancara
Wawancara dilakukan terhadap Ketua Program Studi Teknik Informatika-S1 Universitas Dian Nuswantoro.
2. Studi Pustaka
Studi pustaka dilakukan dengan cara mencari referensi yang terkait dengan penelitian ini berdasarkan dari buku, jurnal, internet, dan sebagainya.
3. Survey
Survey penelitian dilakukan dengan meminta izin untuk pengambilan data mengenai data mahasiswa, KHS, serta transkrip nilai mahasiswa program studi Teknik Informatika-S1 angkatan 2011-2013 kepada PSI Universitas Dian Nuswantoro.

2.2 Metode Analisis

Metodologi pengembangan *data mining* yang digunakan ialah CRISP-DM. *Cross-Industry Standart Process for Data Mining* (CRISP-DM) menyediakan standar proses *data mining* sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian [12]. Berikut ini merupakan langkah-langkah yang dilakukan berdasarkan dari fase CRISP-DM :

1. *Business understanding*

Pada tahapan pertama penulis mencoba untuk memahami permasalahan yang ada dalam pemilihan jalur peminatan pada program studi Teknik Informatika-S1 Universitas Dian Nuswantoro. Sehingga dapat menentukan tujuan dan pola yang akan didapatkan dengan *data mining*.

2. *Data understanding*

Pada tahap ini penulis melakukan pemahaman terhadap data yang dibutuhkan, untuk kemudian mengambil data yang relevan dan memiliki keterkaitan dengan tujuan penelitian. Adapun data yang digunakan yaitu data mahasiswa, KHS, dan transkrip nilai.

3. *Data preparation*

Pada tahap ini penulis mengolah data yang didapat dengan melakukan pembersihan data / *data cleaning*, melakukan integrasi data / *data integration*, melakukan pemilihan data / *data selection*, dan transformasi pada data / *data transformation*. *Tools* yang digunakan dalam menunjang pengolahan datanya ialah *microsoft excel*.

4. *Modeling*

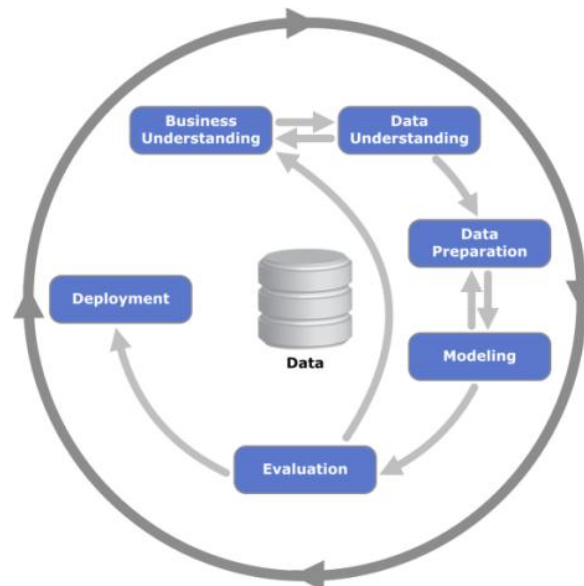
Pada tahap ini penulis menentukan teknik *data mining* yang digunakan untuk mengolah data yang sudah disiapkan sebelumnya. Teknik yang dilakukan yaitu dengan *clustering* menggunakan algoritma *K-Means*.

5. *Evaluation*

Pada tahap evaluasi, akan diketahui apakah hasil daripada tahap pemodelan dapat menjawab tujuan yang telah ditetapkan pada fase pertama. Untuk itu akan dilakukan profilisasi pada setiap *cluster* yang telah terbentuk, untuk diketahui karakteristik pada kelompok tersebut. Disamping itu untuk diketahui kesesuaian dengan jalur peminatan akan dilakukan analisis lebih lanjut untuk dihubungkan dengan atribut peminatan, rerata nilai MK peminatan SC, dan rerata nilai MK peminatan RPLD. Sehingga diharapkan mendapatkan informasi atau pola yang berguna sebagai rekomendasi pemilihan jalur peminatan yang sesuai dengan kemampuan akademis mahasiswa.

6. *Deployment*

Pada tahap ini penulis mencoba menyajikan hasil dari penerapan *data mining* metode *clustering* dengan algoritma *K-Means* yang telah dilakukan.



Gambar 1.Proses CRISP-DM (Sumber : Kenneth Jensen,2012)

2.3 Algoritma K-Means

Algoritma *K-Means* merupakan algoritma pengelompokan iteratif yang melakukan partisi set data ke dalam sejumlah k cluster yang sudah ditetapkan di awal [3]. Berikut merupakan langkah-langkah dari algoritma *K-Means* [3][8]:

