

Implementasi Algoritma C-Means dan Algoritma Mixture dalam Pengclusteran Data Mahasiswa Drop Out

Jhoanne Fredricka¹, Lena Elfianty², Jusuf Wahyudi*³

^{1,2,3}Universitas Dehasen Bengkulu, Jalan Meranti Raya No. 32 Sawah Lebar, Bengkulu, Indonesia

Email : Jusuf.wahyudi@unived.ac.id

Abstrak. Clustering merupakan salah satu metode machine learning dan termasuk dalam unsupervised learning. Tujuan dari clustering yaitu mencari pola data yang mirip sehingga memiliki kemungkinan dalam mengelompokkan data data yang mirip. Banyak Algoritma yang dapat digunakan untuk melakukan proses clustering data. Algoritma C-Means dan algoritma Mixture merupakan bentuk algoritma yang dapat digunakan dalam melakukan proses clustering data. Algoritma C-Means adalah suatu tehnik pengklusteran data yang mana keberadaan tiap – tiap titik data dalam suatu cluster ditentukan oleh derajat keanggotaan. Sedangkan Algoritma Mixture merupakan salah satu jenis data clustering dimana dalam permodelannya, data dalam satu kelompok diasumsikan terdistribusi sesuai dengan salah satu jenis distribusi statistik yang ada. Pada penelitian ini akan dilakukan perbandingan antara Algoritma C-Means dan Algoritma Mixture dalam hal klasterisasi data mahasiswa drop out, dengan melihat hasil performance vector yang di dihasilkan pada metode c-means yaitu avg.within centroid distance pada setiap cluster, Sedangkan pada algoritma mixture melihat hasil performance vector dan jumlah number of clusters.

Kata Kunci : Clustering, Algoritma C-Means, Algoritma Mixture

Abstract. Clustering is one of the machine learning methods and is included in unsupervised learning. The purpose of clustering is to look for similar data patterns so that it has the possibility of identical grouping data. Many algorithms can be used to process data clustering. The C-Means algorithm and the Mixture algorithm are forms of algorithms that can be used in carrying out the data clustering process. The C-Means algorithm is a data clustering technique in which the existence of each data point in a cluster is determined by the degree of membership. Meanwhile, the Mixture Algorithm is a type of data clustering. In its modeling, data in one group is assumed to be distributed according to one of the existing types of statistical distributions. In this study, a comparison will be made between C-Means Algorithm and Mixture Algorithm in terms of clustering dropout student data by looking at the results of the performance vector produced in the c-means method, namely avg. Within the centroid distance in each cluster, the mixture algorithm looks at the result performance vector and number of sets.

Keyword : Clustering, C-Means Algorithm, Mixture Algorithm

PENDAHULUAN

Pendidikan merupakan suatu kebutuhan primer yang sejak dini hingga dewasa hendaknya dirasakan oleh seluruh masyarakat. Hal ini sesuai dengan amanat UUD Negara Kita, anjuran agama dan menjadi penentu kemajuan suatu bangsa. Pendidikan juga merupakan variabel vital untuk pembangunan suatu bangsa. Suatu bangsa bisa maju dengan cepat dibandingkan dengan negara lain karena penyebaran pengetahuan (knowledge) yang merata keseluruh lapisan masyarakatnya. Institusi yang paling bertanggung jawab untuk penyebaran pengetahuan adalah institusi pendidikan.

Pada zaman sekarang ini pendidikan dimulai dari sejak anak berusia dini sampai ke jenjang yang paling tinggi, dari pendidikan Paud untuk anak usia dini hingga ke bangku kuliah baik itu untuk meraih gelar S1, S2, dan S3. Perguruan tinggi merupakan suatu institusi yang sudah pasti memiliki data yang tidak kecil volumenya. Database pada suatu perguruan tinggi menyimpan berbagai data, misalnya data akademik, data administrasi dan juga data mahasiswa. Dari data – data tersebut jika digali dengan tepat maka dapat diketahui pengetahuan baru yang dapat dikembangkan untuk diterapkan pada perguruan tinggi tersebut, diantaranya mengenai data mahasiswa drop out.

