

OAJIS

Open Access
Journal of
Information
Systems

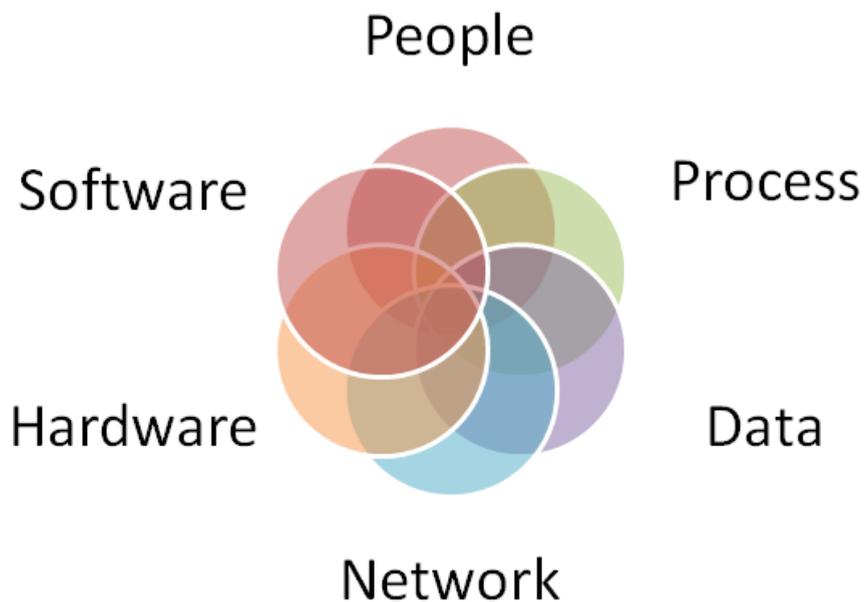
is.its.ac.id/pubs/oajis/

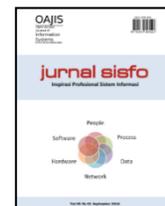
ISSN 1979-3979



jurnal sisfo

Inspirasi Profesional Sistem Informasi





Pimpinan Redaksi

Eko Wahyu Tyas Darmaningrat

Dewan Redaksi

Amna Shifia Nisafani

Arif Wibisono

Faizal Mahananto

Tata Pelaksana Usaha

Achmad Syaiful Susanto

Ricky Asrul Sani

Rini Ekowati

Sekretariat

Jurusan Sistem Informasi – Fakultas Teknologi Informasi

Institut Teknologi Sepuluh Nopember (ITS) – Surabaya

Telp. 031-5999944 Fax. 031-5964965

Email: editor@jurnalsisfo.org

Website: <http://jurnalsisfo.org>

Jurnal SISFO juga dipublikasikan di *Open Access Journal of Information Systems* (OAJIS)

Website: <http://is.its.ac.id/pubs/oajis/index.php>



Mitra Bestari

Aditya Rachmadi, S.ST, M.TI (Universitas Brawijaya)

Ahmad Mukhlason, S.Kom, M.Sc, Ph.D (Institut Teknologi Sepuluh Nopember)

Amalia Utamima, S.Kom, MBA (Institut Teknologi Sepuluh Nopember)

Anisah Herdiyanti, S.Kom, M.Sc, ITILF (Institut Teknologi Sepuluh Nopember)

Ari Widyanti, S.T, M.T, Ph.D (Institut Teknologi Bandung)

Dany Primanita Kartikasari, S.T, M.Kom (Universitas Brawijaya)

Dewi Yanti Liliana, S.Kom, M.Kom (Politeknik Negeri Jakarta)

Erma Suryani, S.T, M.T, Ph.D (Institut Teknologi Sepuluh Nopember)

Faizal Johan Atletiko, S.Kom, M.T (Institut Teknologi Sepuluh Nopember)

Feby Artwodini Muqtadiroh, S.Kom, M.T (Institut Teknologi Sepuluh Nopember)

Hatma Suryotrisongko, S.Kom, M.Eng (Institut Teknologi Sepuluh Nopember)

Irmasari Hafidz, S.Kom, M.Sc (Institut Teknologi Sepuluh Nopember)

Mahendrawathi ER., S.T, M.Sc, Ph.D (Institut Teknologi Sepuluh Nopember)

Mudjahidin, S.T, M.T (Institut Teknologi Sepuluh Nopember)

Nur Aini R., S.Kom, M.Sc.Eng, Ph.D (Institut Teknologi Sepuluh Nopember)

Radityo Prasetyanto W., S.Kom, M.Kom (Institut Teknologi Sepuluh Nopember)

Rahadian Bisma, S.Kom, M.Kom, ITILF (Universitas Negeri Surabaya)

OAJIS

Open Access
Journal of
Information
Systems
is.its.ac.id/pubs/oajis/

jurnal sisfo

Jurnal Sisfo Vol. 6 No. 1 (2017) i–iii



Mitra Bestari

Raras Tyasnurita, S.Kom, MBA (Institut Teknologi Sepuluh Nopember)

Renny Pradina Kusumawardani, S.T, M.T (Institut Teknologi Sepuluh Nopember)