1. Tentukan k sebagai jumlah cluster yang dibentuk.
 Dalam menentukan banyaknya cluster k dapat dilakukan dengan pertimbangan konseptual ataupun teoritis dan sesuai dengan tujuan dari pengelompokan.
2. Tentukan titik pusat (*centroid*) dari setiap cluster.
 Untuk menentukan *centroid* awal setiap cluster dapat dilakukan secara *random* / acak.
3. Alokasikan semua data ke *centroid* terdekat dengan matrik jarak yang sudah ditetapkan.
 Rumus dalam menghitung jarak antara lain : jarak *Euclidean*, jarak *Manhattan / Block City*, dan jarak *Minkowski*. Dalam penelitian ini untuk menghitung jarak antara objek dengan *centroid* menggunakan *Euclidean Distance* dengan perhitungan seperti berikut :

$$d(x, y) = \|x - y\| \quad (2.1)$$

$$= \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad ; i = 1, 2, 3 \dots n$$

dimana x_i : objek x ke-i
 y_i : daya y ke-i
 N : banyaknya objek

4. Hitung kembali titik pusat (*centroid*) berdasarkan data yang mengikuti cluster masing-masing.
 Untuk menghitung titik pusat (*centroid*) cluster ke-i berikutnya, menggunakan rumus sebagai berikut :

$$v = \sum_{i=1}^n x_i \quad ; i = 1, 2, 3 \dots n \quad (2.2)$$

dimana V : *centroid* pada cluster
 x_i : objek ke-i
 N : banyaknya objek

5. Ulangi langkah 3 dan 4 jika posisi *centroid* baru berubah / tidak sama.
 Algoritma *K-Means* mencapai kondisi konvergen ketika pengalokasian kembali titik data dan juga lokasi *centroid* tidak berubah [3].

2.4 Rapidminer

Rapidminer merupakan *software* berlisensi gratis yang digunakan sebagai alat bantu dalam *data mining* dan *knowledge discovery*. Pada *rapidminer* terdapat 400 prosedur *data mining*, termasuk operator untuk masukan, *output*, *data preprocessing* dan visualisasi [10].

3. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Tahapan pertama ini, berguna dalam memahami tujuan dari unit penelitian yang dilakukan. Langkah awal yang dilakukan dalam mencapai tujuan tersebut ialah, meminta ijin terhadap PSI UDINUS dalam melakukan permintaan data mahasiswa program studi Teknik Informatika-S1 Universitas Dian Nuswantoro angkatan 2011, 2012, dan 2013.

3.2 Data Understanding

Pada tahap ini penulis melakukan pemahaman terhadap data yang dibutuhkan, untuk kemudian mengambil data yang relevan dan memiliki keterkaitan dengan tujuan penelitian. Adapun data yang digunakan yaitu data mahasiswa, KHS, dan transkrip nilai. Dari data tersebut peneliti dapat menentukan atribut yang akan digunakan ialah IPS 1-4, rerata nilai matakuliah penunjang SC, rerata nilai matakuliah penunjang RPLD, NIM, IPK, rerata nilai matakuliah peminatan SC, rerata nilai matakuliah peminatan RPLD, dan peminatan. Sedangkan atribut yang dipilih sebagai parameter dalam melakukan *clustering* ialah IPS 1-4, rerata nilai matakuliah penunjang SC, dan rerata nilai matakuliah penunjang RPLD. Pemilihan atribut tersebut sebagai parameter dengan pertimbangan dapat merepresentasikan kemampuan akademis mahasiswa sebelum melakukan pemilihan jalur peminatan.

3.3 Data Preparation

3.3.1 Pembersihan Data dan Integrasi Data

Pada proses pembersihan data yang dilakukan terhadap data mahasiswa, KHS dan transkrip nilai. Dimana akan dilakukan proses penghapusan terhadap data yang tidak memiliki nilai (null), tidak relevan, tidak konsisten, dan data salah *input*. Sedangkan untuk proses integrasi data dilakukan pembuatan dataset baru. Dimana atribut-atribut yang digunakan ialah NIM, IPK, IPS 1-4, rerata nilai matakuliah penunjang SC, rerata nilai matakuliah penunjang RPLD, rerata nilai matakuliah peminatan SC, rerata nilai matakuliah peminatan RPLD, dan peminatan.

3.3.2 Seleksi Data dan Transformasi Data

Pada proses seleksi data dan transformasi data, peneliti melakukan perhitungan rerata IPS ke-1 sampai 4, yang kemudian akan menghasilkan atribut IPS 1-4. Dan juga perhitungan rerata nilai matakuliah yang telah dilakukan inisialisasi sebelumnya sehingga menghasilkan atribut rerata nilai matakuliah penunjang SC, rerata nilai matakuliah penunjang RPLD, rerata nilai matakuliah peminatan SC, rerata nilai matakuliah peminatan RPLD. Sedangkan atribut peminatan didapatkan berdasarkan matakuliah peminatan yang telah dipilih oleh mahasiswa.