Pemahaman informasi mahasiswa yang berpotensi terkena drop out ini penting untuk diketahui dan dipahami. Pemahaman tersebut dapat dilakukan dengan menggali data – data yang dimiliki dan kemudian dilakukan pengelompokan terhadap hasil penggalan data tersebut sehingga memunculkan suatu pola atau kelompok mahasiswa yang berpotensi terkena drop out. Terdapat banyak alasan mengapa seorang mahasiswa bisa di drop out dari kampusnya antara lain mahasiswa tersebut tidak mengikuti aturan yang ada, terlalu lama masa



penyelesaian kuliah sehingga sudah melewati rentang waktu yang seharusnya, mahasiswa terlalu aktif dalam berkomunitas/UKM, kuliah sambil bekerja, bermasalah dengan kehidupan di kampus, masalah keuangan, nilai yang tidak memuaskan, melakukan pengajuan cuti dua semester berturut - turut dan lainnya. Pencegahan kegagalan adalah sangat penting bagi manajemen perguruan tinggi. Pengetahuan ini dapat digunakan dalam membantu pihak perguruan tinggi untuk lebih mengenal situasi para mahasiswanya dan dapat dijadikan sebagai pengetahuan dini dalam proses pengambilan keputusan untuk tindakan preventif dalam hal mengantisipasi mahasiswa drop out, untuk meningkatkan prestasi mahasiswa, untuk meningkatkan kurikulum, meningkatkan proses kegiatan belajar dan mengajar dan banyak lagi keuntungan lain yang bisa diperoleh dari hasil penambangan data tersebut.

Proses penclustering data sangat membantu dalam pengerjaan komputerisasi sehingga mempermudah pihak akademik untuk menarik data jika diperlukan suatu saat untuk melihat seberapa banyak jumlah mahasiswa yang drop out setiap tahunnya beserta dengan alasannya. Sehingga data – data tersebut bisa menjadi pemicu pihak akademik untuk mencariantisipasi mengurangi jumlah mahasiswa yang drop out. Pengelompokan data tersebut akan dilakukan menggunakan dua metode algoritma yaitu algoritma C-Means dan algoritma Mixture, karena algoritma C-means merupakan metode yang paling sederhana dan umum yang mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang cepat dan efisien, sedangkan algoritma Mixture dapat melakukan pengelompokan data – data didalam suatu dataset menjadi kelompok – kelompok data yang sebelumnya tidak terdefiniskan.

Penelitian ini bertujuan untuk melakukan pengelompokan data mahasiswa drop out menggunakan algoritma C-Means dan Algoritma Mixture serta mengevaluasi performansi Algoritma C-Means dan Algoritma Mixture dalam menghasilkan tingkat keakuratan alasan mengapa mahasiswa di drop out pada universitas.

METODOLOGI PENELITIAN

Logika *fuzzy* pertama kali diperkenalkan oleh Lothi Zadeh (UC Barkeley) pada tahun 1965, sebagai suatu cara matematis untuk menyatakan keadaan yang tidak menentu (samara) dalam kehidupan sehari – hari. Ide ini didasarkan pada kenyataan bahwa di dunia ini suatu kondisi sering di interpretasikan dengan ketidakpastian atau tidak memiliki ketetapan secara kuantitatif, misalnya : panas, dingin dan cepat. Dengan logika *fuzzy*, kita dapat menyatakan informasi – informasi yang samara tersebut (kurang spesifik), kemudian memanipulasinya dan menarik suatu kesimpulan dari informasi tersebut. Logika *fuzzy* telah banyak diterapkan dalam proses pengendalian suatu sistem, hal itu karena *fuzzy* dapat menyimpan dan mengaplikasikan pengetahuan – pengetahuan tentang suatu masalah. Pengetahuan – pengetahuan ini terdapat pada aturan – aturan logika *fuzzy*. Aturan – aturan inilah yang dijadikan tempat menyimpan pengetahuan tersebut, dan juga berfungsi sebagai domain dalam sistem pengambilan keputusan. Ada beberapa alasan mengapa orang menggunakan logika *fuzzy*, antara lain:

- Konsep logika *fuzzy* mudah dimengerti. Konsep matematis yang mendasari penalaran *fuzzy* sangat sederhana dan mudah dimengerti.
- Logika *fuzzy* sangat fleksibel.
- Logika *fuzzy* mampu memodelkan fungsi-fungsi *non linier* yang sangat kompleks.
- Logika *fuzzy* dapat membangun dan mengaplikasikan pengalaman–pengalaman para pakar secara langsung tanpa harus melalui proses pelatihan.
- Logika *fuzzy* memiliki toleransi terhadap data – data yang tidak tepat.
- Logika *fuzzy* didasarkan pada bahasa alami.