Retno Aulia Vinarti, S.Kom, M.Kom (Institut Teknologi Sepuluh Nopember)

Sholiq, S.T, M.Kom, M.SA (Institut Teknologi Sepuluh Nopember)

Wiwik Anggraeni, S.Si, M.Kom (Institut Teknologi Sepuluh Nopember)



Daftar Isi

Pembuatan Perangkat Lunak Berbasis Lokasi untuk Berbagi Kendaraan

Arif Wibisono, Amril Azhar..... 265

Pemetaan Proses Bisnis dengan Pendekatan Klasifikasi Proses CIMOSA: Studi Kasus Perusahaan Pengelola Kawasan Industri

Effi Latiffianti, Stefanus Eko Wiratno, Dewanti Anggrahini, Muhammad Saiful Hakim..... 283

Sistem Penginderaan Berbasis UAV untuk Membantu Operasi Pencarian dan Penyelamatan Korban Kecelakaan di Wilayah Pegunungan

Ketut Bayu Yogha, Rajalida Lipikorn..... 293

Analisis dan Perancangan Sistem Informasi Data Siswa Pada Sekolah Menengah Kejuruan (SMK) PGRI 8 Medan dengan *Zachman Framework*

Safrian Aswati, Ada Udi Firmansyah, William Ramdhan, Suhendra..... 309

Desain dan Evaluasi Prototipe Jaringan Sensor Nirkabel untuk Monitoring Lahan Persawahan di Kabupaten Gowa

Mohammad Fajar, Agus Halid, Syaiful Rahman 319

Evaluasi Kebergunaan (*Usability*) pada Aplikasi Daftar Online Rumah Sakit Umum Daerah Gambiran Kediri

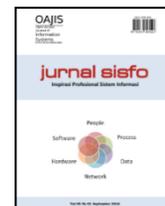
Fithrotu Khoirina, Anisah Herdiyanti, Tony Dwi Susanto..... 331

Sistem Pakar untuk Menentukan Penyakit Hernia dengan Menggunakan Metode *Dempster-Shafer*

Tumingan, Yessy Yanitasari, Dedih..... 347

Pengelompokan Peminatan Program Menggunakan *K-Means Clustering* Berdasarkan Asal Sekolah

C. Purnama Yanti..... 383



Pengelompokan Peminatan Program Menggunakan K-Means Clustering Berdasarkan Asal Sekolah

C. Purnama Yanti*

Pascasarjana Ilmu Komputer, Universitas Pendidikan Ganesha

Abstract

The abundance of new students' data when processed will yield hidden information that can be useful in the future. To obtain the information is needed Data Mining techniques. algorithm used a K-Means Clustering algorithm. This algorithm classifies the data with similar characteristics quickly and efficiently. This study took some new students data sources in PPLP Dhyana Pura as 640 data. From the analysis concluded that in this study resulted in three clusters. In the first cluster, it can be seen that the characteristics of students in cluster 1 are dominated by students from SMA who are more interested in enrolling in 2-years program with percentage of 39.66%. Then in the second cluster can be seen the characteristics of students dominated by students from SMK choose 2-years program with a percentage of 53.43%. In the third cluster, Kejar Paket C to choose 2-years program with 6.9% percentage. The result of this research is expected to make reference by campus marketing party in promoting PPLP Dhyana Pura so that later can increase the number of students every year.

Keywords: Data Mining, K-Means Clustering, Programs, Origin of School

Abstrak

Berlimpahnya data mahasiswa baru apabila diolah akan menghasilkan informasi yang tersembunyi yang dapat bermanfaat dimasa yang akan datang. Untuk mendapatkan informasi tersebut dibutuhkan teknik yaitu *Data Mining*. Algoritma yang digunakan adalah algoritma *K-Means Clustering*. Algoritma ini mengelompokkan data-data yang memiliki kesamaan karakteristik dengan cepat dan efisien. Penelitian ini mengambil sumber data mahasiswa baru di PPLP Dhyana Pura sebanyak 2231 data. Dari hasil analisis didapatkan tiga *cluster* dari hasil perhitungan *K-Means Clustering*. *Cluster* pertama didominasi oleh mahasiswa dengan asal sekolah SMA yang lebih berminat mendaftar pada program 2 tahun dengan persentase 39.66%. Kemudian pada *cluster* kedua didominasi oleh mahasiswa dengan asal sekolah SMK yang memilih program 2 tahun dengan persentase 53,43%. Pada *cluster* 3, Kejar Paket C lebih dominan memilih program 2 tahun dengan persentase 6,9%. Hasil dari penelitian ini diharapkan bisa dijadikan referensi oleh pihak marketing kampus dalam mempromosikan PPLP Dhyana Pura sehingga nantinya bisa meningkatkan jumlah mahasiswa setiap tahunnya.

Kata kunci: Data Mining, K-Means Clustering, Program, Asal Sekolah

© 2016 Jurnal SISFO.