	A	B	C	D	E	F	G	H
1	NIM	IPK	IPS 1-4	rerata nilai mtk penunjang SC	rerata nilai mtk penunjang RPLD	rerata nilai MK peminatan SC	rerata nilai MK peminatan RPLD	peminatan
2	A11.2011.05830	3,21	3,03	4	3,75	0	4	RPLD
3	A11.2011.05832	2,75	2,07	4	3,5	4,5	0	SC
4	A11.2011.05833	3,08	2,39	3,75	3,75	0	4,3	RPLD
5	A11.2011.05835	3,37	3,16	4,75	3,5	0	4,3	RPLD
6	A11.2011.05837	3,29	3,18	4,75	4	4	0	SC

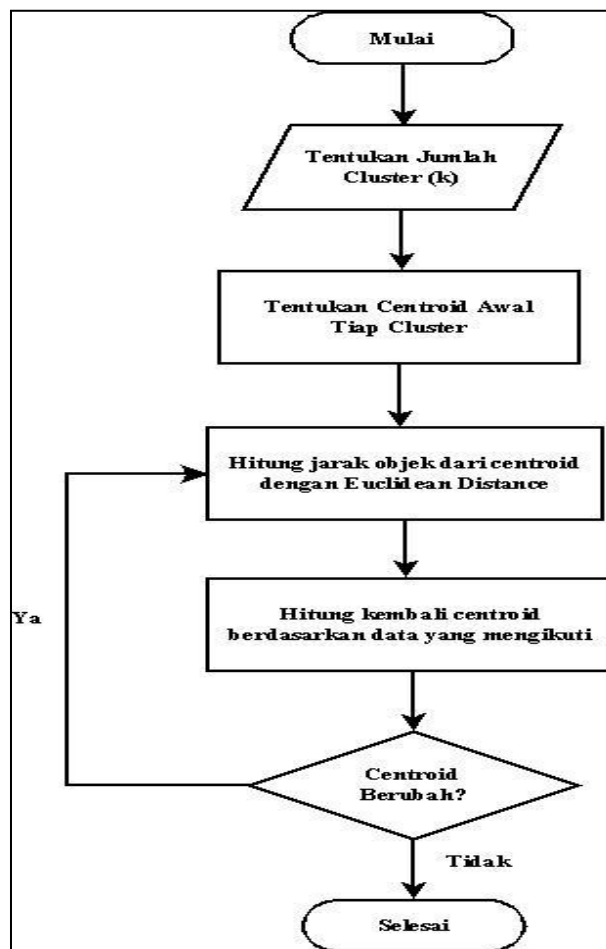
Gambar 2. Contoh Dataset Hasil Akhir pada Tahap *Data Preparation*

3.4 Modeling

Tahap ini ditentukan teknik yang digunakan dalam melakukan *data mining*. Dengan teknik *clustering* menggunakan algoritma *K-Means*.

3.4.1 Perhitungan K-Means Secara Manual

Data yang digunakan dalam perhitungan *K-Means* secara manual diambil secara acak yang berjumlah 50 sampel data. Langkah-langkah dalam melakukan perhitungan *K-Means* yaitu :



Gambar 3. Flowchart Algoritma *K-Means*

1. Menentukan jumlah *cluster* yang dibentuk.
Cluster yang terbentuk yaitu $k=5$; dengan pertimbangan pemilihan 5 *cluster* dapat diketahui tingkatan yang lebih bervariasi dari kelompok mahasiswa berdasarkan kemampuan akademisnya.
2. Menentukan titik pusat (*centroid*) pada *cluster*.

Tabel 1: Penentuan Titik Pusat (*Centroid*) Awal

Atribut	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
IPS 1-4	2.985	2.414	3.436	3.018	2.539
Rerata Nilai MK Penunjang SC	4.114	3.375	4.625	4.469	3.909
Rerata Nilai MK Penunjang RPLD	4.068	3.6	4.4	3.750	3.545

3. Menghitung jarak data dengan *centroid* terdekat menggunakan *Euclidean Distances*.
Persamaan yang digunakan ialah :

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2} \quad (3.1)$$

Sebagai contohnya, dilakukan perhitungan jarak terhadap data mahasiswa pertama dengan titik pusat (*centroid*) pada *cluster* pertama.

$$\begin{aligned} d(1,1) &= \sqrt{(x_{11} - y_{11})^2 + (x_{12} - y_{12})^2 + (x_{13} - y_{13})^2} \\ &= \sqrt{(2.2 - 2.985)^2 + (3.5 - 4.114)^2 + (3.5 - 4.068)^2} \\ &= \sqrt{1.284845} \\ &= 1.134 \end{aligned} \quad (3.2)$$

Perhitungan jarak mahasiswa pertama dengan *centroid cluster* kedua, seperti berikut :

$$\begin{aligned} d(1,2) &= \sqrt{(x_{11} - y_{21})^2 + (x_{12} - y_{22})^2 + (x_{13} - y_{23})^2} \\ &= \sqrt{(2.2 - 2.414)^2 + (3.5 - 3.375)^2 + (3.5 - 3.6)^2} \\ &= \sqrt{0.063261} \\ &= 0.252 \end{aligned} \quad (3.3)$$

Perhitungan jarak mahasiswa pertama dengan *centroid* pada *cluster* ketiga, seperti berikut :

$$\begin{aligned} d(1,3) &= \sqrt{(x_{11} - y_{31})^2 + (x_{12} - y_{32})^2 + (x_{13} - y_{33})^2} \\ &= \sqrt{(2.2 - 3.436)^2 + (3.5 - 4.625)^2 + (3.5 - 4.4)^2} \\ &= \sqrt{3.554281} \\ &= 1.885 \end{aligned} \quad (3.4)$$