1. Algoritma C-Means

Fuzzy C-Means (FCM) adalah suatu tehnik pengklusteran data yang mana keberadaan tiap – tiap titik data dalam suatu cluster ditentukan oleh derajat keanggotaan. Konsep dasar FCM pertama kali adalah menentukan pusat cluster yang akan menandai lokasi rata – rata untuk tiap – tiap cluster. Dalam algoritma C-Means terdapat beberapa langkah yang harus dilakukan yaitu:

- Masukan data yang akan di cluster X, berupa matriks berukuran $n \times m$ (n = jumlah sample data, m = atribut setiap data). X_{ik} = data sample ke – i ($i = 1, 2, \dots, n$), atribut ke k ($k = 1, 2, \dots, m$).
- Menentukan :
 - Jumlah cluster = c
 - Pangkat = w



- c. Maximum Iterasi = MaxIter
- d. Error terkecil yang di harapkan = ξ
- e. Fungsi Objektif awal = $P_0 = 0$
- f. Iterasi awal = $t = 1$

3. Membangkitkan bilangan random μ_{ik} , $i = 1,2,3,\dots,n$; $k = 1,2,3,\dots,c$; sebagai elemen – elemen matriks partisi awal U. Menghitung jumlah tiap kolom :

$$Q_i = \sum_{k=1}^c \mu_{ik} \dots\dots\dots (1)$$

Dengan $j = 1,2,\dots,n$; menghitung:

$$\mu_{ik} = \frac{\mu_{ik}}{Q_i} \dots\dots\dots (2)$$

4. Menghitung pusat vektor tiap – tiap klaster untuk matriks partisi yaitu:

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \dots\dots\dots (3)$$

5. Menghitung fungsi objektif pada iterasi ke – t:

$$P_t = \sum_{i=1}^n \sum_{k=1}^c ([\sum_{j=1}^m (x_{ij} - v_{kj})^2] (\mu_{ik})^w) \dots\dots\dots (4)$$

6. Menghitung perubahan matriks partisi:

$$\mu_{ik} = \frac{[\sum_{j=1}^m (x_{ij} - v_{kj})^2]^{w-1}}{\sum_{k=1}^c [\sum_{j=1}^m (x_{ij} - v_{kj})^2]^{w-1}} \dots\dots\dots (5)$$

dengan $i = 1,2,\dots,n$; dan $k = 1,2, \dots,c$.

7. Cek kondisi berhenti:

- Jika $(|P_t - P_{t-1}| < \xi)$ atau $(t > \text{MaxIter})$ maka berhenti
- Jika tidak: $t=t+1$, mengulang langkah ke-4.

2. Algoritma Mixture

Algoritma *Mixture* merupakan salah satu jenis data clustering dimana dalam permodelannya, data dalam satu kelompok diasumsikan terdistribusi sesuai dengan salah satu jenis distribusi statistik yang ada. Mixture model juga mengasumsikan bahwa data merupakan suatu campuran bilangan yang terdistribusi secara statistik (Hermawati, 2013). Algoritma *Mixture* merupakan metode yang mempunyai optimasi yang sama dengan algoritma *C-Means* melalui proses optimazation dan maximization. Berbeda dengan algoritma *Hard C-Means* dan *Fuzzy C-Means*, perbandingan jumlah data yang tercakup didalam masing – masing *cluster* juga mempengaruhi hasil akhir dari suatu proses data *clustering*.

Distribusi statistik yang paling sering digunakan dalam data *clustering* menggunakan metode *mixture* adalah distribusi *Gaussiannormal*. Disamping karena kemudahan penurunan berbagai rumus yang diperlukan, kecenderungan umum yang ada pada saat melakukan observasi adalah bahwa data yang di dapatkan umumnya dalam keadaan terdistribusi secara normal.