Histori Artikel : Disubmit 31 Maret 2017; Diterima 12 Mei 2017; Tersedia online 22 Mei 2017

*Corresponding Author

Email address: christ25.christina@gmail.com (C. Purnama Yanti)

1. Pendahuluan

Dokumen penerapan teknologi informasi pada dunia pendidikan khususnya pada penerimaan siswa baru dapat menghasilkan data yang berlimpah mengenai data siswa. Data tersebut akan bertambah terus menerus dan apabila data tersebut diolah akan menghasilkan informasi yang tersembunyi yang dapat berguna bagi perguruan tinggi [1]. PPLP Dhyana Pura adalah sebuah lembaga pelatihan kerja yang berlokasi di Badung, Bali dimana Bali merupakan salah satu daerah pariwisata yang mendidik dan melatih secara profesional para tenaga kerja yang ingin bekerja dibidang perhotelan dan pariwisata. Tidak hanya siswa yang berasal dari SMK jurusan pariwisata saja yang mendaftarkan diri, ada juga siswa yang berasal dari SMA dan Kejar Paket C.

Untuk mengolah data tersebut agar menghasilkan informasi dibutuhkan teknik yaitu dengan *data mining*. *Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik metode tertentu sehingga menghasilkan informasi yang bermanfaat [2]. Pemilihan metode yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database* (KDD) secara keseluruhan. Adapun metode data mining yang digunakan adalah metode *K-Means Clustering*. Metode *K-Means Clustering* adalah sebuah metode yang populer digunakan dalam *data mining* untuk mendapatkan sekumpulan data yang nantinya akan dikelompokkan berdasarkan pada kemiripan karakteristik individu-individu data yang ada [3].

Dari beberapa penelitian yang telah dilakukan dapat disimpulkan bahwa algoritma *K-Means Clustering* dapat digunakan untuk mengelompokkan data dengan efisien dan efektif [4]. Hal ini lah yang mendasari dalam melakukan penelitian dalam hal penerapan algoritma *K-Means Clustering* pada data penerimaan siswa baru dari tahun 2012 sampai dengan 2016 dengan studi kasus langsung ke target yang dituju dalam hal ini melakukan studi kasus pada PPLP Dhyana Pura.

Berdasarkan data yang diperoleh dari pihak marketing PPLP Dhyana Pura, pada tahun 2012 dan 2013 terjadi peningkatan jumlah mahasiswa namun pada tahun 2014 terjadi penurunan jumlah mahasiswa. Namun pada tahun 2015 dan 2016 sudah kembali meningkatnya jumlah mahasiswa. Salah satu akibat berkurangnya jumlah pendaftar adalah kurangnya pengolahan data dengan tepat sehingga hal tersebut mempengaruhi dalam pengambilan keputusan dalam menentukan promosi di sekolah-sekolah. Pengolahan data yang tepat diharapkan dapat dijadikan acuan oleh pihak marketing kampus untuk melakukan promosi kampus agar mendapatkan lebih banyak mahasiswa setiap tahunnya [5].

2. Tinjauan Pustaka/ Penelitian Sebelumnya

Beberapa peneliti dahulu telah menerapkan metode *K-Means Clustering* sebagai penelitian dalam hal pengelompokan data diantaranya :

- 1) Ramadhani dalam penelitiannya yang berjudul “*Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro*”. Penulis dalam penelitian ini menceritakan bahwa metode *K-Means Clustering* cukup efektif diterapkan dan menjadi salah satu dasar pengambilan keputusan untuk menentukan strategi promosi UDINUS [1].
- 2) Nasari et al. dengan penelitian yang berjudul “*Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Potensi Utama)*”. Tujuan penulisan yang dijelaskan oleh penulis adalah menerapkan metode *K-Means Clustering* pada data penerimaan siswa baru tahun ajaran 2014/2015. Hasil penelitian pun di dapat bahwa asal sekolah SMA rata-rata jurusan diambil sistem informasi dan asal sekolah SMK rata-rata mengambil jurusan Teknik Informatika. Metode tersebut juga efektif dan efisien dalam mengelompokkan data [6].

2.1 Knowledge Discovery in Database (KDD)

Istilah *data mining* dan KDD dipakai untuk menjelaskan proses mencari informasi yang tersembunyi dalam suatu database yang besar. Sebenarnya data mining adalah salah satu tahapan dalam keseluruhan proses KDD. Proses KDD dapat dijelaskan sebagai berikut [6]:

1) Data Selection

Pemilihan data dari sekumpulan data yang dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi akan digunakan untuk proses data mining disimpan terpisah dari basis data operasional.

2) Pre-Processing/Cleaning

Sebelum dilakukan *data mining*, dilakukan proses membuang duplikasi data, memeriksa data inkonsisten dan memperbaiki kesalahan pada data seperti *typografi*.

3) Transformation

Coding merupakan transformasi pada data yang sesuai untuk diproses *data mining*. Proses *coding* dalam KDD sangat tergantung pada jenis atau pola informasi yang dicari dalam database.

4) Data Mining

Data mining adalah proses mencari pola atau informasi yang bermanfaat dalam data terpilih dengan menggunakan metode tertentu. Teknik dan metode dalam data mining sangat bervariasi. Pemilihan metode atau teknik yang tepat bergantung pada tujuan dan proses KDD secara keseluruhan.