Perhitungan jarak mahasiswa pertama dengan *centroidcluster* keempat, seperti berikut :

$$\begin{aligned}
 d(1,4) &= \sqrt{(x_{11} - y_{41})^2 + (x_{12} - y_{42})^2 + (x_{13} - y_{43})^2} \\
 &= \sqrt{(2.2 - 3.018)^2 + (3.5 - 4.469)^2 + (3.5 - 3.75)^2} \\
 &= \sqrt{1.638265} \\
 &= 1.28
 \end{aligned}
 \tag{3.5}$$

Perhitungan jarak mahasiswa pertama dengan *centroidcluster* kelima, seperti berikut :

$$\begin{aligned}
 d(1,5) &= \sqrt{(x_{11} - y_{51})^2 + (x_{12} - y_{52})^2 + (x_{13} - y_{53})^2} \\
 &= \sqrt{(2.2 - 2.539)^2 + (3.5 - 3.909)^2 + (3.5 - 3.545)^2} \\
 &= \sqrt{0.271067} \\
 &= 0.521
 \end{aligned}
 \tag{3.6}$$

Berikut hasil perhitungan jarak data dengan titik pusat pada iterasi 1 menggunakan *Euclidean Distances*.

1	IPS 1-4 rerata nilai mk penunjang SC	rerata nilai mk penunjang RPLD	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	
2	2,22	3,5	3,5	1,134	0,252	1,885	1,28	0,521
3	2,93	4	3,75	0,342	0,824	1,034	0,477	0,451
4	2,99	4,25	4,25	0,227	1,233	0,602	0,547	0,904
5	2,9	3,75	4	0,38	0,733	1,101	0,77	0,602
6	2,75	4	4	0,27	0,815	1,011	0,595	0,51
7	2,31	3,5	3,5	1,075	0,191	1,829	1,226	0,471
8	3,39	4,5	4,25	0,588	1,625	0,201	0,624	1,253
9	2,25	4,25	3,5	0,939	0,896	1,535	0,837	0,449
10	3,08	4,75	3,75	0,717	1,535	0,752	0,288	1,021
11	3,36	5	4,5	1,055	2,085	0,395	0,981	1,666
12	2,5	3,25	3,75	1,041	0,213	1,786	1,324	0,691
13	2,51	4	3,75	0,583	0,65	1,293	0,691	0,226
14	2,48	3,75	3,75	0,699	0,409	1,45	0,898	0,266
15	2,6	4	3,5	0,696	0,66	1,378	0,676	0,118
16	2,4	3,5	3,75	0,906	0,196	1,662	1,149	0,478
17	3,02	4,5	4	0,394	1,339	0,59	0,252	0,888
18	2,49	3,25	3,5	1,146	0,177	1,896	1,352	0,662
19	3,4	4,5	4,5	0,713	1,746	0,164	0,842	1,415
20	3,13	4,5	4	0,418	1,392	0,519	0,276	0,952
21	3,53	4,75	4,25	0,857	1,886	0,217	0,769	1,479
22	2,86	4,5	3,75	0,516	1,219	0,877	0,161	0,703
23	3,04	4,5	3,75	0,503	1,296	0,771	0,038	0,801
24	3,06	4,5	3,75	0,506	1,306	0,761	0,052	0,814
25	2,45	3,5	3,75	0,874	0,199	1,631	1,123	0,466

Gambar 4.Contoh Hasil Perhitungan Jarak Data ke Titik Pusat pada Iterasi 1

Data / objek yang memiliki jarak terkecil dengan *centroid* akan menjadi anggota pada kelompok tersebut. Berikut ini merupakan posisi data dengan tiap *cluster* pada iterasi ke-1, dengan menggunakan bantuan tanda (*) yang menyatakan data menjadi anggota pada *cluster*.

IPS 1-4	rerata nilai mk penunjang SC	rerata nilai mk penunjang RPLD	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
2,22	3,5	3,5	*				
2,93	4	3,75	*				
2,99	4,25	4,25	*				
2,9	3,75	4	*				
2,75	4	4	*				
2,31	3,5	3,5		*			
3,39	4,5	4,25			*		
2,25	4,25	3,5					*
3,08	4,75	3,75				*	
3,36	5	4,5			*		
2,5	3,25	3,75	*				
2,51	4	3,75					*
2,48	3,75	3,75					*
2,6	4	3,5					*
2,4	3,5	3,75	*				
3,02	4,5	4				*	
2,49	3,25	3,5	*				
3,4	4,5	4,5		*			
3,13	4,5	4				*	
3,53	4,75	4,25		*			
2,86	4,5	3,75				*	
3,04	4,5	3,75				*	
3,06	4,5	3,75				*	
2,45	3,5	3,75	*				

Gambar 5.Contoh Posisi Data dengan Tiap *Cluster* pada Iterasi 1

4. Menghitung titik pusat baru menggunakan hasil dari setiap anggota pada masing-masing *cluster*. Contoh perhitungan titik pusat baru pada *cluster 0* ialah :

$$\left(\frac{2.93 + 2.99 + 2.9 + 2.75 + 2.76 + 3 + 2.98 + 2.92 + 3.09 + 2.89}{10}, \frac{4 + 4.25 + 3.75 + 4 + 3.75 + 4 + 4 + 3.75 + 3.75 + 4.25}{10}, \frac{3.75 + 4.25 + 4 + 4 + 4 + 4 + 4.25 + 4 + 4 + 4.25}{10} \right) \quad (3.7)$$

Jadi titik pusat baru pada *cluster 0* ialah (2.921, 3.95, 4.05).