HASIL DAN PEMBAHASAN

1. Analisa Sistem

Analisa sistem pengclusteringan data mahasiswa *drop out* pada Universitas Dehasen yaitu melalui penguraian dari suatu sistem utuh kedalam bagian komponen – komponen dengan maksud agar dapat mengidentifikasi dan melakukan evaluasi terhadap masalah – masalah dan hambatan – hambatan yang terjadi. Data yang digunakan adalah data mahasiswa yang telah di *drop out* pada Universitas Dehasen berdasarkan tahun angkatan yaitu dari tahun angkata 2014 sampai dengan tahun angkatan 2016.

Pengclusteringan data mahasiswa *drop out* dalam menggunakan *Algoritma C-Means* dan *Algoritma Mixture* ini akan dilakukan pengujian data dengan menggunakan masing – masing algoritma, dengan memperoleh hasil yang dapat dibandingkan sehingga didapat sebuah algoritma yang baik digunakan dalam pengclusteringan data mahasiswa *drop out* pada di Universitas Dehasen Bengkulu. Aplikasi yang digunakan dalam pengclusteringan data mahasiswa *drop out* menggunakan metode *Algoritma C-Means*



dan *Algoritma Mixture* adalah *Rapidminer Studio 7.6*, dimana dalam program *rapidminer* ini terdapat panel *Fuzzy Cluster* yang mempunyai kemampuan dalam mengelompokkan data sesuai dengan kelompok data yang telah ditentukan. Sedangkan kriteria yang akan digunakan pada kedua metode ini adalah jumlah cluster, banyaknya iterasi, toleransi dan bobot.

2. Hasil dan Pengujian

Pengujian yang dilakukan dengan menggunakan Bahasa Pemrograman *Rapid Miner Studio 7.6*, data yang digunakan di representasikan dalam bentuk proses pengujian dengan penerapan *algoritma Fuzzy Clustering Means (FCM)* dan *Mixture*. Pengujian dilakukan dengan menggunakan data yang tersimpan dalam *Microsoft Excel*. Dimana sampel data yang digunakan terdapat 50 data dan 6 variabel. Data yang terdapat pada file excel awal mulanya di import ke dalam *Rapidminer*, setelah itu menentukan indikator kedalam masing – masing atribut. Atribut di rubah menjadi label karena penentu, dan field no (nomor urut) tanda centang di hilangkan karena tidak diperlukan. Gambar 1 merupakan hasil pengaturan dan potongan dataset pada proses import kedalam *Rapidminer*.

No	Nilai IPK	Usia Masuk	Pekerjaan O	Penghasilan	Status sekol	Fakultas
1	1.790	19	0	3500000	0	ILMU KO...
2	2.290	20	1	5700000	1	ILMU KO...
3	1.910	23	1	2800000	1	EKONOMI
4	2.410	21	1	2000000	1	EKONOMI
5	2.390	22	0	3200000	1	EKONOMI

Gambar 1. Hasil pengaturan dan potongan dataset proses import Rapidminer

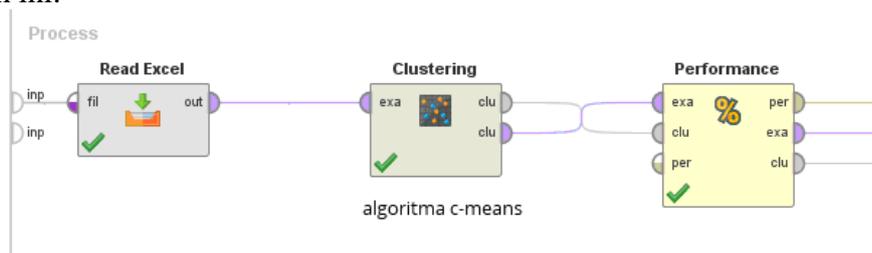
Hasil statistik setelah dataset berhasil ter-upload didapatkan hasil yang ditetapkan sebagai penerima sebanyak 50 examples dengan 3 special attributes dan 5 regular attributes. Detail statistik dapat dilihat pada gambar 2 di bawah ini:

id	Fakultas	cluster	Nilai IPK	Usia Masuk	Pekerjaan O...	Penghasilan...	Star
1	ILMU KOMPU...	cluster_2	1.790	19	0	3500000	0
2	ILMU KOMPU...	cluster_1	2.290	20	1	5700000	1
3	EKONOMI	cluster_0	1.910	23	1	2800000	1
4	EKONOMI	cluster_0	2.410	21	1	2000000	1
5	EKONOMI	cluster_2	2.390	22	0	3200000	1
6	EKONOMI	cluster_0	3.010	20	0	1800000	1
7	EKONOMI	cluster_2	1.790	19	1	3400000	0
8	ILMU KOMPU...	cluster_0	2.510	23	1	2300000	0
9	ILMU KOMPU...	cluster_2	2.290	23	1	3800000	1
10	ILMU KOMPU...	cluster_1	2.290	19	1	6500000	1
11	ILMU KOMPU...	cluster_1	1.790	20	0	5400000	1
12	ILMU KOMPU...	cluster_0	2.390	21	0	2800000	1
13	EKONOMI	cluster_2	2.390	23	0	4300000	1
14	EKONOMI	cluster_2	1.790	23	0	3500000	1
15	EKONOMI	cluster_2	2.290	23	1	3500000	0

Gambar 2. Detail statistik dataset pada Rapidminer

Clustering dengan Algoritma C-Means

Pembangunan model Algoritma *C-Means* menggunakan *Rapidminer Studio 7.6* ditunjukkan pada gambar 3 dibawah ini.



Gambar 3. Pembangunan Algoritma C-means

Dari gambar 3 di atas dapat dijelaskan bahwa konfigurasi dimulai dengan mengimport data dari excel ke *Rapidminer* dengan melakukan *import configuration* agar dapat di baca sistem. Kemudian dilakukan *cluster distanceperformance* terhadap *clustering* model *c-means* untuk melihat hasil pengclusterannya. Adapun hasil cluster model dapat dilihat pada gambar 4.

Cluster Model

```
Cluster 0: 20 items  
Cluster 1: 4 items  
Cluster 2: 26 items  
Total number of items: 50
```

Gambar 4. Hasil cluster model c-means

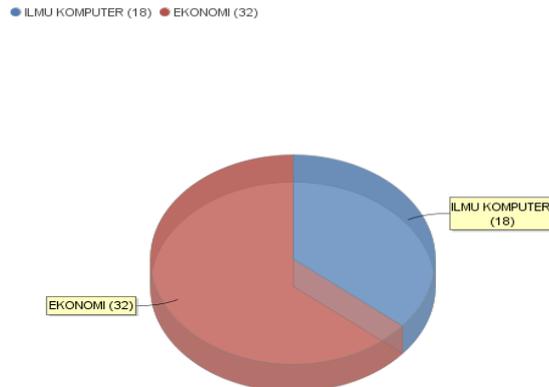
Dapat dilihat pada gambar 4 bahwa pengclusteran dengan model c-means menghasilkan 3 cluster dengan hasil cluster 0 terdiri dari 20 items, cluster 1 terdiri dari 4 items dan cluster 2 terdiri dari 26 items. Sedangkan hasil dari performance vektornya yaitu jumlah davies buldin nya sebesar -0,349 dimana dapat dilihat pada gambar 5.

PerformanceVector

```
PerformanceVector:  
Avg. within centroid distance: -127236923080.016  
Avg. within centroid distance_cluster_0: -126025000002.745  
Avg. within centroid distance_cluster_1: -221875000005.812  
Avg. within centroid distance_cluster_2: -113609467458.563  
Davies Bouldin: -0.349
```

Gambar 5. Hasil Performance Vector c-means

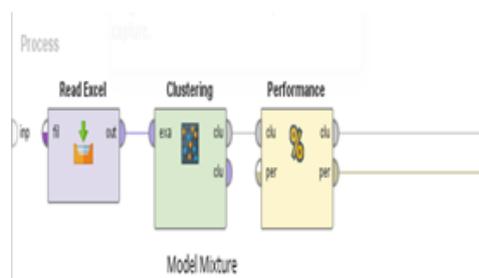
Adapun Hasil grafik dari pengclusteran data mahasiswa drop out menggunakan algoritma *C-Means* dapat dilihat pada gambar 6, dimana pada grafik tersebut dapat dilihat bahwa mahasiswa yang drop out lebih dominan pada fakultas ekonomi dan selanjutnya diikuti oleh fakultas ilmu komputer.