5) Interpretation/Evaluation

Tahap ini merupakan bagian dari proses KDD disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.2 Algoritma K-Means Clustering

K-Means Clustering adalah salah satu metode data *clustering* non-hierarki yang mengelompokkan data dalam bentuk satu atau lebih *cluster*/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan menjadi satu *cluster* dan data yang memiliki karakteristik berbeda dikelompokkan dengan *cluster* yang lain sehingga data yang berada dalam satu *cluster* yang memiliki tingkat variasi yang kecil [6]. Algoritma *K-Means Clustering* hanya bisa digunakan oleh data yang atributnya numerik. Apabila data memiliki atribut yang bukan numerik, maka atribut akan ditransformasikan menjadi data numerik.

Langkah-langkah melakukan *clustering* dengan metode K-Means adalah [6]:

- 1) Menentukan nilai k sebagai jumlah *cluster*
- 2) Inisialisasi k sebagai titik pusat yang dapat dibangkitkan *random*.
- 3) Hitung jarak setiap data ke setiap titik pusat *cluster* dengan teori jarak *Euclidean* sebagai berikut :

$$d(P,Q) = \sqrt{\sum_{j=1}^p (x_j(P) - x_j(Q))^2} \quad (1)$$

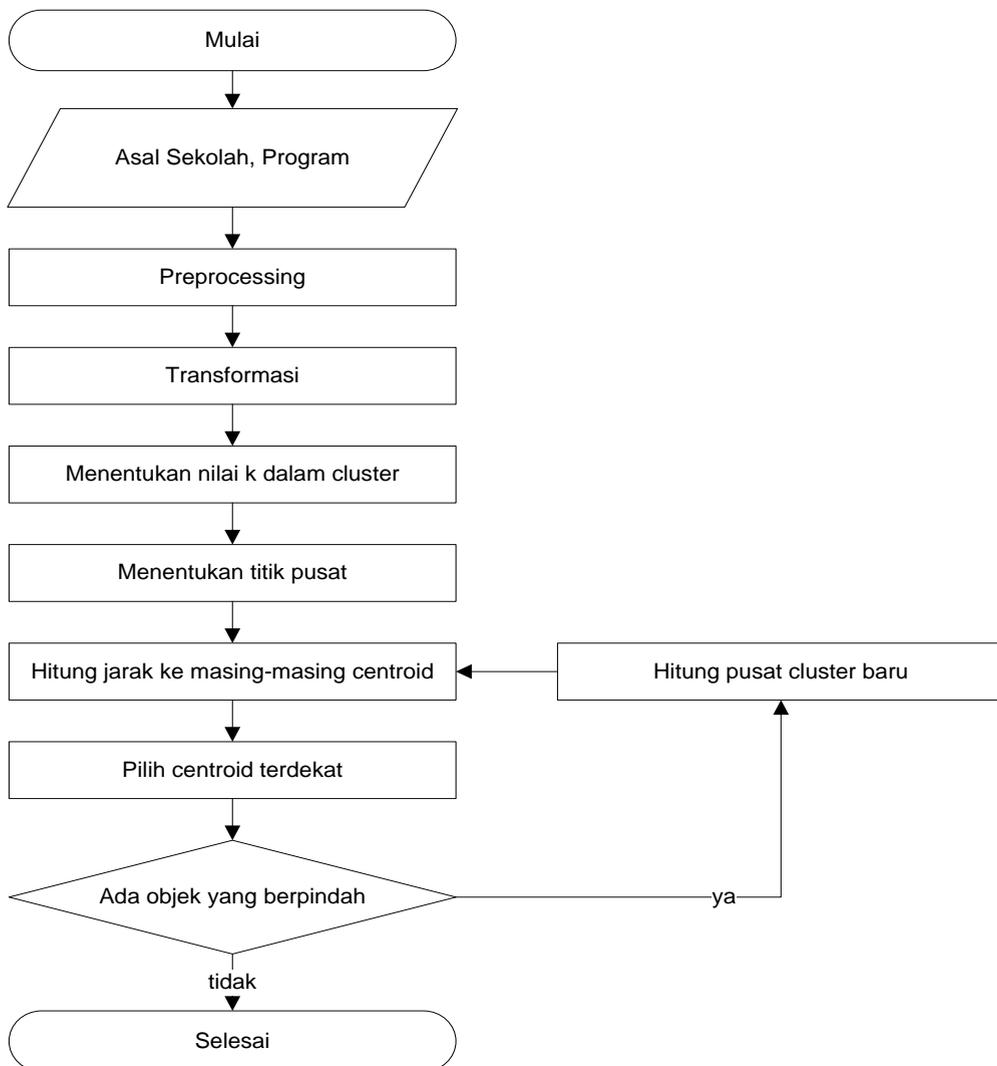
- 4) Kelompokkan data berdasarkan jarak terdekat data dengan titik pusat
- 5) Tentukan posisi titik pusat baru
- 6) Kembali ke langkah 3 jika posisi titik pusat baru dengan titik pusat lama tidak sama.