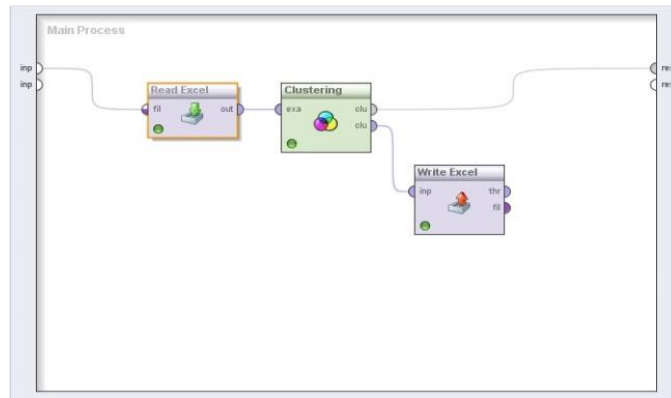
5. Selanjutnya dilakukan kembali langkah 3 dan 4, hingga tidak ditemukan lagi perubahan posisi pada data dan perubahan titik pusat.

Tabel2:Titik Pusat Akhir pada Perhitungan *K-Means* Manual

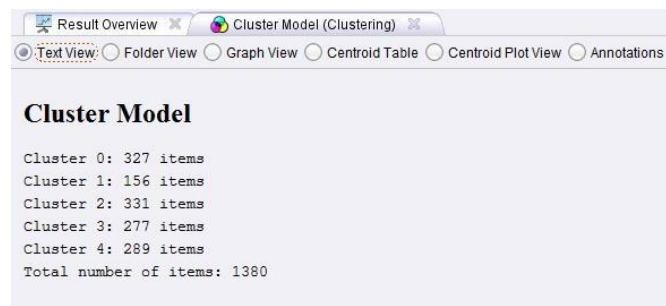
Atribut	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
IPS 1-4	2.921	2.414	3.436	3.052	2.517
Rerata Nilai MK Penunjang SC	3.95	3.375	4.625	4.525	3.925
Rerata Nilai MK Penunjang RPLD	4.05	3.6	4.4	3.825	3.5

3.4.2 Penerapan K-Means Clustering pada Rapidminer

Data yang digunakan dalam perhitungan *K-Means* pada *rapidminer* berjumlah 1380. Berikut ini pada Gambar 6 merupakan proses yang dilakukan dalam penerapan *K-Meansclustering* pada *rapidminer*.



Gambar 6. Proses Clustering pada Rapidminer



Gambar 7. Hasil Cluster pada Rapidminer

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4
IPS 1-4	2.915	2.404	3.372	3.056	2.514
rerata nilai mk penunjang SC	3.933	3.338	4.696	4.513	3.964
rerata nilai mk penunjang RPLD	4.059	3.590	4.418	3.837	3.545

Gambar 8. Hasil Centroid Akhir dari Setiap Cluster

3.5 Evaluation

Pada tahap evaluasi, bertujuan dalam mengetahui apakah hasil dari tahap pemodelan dapat menjawab tujuan yang telah ditetapkan pada fase pertama. Untuk itu dilakukan profilisasi dan analisis lebih lanjut dengan bantuan *MS.Excel*.

3.5.1 Profilisasi pada Cluster

Profilisasi bertujuan untuk mengidentifikasi dan menampilkan karakteristik yang didapatkan dari setiap *cluster* yang terbentuk.

1. *Cluster 0*
Merupakan kelompok mahasiswa yang memiliki IPS 1-4 sedang, rerata nilai MK penunjang SC sedang, rerata nilai MK penunjang RPLD sedang.
2. *Cluster 1*
Merupakan kelompok mahasiswa yang memiliki IPS 1-4 rendah, rerata nilai MK penunjang SC rendah, rerata nilai MK penunjang RPLD rendah.
3. *Cluster 2*
Merupakan kelompok mahasiswa yang memiliki IPS 1-4 tinggi, rerata nilai MK penunjang SC tinggi, rerata nilai MK penunjang RPLD tinggi.
4. *Cluster 3*
Merupakan kelompok mahasiswa yang memiliki IPS 1-4 sedang, rerata nilai MK penunjang SC tinggi, rerata nilai MK penunjang RPLD sedang.
5. *Cluster 4*
Merupakan kelompok mahasiswa yang memiliki IPS 1-4 rendah, rerata nilai MK penunjang SC sedang, rerata nilai MK penunjang RPLD rendah.

3.5.2 Menentukan Pemilihan Jalur Peminatan

Berdasarkan hasil karakteristik dari 5 *cluster* yang terbentuk, akan dianalisis lebih lanjut untuk didapatkan informasi mengenai jalur peminatan pada setiap *cluster* tersebut. Cara yang dilakukan untuk mengetahui kesesuaian dengan jalur peminatan ialah menghubungkan dengan atribut peminatan, atribut rerata nilai MK peminatan SC, dan atribut rerata nilai MK peminatan RPLD. Sehingga akan diketahui informasi dan pola karakteristik mahasiswa berdasarkan tingkat kemampuan akademisnya yang sesuai dengan jalur peminatan.