Gambar 6. Hasil Grafik Pengclusteran dengan C-Means

Clustering dengan Algoritma Mixture

Pembangunan model pada algoritma *mixture* menggunakan *Rapidminer Studio 7.6* dapat dilihat pada gambar 7.



Gambar 7. Pembangunan Algoritma Mixture

Dari gambar 7 di atas dapat dijelaskan bahwa konfigurasi dimulai dengan mengimport data dari *excel* ke *rapid miner* dengan melakukan *import configuration* agar dapat di baca sistem. Kemudian dilakukan *cluster count performance* terhadap *clustering model mixture* untuk melihat hasil pengclusterannya. Adapun hasil *cluster* model dapat dilihat pada gambar 8.

```
Cluster Model

Cluster 0: 17 items
Cluster 1: 29 items
Cluster 2: 4 items
Total number of items: 50

-----

cluster probabilities:
Cluster 0: 0.36515678103306454
Cluster 1: 0.5545145519459156
Cluster 2: 0.08032866702101962

cluster means:
Cluster 0: 2.191893394017597; 22.00301992148562; 0.5396390211357598; 2072961.970910545
Cluster 1: 2.2577681417495414; 21.493588237814; 0.6179974799179001; 3533987.500574226;
Cluster 2: 2.0422344903340686; 21.000480695404388; 0.5011729676033986; 5718959.3836376

cluster standard deviations:
Cluster 0: 9.367024441740959E10
Cluster 1: 1.4403446505749536E11
Cluster 2: 2.2963409233987097E11
```

Gambar 8. Hasil Cluster Model Mixture

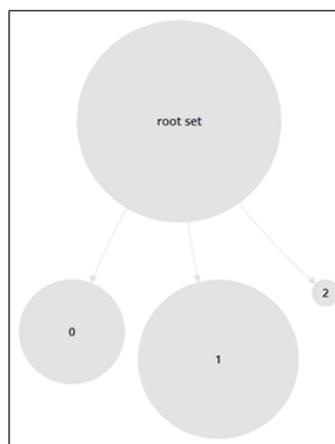
Dari gambar 8 dapat dilihat bahwa pada cluster model terdapat 3 cluster dimana terdiri dari cluster 0 ada 17 items, cluster 1 ada 29 items dan cluster 2 ada 4 items. Ketiga cluster tersebut memiliki jumlah *cluster probabilities*, *cluster means* dan *cluster standard deviations* yang berbeda – beda. Adapun *performance vector* yang dihasil pada model *mixture* yaitu tingkat *number of cluster* sebesar 3.000 dan tingkat *cluster number index* nya sebesar 0.940, hasil *performance vector* dapat dilihat pada gambar 9.

PerformanceVector

```
PerformanceVector:
Number of clusters: 3.000
Cluster Number Index: 0.940
```

Gambar 9. Hasil performance vector model mixture.

Bentuk grafik pengclusteran data mahasiswa drop out menggunakan algoritma *Mixture* dapat dilihat pada gambar 10, dimana pada grafik tersebut hanya terlihat berapa besar jumlah pengelompokan alasannya saja berdasarkan root set yang telah di tentukan.



Gambar 10. Grafik Hasil Pengclusteran dengan Mixture

Dari hasil performance yang telah dilakukan terhadap masing – masing model diperoleh hasil yang berbeda - beda sesuai dengan *performance* yang di gunakan, dan hasil pengclusteringnya pun berbeda antara menggunakan *metode c-means* dan *metode mixture*. Pada *metode c-means* setelah di lakukan pengclustering dapat di analisa bahwa faktor yang lebih dominan dapat terlihat dari faktor nilai sedangkan pada *metode mixture* di analisa terdapat faktor keuangan yang lebih dominan mempengaruhi terjadinya drop out.