Karakteristik dari algoritma K-Means salah satunya adalah sangat sensitif dalam penentuan titik pusat awal kluster karena K-Means membangkitkan secara *random*. Pada saat membangkitkan awal titik pusat secara *random* mendekati solusi akhir pusat *cluster*, K-Means memiliki kemungkinan yang tinggi untuk menemukan

titik pusat *cluster* yang tepat. Sebaliknya jika awal titik pusat jauh dari solusi akhir pusat *cluster*, maka besar kemungkinan menyebabkan hasil yang tidak tepat. Akibatnya algoritma *K-Means Clustering* tidak menjamin hasil yang unik. Inilah yang menyebabkan K-Means sulit mencapai optimum global, tetapi hanya minimum lokal [7].

3. Metodologi

Metode yang digunakan dalam penelitian ini adalah metode K-Means Clustering seperti yang ditunjukkan pada Gambar 1. Beberapa parameter yang dimasukkan yaitu variabel asal sekolah dan program. Langkah pertama yang dilakukan adalah *preprocessing* untuk mengeliminasi *noise* (kekosongan nilai) pada data. Selanjutnya dilakukan transformasi data yaitu mengubah data yang dibutuhkan menjadi bentuk nominal atau angka karena algoritma *K-Means Clustering* hanya bisa memproses data berupa angka. Setelah itu baru dilakukan proses K-Means Clustering dimulai dari menentukan nilai k dalam *cluster*, menentukan titik pusat, menghitung jarak ke masing-masing *centroid*, memilih *centroid* terdekat dan menghitung pusat *cluster* baru hingga tidak ada objek yang berpindah.



Gambar 1 Gambaran Umum Prediksi Program menggunakan *K-Means Clustering*

3.1 Preprocessing Data

Sumber data dari penelitian ini adalah dari PPLP Dhyana Pura. Data mahasiswa baru diambil dengan jumlah data sebanyak 2231 data dari tahun 2012 sampai dengan 2016 dengan *field* yaitu No. Pendaftaran, Program Studi, Tahun Program, Provinsi dan Asal Sekolah. Berikut adalah beberapa sampel dari data pendaftaran siswa baru pada Tabel 1.

Tabel 1. Data Mahasiswa Baru

No	No. Pendaftaran	Tahun Program	Provinsi	Asal Sekolah
1.	201610001	2 tahun	Bali	SMA Negeri 1 Marga
2.	201610004	2 tahun	Bali	SMK Kencana Sakti
3.	201610005	2 tahun	Bali	SMA Negeri 1 Marga
4.	201610006	1 tahun	Bali	SMA Negeri 1 Petang
5.	201610007	1 tahun	Bali	SMA Negeri 1 Petang
6.	201610008	2 tahun	Bali	SMKN 1 Kuta Selatan
7.	2016100012	1 tahun	Bali	SMA 2 Negara
8.	2016100013	2 tahun	Bali	SMK P 1 Badung
9.	2016100014	2 tahun	Bali	SMA Negeri 1 Marga
10.	2016100015	1 tahun	Bali	SMA 6 Denpasar

Selanjutnya dilakukan *preprocessing* data yaitu menyaring data yang dibutuhkan serta membuang data yang tidak dibutuhkan dalam proses *data mining*.

3.2 Transformasi Data

Pada tahap ini dilakukan perubahan data. Tujuannya memilih data yang sesuai dan dapat diolah menggunakan metode *K-Means Clustering*. Variabel yang dipilih pada data mahasiswa baru tahun 2016/2017 adalah Asal Sekolah dan Tahun Program (Jurusan). Untuk variabel Asal sekolah dikelompokkan menjadi 3 kelompok yaitu SMA (Sekolah Menengah Atas) yang ditransformasi dengan nilai 1, dan SMK (Sekolah Menengah Kejuruan) yang ditransformasi dengan nilai 2 dan Kejar Paket C dengan nilai 3. Untuk Tahun Program dibagi menjadi 3 kelompok. Untuk pembagian transformasi program studi bisa dilihat pada Tabel 2. Hasil Transformasi dapat dilihat pada Tabel 3.

Tabel 2. Pembagian Transformasi Program

No.	Tahun Program	Nilai
1.	1 Tahun	1
2.	2 Tahun	2
3.	3 Tahun	3

Tabel 3. Tabel Hasil Transformasi

Asal Sekolah	Tahun Program
1	2
2	2

Asal Sekolah	Tahun Program
1	2
1	1
1	1
2	2
1	1
2	2
1	2
1	1

3.3 Pengolahan K-Means Clustering

Setelah dilakukan transformasi data, selanjutnya dilakukan pengolahan data menggunakan metode *K-Means Clustering*. Adapun langkah-langkah dalam melakukan *cluster* menggunakan *K-Means Clustering* adalah sebagai berikut :

- 1) Menentukan nilai k dari jumlah *cluster* yang ingin dibentuk. Jumlah *cluster* yang dibuat berjumlah 3 *cluster* berdasarkan asal sekolah yang memiliki 3 kelompok yaitu *cluster* 1 yang memiliki ijazah terakhir SMA, *cluster* 2 yang memiliki ijazah SMK dan *cluster* 3 yang memiliki ijazah kejar paket C.
- 2) Menentukan titik pusat awal pada setiap *cluster*. Dalam penelitian ini titik pusat awal ditentukan secara *random*. Hasil titik pusat awal dapat dilihat pada Tabel 4.