1. *Cluster 0*
Terdapat 72.5% dari 327 mahasiswa yang telah memilih peminatan RPLD. Sedangkan 27.5% dari 327 mahasiswa memilih peminatan SC. Dapat diartikan bahwa *cluster 0* lebih didominasi oleh mahasiswa yang memilih peminatan RPLD dibandingkan peminatan SC. Pada *cluster* ini juga didapatkan 64 dari 90 mahasiswa (71.1%) yang memiliki nilai 4-5 pada peminatan SC. Dan 172 dari 237 mahasiswa (72.6%) memiliki rentang nilai 4-5 pada peminatan RPLD. Dapat diartikan bahwa mahasiswa yang mengambil peminatan RPLD pada *cluster 0* memiliki nilai peminatan yang lebih baik dibandingkan mahasiswa yang mengambil peminatan SC. Oleh karena itu pada *cluster 0* lebih direkomendasikan untuk jalur peminatan RPLD.
 2. *Cluster 1*
Terdapat 67.5% dari 156 mahasiswa yang memilih peminatan RPLD. Sedangkan 32.7% mahasiswa memilih peminatan SC. Dapat diartikan bahwa *cluster 1* masih didominasi oleh mahasiswa yang mengambil peminatan RPLD dibandingkan peminatan SC. Pada *cluster* ini juga didapatkan 23 dari 51 mahasiswa (45.1%) memiliki rentang nilai 4-5 pada peminatan SC. Dan 55 dari 105 mahasiswa (52.4%) memiliki rentang nilai 4-5 pada peminatan RPLD. Dapat diartikan bahwa, pada *cluster 1* mahasiswa yang mengambil peminatan RPLD memiliki nilai peminatan yang lebih baik dibandingkan mahasiswa yang mengambil peminatan SC. Oleh karena itu pada *cluster 1* lebih direkomendasikan untuk jalur peminatan RPLD.
 3. *Cluster 2*
Lebih banyak ditempati oleh mahasiswa yang memilih peminatan SC sebesar 64.4% dari 331 mahasiswa. Sisanya 35.6% dari 331 mahasiswa memilih peminatan RPLD. Dapat diartikan bahwa *cluster 2* didominasi oleh mahasiswa yang memilih peminatan SC. Pada *cluster* ini juga didapatkan 199 dari 213 mahasiswa (93.4%) memiliki rentang nilai 4-5 pada peminatan SC. Dan 109 dari 118 mahasiswa (92.4%) memiliki rentang nilai 4-5 pada peminatan RPLD. Dapat diartikan bahwa
-

mahasiswa yang mengambil peminatan SC pada *cluster 2* memiliki nilai peminatan yang lebih baik dibandingkan mahasiswa yang mengambil peminatan RPLD. Oleh karena itu pada *cluster 2* lebih direkomendasikan untuk jalur peminatan SC.

4. *Cluster 3*

Lebih banyak ditempati oleh mahasiswa yang memilih peminatan SC sebesar 75.1% dari 277 mahasiswa. Sedangkan sisanya 24.9% mahasiswa memilih RPLD. Dapat diartikan bahwa *cluster 3* didominasi oleh mahasiswa yang memilih peminatan SC. Pada *cluster* ini jugadidapatkan 186 dari 208 mahasiswa (89.4%) memiliki rentang nilai 4-5 pada peminatan SC. Sedangkan 52 dari 69 mahasiswa (75.4%) memiliki rentang nilai 4-5 pada peminatan RPLD. Dapat diartikan bahwa mahasiswa yang mengambil peminatan SC pada *cluster 3* memiliki nilai peminatan yang lebih baik dibandingkan mahasiswa yang mengambil peminatan RPLD. Oleh karena itu pada *cluster 3* lebih direkomendasikan untuk jalur peminatan SC.

5. *Cluster 4*

Lebih banyak ditempati oleh mahasiswa yang memilih peminatan SC sebesar 70.9% dari 289 mahasiswa. Sedangkan sisanya 29.1% mahasiswa memilih peminatan RPLD. Dapat diartikan bahwa *cluster 4* lebih didominasi oleh mahasiswa yang memilih peminatan SC. Pada *cluster* ini juga didapatkan 122 dari 205 mahasiswa (59.5%) memiliki rentang nilai 4-5 pada peminatan SC. Sedangkan 40 dari 84 mahasiswa (47.6%) memiliki rentang nilai 4-5 pada peminatan RPLD. Dapat diartikan bahwa mahasiswa yang mengambil peminatan SC pada *cluster 4* memiliki nilai peminatan yang lebih baik jika dibandingkan mahasiswa yang mengambil peminatan RPLD. Oleh karena itu pada *cluster 4* lebih direkomendasikan untuk jalur peminatan SC.

3.5.3 Mengukur Tingkat Akurasi Rekomendasi

Dalam mengukur tingkat akurasi dari rekomendasi yang dihasilkan, akan dilakukan perbandingan antara hasil rekomendasi dengan sistem rekomendasi lama untuk diketahui keakuratannya dengan peminatan yang dipilih mahasiswa.

Dari hasil perbandingan menggunakan 1380 data, didapatkan bahwa keakuratan dari rekomendasi yang dihasilkan jika dibandingkan dengan peminatan yang telah dipilih sebesar 70.1%. Sedangkan dengan sistem rekomendasi lama jika dibandingkan dengan peminatan yang dipilih didapatkan keakuratannya sebesar 55.9%.

3.6 Deployment

Pada tahap ini dilakukan untuk penyusunan laporan atau presentasi hasil pengetahuan yang didapatkan dari proses *data mining*. Berdasarkan penerapan data mining metode *clustering* menggunakan algoritma *K-Means* dari data mahasiswa program studi Teknik Informatika-S1 angkatan 2011, 2012, dan 2013 dihasilkan suatu rekomendasi pemilihan jalur peminatan yang sesuai dengan kemampuan. Untuk kedepannya hasil tersebut dapat digunakan program studi Teknik Informatika-S1 sebagai salah satu alternatif rekomendasi untuk diberikan kepada mahasiswa. Sebagai bahan acuan untuk mahasiswa yang merasa kebingungan dan kesulitan dalam menentukan pilihan, sehingga dapat dijadikan pertimbangan dalam pemilihan jalur peminatan yang sesuai tingkat kemampuan akademis yang dimilikinya.

No	Keterangan Karakteristik Kelompok Mahasiswa	Titik Pusat Cluster			Rekomendasi Jalur Peminatan
		IPS 1-4	Rerata Nilai MK Penunjang SC	Rerata Nilai MK Penunjang RPLD	
1	Kelompok mahasiswa yang memiliki : a. IPS 1-4 Sedang b. Rerata Nilai MK Penunjang SC Sedang c. Rerata Nilai MK Penunjang RPLD Sedang	2.915	3.933	4.059	RPLD
2	Kelompok mahasiswa yang memiliki : a. IPS 1-4 Rendah b. Rerata Nilai MK Penunjang SC Rendah c. Rerata Nilai MK Penunjang RPLD Rendah	2.404	3.338	3.590	RPLD
3	Kelompok mahasiswa yang memiliki : a. IPS 1-4 Tinggi b. Rerata Nilai MK Penunjang SC Tinggi c. Rerata Nilai MK Penunjang RPLD Tinggi	3.372	4.696	4.418	SC
4	Kelompok mahasiswa yang memiliki : a. IPS 1-4 Sedang b. Rerata Nilai MK Penunjang SC Tinggi c. Rerata Nilai MK Penunjang RPLD Sedang	3.056	4.513	3.837	SC
5	Kelompok mahasiswa yang memiliki : a. IPS 1-4 Rendah b. Rerata Nilai MK Penunjang SC Sedang c. Rerata Nilai MK Penunjang RPLD Rendah	2.514	3.964	3.543	SC

Gambar 9.Rekomendasi Pemilihan Jalur Peminatan

4. KESIMPULAN

Dari hasil penelitian dapat ditarik kesimpulan bahwa :

1. Dari penerapan *data mining* menggunakan metode *clustering* dengan algoritma *K-Means*, didapatkan 5 kelompok mahasiswa berdasarkan kemampuan akademisnya. Hasil dari analisis diharapkan dapat membantu program studi Teknik Informatika-S1 untuk kedepannya, sebagai salah satu alternatif rekomendasi untuk diberikan kepada mahasiswa mengenai pemilihan jalur peminatan yang sesuai dengan kemampuan akademisnya. Berikut ini penjelasan dari tiap *cluster* yang terbentuk :
 - a. *Cluster 0*, memiliki titik pusat (2.915, 3.933, 4.059). Yang dimaksudkan agar pada *cluster 0* terdapat mahasiswa yang memiliki IPS 1-4 sedang, rerata nilai MK penunjang SC sedang, dan rerata nilai MK penunjang RPLD sedang. Pada *cluster 0* lebih didominasi oleh mahasiswa yang memilih peminatan RPLD dan juga memiliki nilai peminatan RPLD yang lebih baik. Oleh karena itu *cluster 0* direkomendasikan untuk jalur peminatan RPLD.
 - b. *Cluster 1*, memiliki titik pusat (2.404, 3.338, 3.590). Merupakan kelompok mahasiswa yang memiliki nilai IPS 1-4 rendah, rerata nilai MK penunjang SC rendah, dan rerata nilai MK penunjang RPLD rendah. Pada *cluster 1* didominasi mahasiswa yang memilih RPLD dan memiliki nilai peminatan RPLD yang lebih baik. Oleh karena itu *cluster 1* masih direkomendasikan untuk jalur peminatan RPLD.
 - c. *Cluster 2*, memiliki titik pusat (3.372, 4.696, 4.418). Merupakan kelompok mahasiswa yang memiliki nilai IPS 1-4 tinggi, rerata nilai MK penunjang SC tinggi, rerata nilai MK penunjang RPLD tinggi. Pada *cluster 2* didominasi mahasiswa yang memilih SC dan memiliki nilai peminatan SC yang lebih baik. Oleh karena itu *cluster 2* direkomendasikan untuk jalur peminatan SC.