Tabel 4. Titik Pusat Awal Cluster

Titik Pusat Cluster	Asal Sekolah	Tahun Program
Cluster 1	1	1
Cluster 2	2	1
Cluster 3	3	1

- 3) Selanjutnya dilakukan perhitungan menggunakan perhitungan jarak *Euclidean* untuk menghitung setiap data ke dalam suatu *cluster* yang memiliki jarak paling dekat dengan titik pusat *cluster*

$$D(p,q)=\sqrt{(X1p - X1q)^2 + (X2p - X2q)^2 + \dots + (Xkp - Xkq)^2} \quad (2)$$

Dimana :

$D(p,q)$ = jarak data ke p ke pusat *cluster* q

Xkp = Data ke p pada atribut ke k

Xkq = Titik pusat ke j pada atribut ke k

Berikut adalah perhitungan jarak dari data mahasiswa pertama ke pusat *cluster* pertama dengan persamaan :

$$D(1,1) = \sqrt{(1 - 1)^2 + (2 - 1)^2} = \sqrt{(0)^2 + (1)^2} = 1$$

Dari hasil perhitungan tersebut dapat disimpulkan bahwa jarak data mahasiswa pertama dengan pusat *cluster* pertama adalah 1.

Perhitungan jarak data mahasiswa pertama ke pusat *cluster* kedua adalah

$$D(1,2) = \sqrt{(1-2)^2 + (2-1)^2} = \sqrt{(-1)^2 + (1)^2} = \sqrt{2} = 1,41421$$

Dari hasil perhitungan tersebut dapat disimpulkan bahwa jarak data mahasiswa pertama dengan pusat *cluster* kedua adalah 1,41421.

Perhitungan jarak data mahasiswa pertama ke pusat *cluster* ketiga adalah :

$$D(1,3) = \sqrt{(1-3)^2 + (2-1)^2} = \sqrt{(-2)^2 + (1)^2} = 5$$

Adapun hasil perhitungan setiap data pada iterasi 1 terlihat pada Tabel 5.

Tabel 5. Hasil Perhitungan Setiap Data *Cluster* Pada Iterasi 1

No Pendaftaran	Asal Sekolah	Program	Cluster 1	Cluster 2	Cluster 3	Jarak Terdekat
201610001	1	2	1	1,41421	2,23607	1
201610004	2	2	1,41421	1	1,41421	2
201610005	1	2	1	1,41421	2,23607	1
201610006	1	1	0	1	2	1
201610007	1	1	0	1	2	1
201610008	2	2	1,41421	1	1,41421	2
2016100012	1	1	0	1	2	1
2016100013	2	2	1,41421	1	1,41421	2
2016100014	1	2	1	1,41421	2,23607	1
2016100015	1	1	0	1	2	1

- 4) Setelah semua data diletakkan pada jarak terdekat, lalu menghitung jarak titik pusat berdasarkan rata-rata anggota hingga nilai pusat *cluster* tidak berubah lagi. Apabila nilai pusat *cluster* berubah, selanjutnya iterasi tetap dilakukan namun apabila nilai pusat *cluster* tidak berubah maka iterasi dihentikan.

$$\text{Untuk cluster 1 ada 885 data } C1 = \frac{1+1+1+\dots+1+1+1}{885}, \frac{2+2+2+\dots+1+2+1}{885}$$

Jadi hasil titik pusat untuk *cluster* 1 adalah (1, 1,63)

$$\text{Lalu cluster 2 ada 1192 data } C2 = \frac{2+2+2+\dots+2+2+2}{1192}, \frac{2+2+2+\dots+2+2+2}{1192}$$

Jadi hasil titik pusat untuk *cluster* 2 adalah (2, 1.73)

$$\text{Untuk cluster 3 ada 154 data } C1 = \frac{3+3+3+\dots+3+3+3}{154}, \frac{2+2+1+\dots+2+2+2}{154}$$

Jadi hasil titik pusat untuk *cluster* 3 adalah (3, 1,63)

Dari hasil perhitungan titik pusat yang baru, ternyata hasil tidak konvergen dengan hasil titik pusat yang lama, maka lanjutkan proses iterasi ke iterasi kedua. Berikut adalah perhitungan untuk iterasi kedua:

Untuk *cluster* 1 ada 885 data $C1 = \frac{1 + 1 + 1 + \dots + 1 + 1 + 1}{885}, \frac{2 + 2 + 2 + \dots + 3 + 3 + 3}{885}$

Jadi hasil titik pusat untuk *cluster* 1 adalah (1, 1,63)

Lalu *cluster* 2 ada 1192 data $C2 = \frac{2 + 2 + 2 + \dots + 2 + 2 + 2}{1192}, \frac{2 + 2 + 2 + \dots + 2 + 2 + 2}{1192}$

Jadi hasil titik pusat untuk *cluster* 2 adalah (2, 1,73)

Untuk *cluster* 3 ada 154 data $C1 = \frac{3 + 3 + 3 + \dots + 3 + 3 + 3}{154}, \frac{2 + 2 + 2 + \dots + 1 + 1 + 1}{154}$

Jadi hasil titik pusat untuk *cluster* 1 adalah (3, 1,63)

Adapun hasil perhitungan setiap data pada iterasi 2 terlihat pada Tabel 6.

Tabel 6. Hasil Perhitungan Setiap Data Cluster Pada Iterasi 2

No Pendaftaran	Asal Sekolah	Program	Cluster 1	Cluster 2	Cluster 3	Jarak Terdekat
201610001	1	2	0,37	1,0358	2,0339	1
201610004	2	2	1,0663	0,27	1,0662	2
201610005	1	2	0,37	1,0358	2,0339	1
201610006	1	1	0,63	1,2381	2,0969	1
201610007	1	1	0,63	1,2381	2,0969	1
201610008	2	2	1,0663	0,27	1,0662	2
2016100012	1	1	0,63	1,2381	2,0969	1
2016100013	2	2	1,0663	0,27	1,0662	2
2016100014	1	2	0,37	1,0358	2,0339	1
2016100015	1	1	0,63	1,2381	2,0969	1

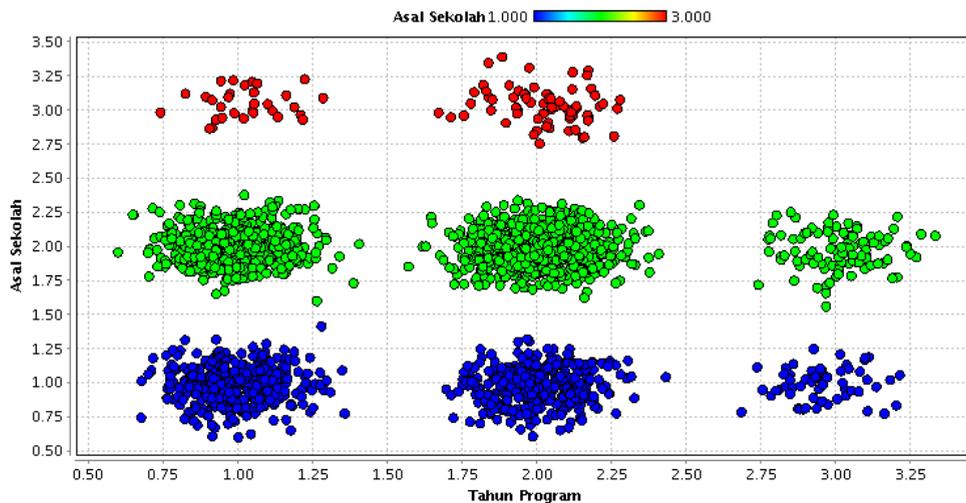
Setelah dilihat hasil iterasi kedua menghasilkan nilai titik pusat yang bersifat konvergen yaitu nilai tidak berubah, maka iterasi dihentikan. Berikut adalah gambar hasil perhitungan K-Means menggunakan *Excel* terlihat pada Gambar 2.

Tahun Program	Asal Sekolah	Cluster 1	Cluster 2	Cluster 3	Jarak Terdekat
1	2	0,37	1,035808863	2,033937069	1
1	2	0,37	1,035808863	2,033937069	1
1	1	0,63	1,238103388	2,096878633	1
1	1	0,63	1,238103388	2,096878633	1
1	1	0,63	1,238103388	2,096878633	1
1	2	0,37	1,035808863	2,033937069	1
1	1	0,63	1,238103388	2,096878633	1
1	2	0,37	1,035808863	2,033937069	1
1	3	1,37	1,61644672	2,424231837	1
1	1	0,63	1,238103388	2,096878633	1
1	1	0,63	1,238103388	2,096878633	1
1	1	0,63	1,238103388	2,096878633	1
1	1	0,63	1,238103388	2,096878633	1
1	1	0,63	1,238103388	2,096878633	1
1	1	0,63	1,238103388	2,096878633	1
1	2	0,37	1,035808863	2,033937069	1
1	1	0,63	1,238103388	2,096878633	1
1	1	0,63	1,238103388	2,096878633	1
1	2	0,37	1,035808863	2,033937069	1

Gambar 2 Hasil Perhitungan K-Means Menggunakan *Ms. Excel*

4. Hasil dan Pembahasan

Pada tahap ini penulis mengimplementasikan metode *K-Means Clustering* pada data mahasiswa baru di PPLP Dhyana Pura. Pada *cluster 1*, terlihat bahwa karakteristik mahasiswa pada *cluster 1* didominasi oleh mahasiswa dengan asal sekolah SMA lebih berminat mendaftar pada program 2 tahun dengan persentase 39.66%. Kemudian pada *cluster* kedua dapat dilihat karakteristik mahasiswa didominasi oleh mahasiswa dengan asal sekolah SMK memilih program 2 tahun dengan persentase 53,43%. Pada *cluster 3*, Kejar Paket C lebih dominan memilih program 2 tahun dengan persentase 6.9%. Itu berarti baik SMA, SMK, ataupun Kejar Paket C sama sama lebih dominan memilih program 2 tahun. Dari hasil perhitungan yang dilakukan sampai 2 iterasi dengan pusat *cluster 1* = (1, 1.63) dan *cluster 2* = (2, 1.73) dan *cluster 3* = (3, 1.63). Adapun hasil persebaran masing-masing *cluster* dapat dilihat pada Gambar 3.