- d. *Cluster 3*, memiliki titik pusat (3.056, 4.513, 3.837). Merupakan kelompok mahasiswa yang memiliki nilai IPS 1-4 sedang, rerata nilai MK penunjang SC tinggi, rerata nilai MK penunjang RPLD sedang. Pada *Cluster 3* didominasi mahasiswa yang memilih SC dan memiliki nilai peminatan SC yang lebih baik. Oleh karena itu *cluster 3* direkomendasikan untuk jalur peminatan SC.
- e. *Cluster 4*, memiliki titik pusat (2.514, 3.964, 3.545). Merupakan kelompok mahasiswa yang memiliki nilai IPS 1-4 rendah, rerata nilai MK penunjang SC sedang, rerata nilai penunjang MK RPLD rendah. Pada *cluster 4* didominasi mahasiswa yang memilih SC dan memiliki nilai peminatan SC yang lebih baik. Oleh karena itu *cluster 4* direkomendasikan untuk jalur peminatan SC.

Berdasarkan hasil perbandingan dalam mengukur tingkat akurasi dari rekomendasi yang dihasilkan, didapatkan bahwa keakuratan rekomendasi yang dihasilkan dengan peminatan yang dipilih mahasiswa sebesar 70,1%. Sedangkan keakuratan dari sistem rekomendasi lama dengan peminatan yang dipilih mahasiswa sebesar 55.9%.

5. SARAN

1. Untuk penelitian selanjutnya diharapkan dapat menghubungkan peminatan dengan topik penelitian yang diambil oleh mahasiswa. Untuk mengetahui apakah sesuai topik penelitian yang telah diambil mahasiswa dengan peminatannya, sehingga didapatkan informasi yang lebih lengkap mengenai jalur peminatan.
2. Untuk penelitian selanjutnya diharapkan dapat melakukan langkah uji coba seperti : pengujian jumlah *cluster*, pengujian validitas *cluster*. Agar *cluster* yang didapat akan lebih optimal.
3. Penerapan tersebut sebaiknya diperbaharui secara berkala, agar dapat memberikan hasil rekomendasi yang akurat.

DAFTAR PUSTAKA

- [1] Universitas Dian Nuswantoro, *Buku Panduan Akademik 2015 / 2016*. Semarang: Universitas Dian Nuswantoro, 2015.
- [2] Muhammad Faisal Mirza, "Metode Clustering Dengan Algoritma Fuzzy C-Means Untuk Rekomendasi Pemilihan Bidang Keahlian Pada Program Studi Teknik Informatika," 2013.
- [3] Eko Prasetyo, *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: ANDI, 2014.
- [4] Marlindawati and Andri, "Model Data Mining Dalam Pengklasifikasian Ketertarikan Belajar Mahasiswa Menggunakan Metode Clustering," *Seminar Nasional Teknologi Informasi dan Multimedia*, vol. 2, no. 1, pp. 7-12, 2015.
- [5] Fajar Astuti Hermawati, *Data Mining*. Yogyakarta: CV ANDI OFFSET, 2013.
- [6] Totok Suprawoto, "Klasifikasi Data Mahasiswa Menggunakan Metode K-Means Untuk Menunjang Pemilihan Strategi Pemasaran," *Jurnal Informatika dan Komputer (JIKO)*, vol. 1, no. 1, pp. 12-18, Februari 2016.
- [7] Tamsir Hasudungan Sirait and Johan Oscar Ong, "Analisis Keberhasilan Mahasiswa Dengan Metode Clustering K-Means," *SNASTIA*, pp. 1-6, 2011.
- [8] Ediyanto, Muhlasah Novitasari Mara, and Neva Satyahadewi, "Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis," *Buletin Ilmiah Mat. Stat dan Terapannya (Bimaster)*,

- vol. 02, no. 2, pp. 133-136, 2013.
- [9] Sofi Defiyanti and Mohamad Jajuli, "Integrasi Metode Klasifikasi Dan Clustering dalam Data Mining," *Konferensi Nasional Informatika (KNIF)*, pp. 39-44, 2015.
- [10] Lee Finn and Juan Santana, *Data Mining : Meramalkan Bisnis Perusahaan*. Jakarta: PT Elex Media Komputindo, 2010.
- [11] Hasan Ziafat and Majid Shakeri, "Using Data Mining Techniques in Customer Segmentation," *Journal of Engineering Research and Applications*, vol. 4, no. 9, pp. 70-79, 2014. [Online]. www.ijera.com
- [12] Kusriani and Emha Taufiq, *Algoritma Data Mining*. Yogyakarta: CV ANDI OFFSET, 2009.
- [13] Fakultas Ilmu Komputer, *Buku Panduan Minat Studi Mahasiswa Prodi Teknik Informatika Fakultas Ilmu Komputer 2012/2013*. Semarang: Universitas Dian Nuswantoro, 2012.
- [14] Sumanto and Romi Satrio Wahono, "Penerapan Fuzzy C-Means Dalam Pemilihan Peminatan Tugas Akhir Mahasiswa," *Proceeding ISIT*, vol. 1, no. 1, pp. 1-8, 2011.
-