Gambar 3 Hasil Pengelompokan K-Means menggunakan *Rapid Miner*

Pada Gambar 3, terlihat gambar pengelompokan berdasarkan masing-masing *cluster* dimana yang berwarna biru merupakan *cluster 1*, warna hijau merupakan *cluster 2* dan warna merah merupakan *cluster 3*. Adapun hasil persebaran *cluster* dapat dilihat pada Tabel 7.

Tabel 7. Hasil Persebaran *Cluster*

Cluster 1	Cluster 2	Cluster 3
<i>Cluster 1</i> terdiri dari 885 mahasiswa yang berasal dari SMA.	<i>Cluster 2</i> terdiri dari 1192 mahasiswa yang berasal dari SMK.	<i>Cluster 3</i> terdiri dari 154 mahasiswa yang berasal dari Kejar Paket C.
Program :	Program :	Program :
1 tahun = 365 orang	1 tahun = 388 orang	1 tahun = 56 orang
2 tahun = 493 orang	2 tahun = 734 orang	2 tahun = 98 orang
3 tahun = 31 orang	3 tahun = 70 orang	3 tahun = 0 orang

5. Kesimpulan

5.1 Simpulan

Setelah dilakukan pengelompokan data mahasiswa baru pada penelitian ini, dapat disimpulkan bahwa dalam penelitian ini menghasilkan tiga *cluster* dan dari hasil perhitungan yang dilakukan didapat pusat

cluster 1 = (1,1.63), *cluster 2* = (2 , 1.73) dan *cluster 3* = (3, 1.63). Jadi, mahasiswa yang berasal dari SMA lebih memilih program 2 tahun dengan persentase sebesar 39.66 % dari keseluruhan mahasiswa dengan pembagian mahasiswa yang berasal dari SMA memilih program 1 tahun sebesar 41 % memilih program 2 tahun sebesar 56 % dan yang memilih program 3 tahun sebesar 3%. Untuk mahasiswa yang berasal dari SMK lebih memilih program 2 tahun dengan persentase sebesar 53.43% dari keseluruhan mahasiswa dengan pembagian mahasiswa yang berasal dari SMK memilih program 1 tahun sebesar 32 % memilih program 2 tahun sebesar 62% dan yang memilih program 3 tahun sebesar 6%. Sedangkan untuk mahasiswa Kejar Paket C memilih program 1 tahun sebesar 36 % dan memilih program 2 tahun sebesar 64% dan tidak ada yang memilih program 3 tahun. Hasil dari penelitian ini bisa dijadikan referensi oleh pihak marketing kampus dalam melakukan promosi ke sekolah–sekolah.

5.2 Saran

Pengelompokan data mahasiswa baru PPLP Dhyana Pura sebaiknya dilakukan setiap tahun sehingga bisa dilihat program mana yang lebih diminati setiap tahun dan hal tersebut dapat dijadikan pedoman atau referensi pihak marketing kampus dalam merekrut calon mahasiswa baru setiap tahunnya sehingga setiap tahunnya jumlah mahasiswa semakin meningkat.

6. Daftar Rujukan

- [1] Ramadhani, R.D., Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro. pp 1-9, 2014
- [2] Muflikhah, L., Prediksi Nilai Mata Kuliah Mahasiswa Menggunakan Algoritma K-Apriori. Jurnal SISFO Vol 6 no. 02, pp 157-172, 2017
- [3] Prilianti, K.R., Aplikasi Text Mining Untuk Automasi Penentuan Tren Topik Skripsi dengan Metode K-Means Clustering. Cybermatika Vol. 2 No. 1, pp 1-6, 2014
- [4] Adrian, A.R., Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang. Vol 18 no. 1, pp 78-62, 2015
- [5] Ong, J.O., Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University. JUTI Vol. 12 no. 1, pp10-20, 2013
- [6] Nasari, F., Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Potensi Utama). Seminar Nasional Teknologi Informasi dan Multimedia, pp 73-78, 2015
- [7] Kusuma, D.T., Prototipe Komparasi Model Clustering Menggunakan Metode K-Means dan FCM Untuk Menentukan Strategi Promosi: Studi Kasus Sekolah Tinggi Teknik-PLN Jakarta. TICOM Vol. 3 no. 3, pp 1-10, 2015
- [8] Larose, T.D., Discovering Knowledge in Data An Introduction to Data Mining. Central Connecticut State University, 2005