Sehingga dapat dilihat perbandingan melakukan pengclustering data menggunakan algoritma *c-means* dan algoritma *mixture* yaitu dari waktu pengerjaan, hasil cluster dan hasil performance vector yang diberikan. Pada pengclustering menggunakan algoritma *c-means* hanya menghasilkan hasil clusternya saja tanpa menunjukkan nilai *probabilitas*, *means* dan *standart deviationsnya* dan menampilkan nilai *performance vector* dengan jumlah hasil *centroid distance* dari masing – masing cluster, sedangkan pada pengclustering dengan algoritma *mixture* menghasilkan jumlah cluster data dengan menampilkan nilai *probabilitas*, *means* dan nilai *standart deviations* dari masing – masing cluster, pada performance vectornya hanya menunjukkan no cluster dan no index cluster.

KESIMPULAN

Dari hasil pengujian dan pembahasan yang telah dijelaskan maka dapat di ambil kesimpulan antara lain:

1. *Performance vector* yang di hasilkan pada *metode c-means* yaitu *avg.within centroid distance* : - 127236923080,016, *avg. Within centroid distance cluster 0* : -126025000002,745, *avg.within centroid distance cluster 1*: - 221875000005,812 dan *avg.within centroid distance cluster 2*: - 113609467458,563 dengan besar *davies bouldin* : - 0,349, dimana terdapat faktor nilai yang lebih dominan dalam pengelompok. Sedangkan hasil *performance vector* pada *metode mixture* yaitu dengan jumlah *number of clusters* sebesar 3.000 dan *cluster number index* sebesar 0,940, terdapat faktor keuangan yang lebih dominan dalam pengelompokan.
2. Dari hasil pengelompokan yang telah dilakukan maka dapat di analisa bahwa yang menjadi faktor penyebab terjadinya mahasiswa drop out yaitu adanya perbedaan asal sekolah, penghasilan orang tua, pekerjaan orang tua dan juga faktor nilai.

DAFTAR PUSTAKA

- [1]. Adnyana, B, M, I., 2015, Segmentasi Citra Berbasis Clustering Menggunakan Algoritma Fuzzy C-Means dan CAT Swarm Optimization, Thesis, Magister Teknik Elektro, Universitas Udayana, Denpasar
- [2]. Afifah, N., dkk., 2016, Pengklasteran Lahan Sawah Di Indonesia Sebagai Evaluasi Ketersediaan Produksi Pangan Menggunakan Fuzzy C-Means, Jurnal Matematika “MANTIK”, ISSN: 2527-3159, E-ISSN: 2527-3167, Vol. 02, No. 01, Oktober 2016
- [3]. Agusta, Y, 2007, K-Means Penerapan, Permasalahan dan Metode Terkait, Journal Sistem dan Informatika Vol. 3, Februari 2007
- [4]. Gupta, S., Vinay, B., 2015, *GMMC: Gaussian Mixture Model Based Clustering Hierarchy Protocol In Wireless Sensor Network*, International Journal Of Scientific Engineering And Research (IJSER), ISSN: 2347-3878, Impact Factor (2014): 3.05, Vol. 3, Issue 7, July 2015
- [5]. Hermawati, F.A., 2013, Data Mining, Andi Yogya, Yogyakarta
- [6]. Khoirunnisak, M., Iriawan, N., Pemodelan Faktor – Faktor Yang Mempengaruhi Mahasiswa Berhenti Studi (Drop Out) Di Institut Teknologi Sepuluh November Menggunakan Analisis Bayesian Mixture Survival, Paper, Statistika FMIPA-ITS
- [7]. Kusrini, Emha, 2009, Algoritma Data Mining, Andi Yogya, Yogyakarta
- [8]. Kusumadewi, S., Purnomo, H., 2004, Aplikasi Logika Fuzzy Untuk Pendukung Keputusan, Graha Ilmu, Yogyakarta



- [9]. Mathew, M, 2015, *CCTS With Two Clustering Techniques For Time Synchronization In Wireless Sensor Network*, International Journal Of Scientific Engineering And Research (IJSER), ISSN: 2347-3878, Vol. 3, ISSUE. 9, September 2015
- [10]. Nurhayati, S., dkk., 2015, *Prediksi Mahasiswa Drop Out Menggunakan Metode Support Vector Machine*, Jurnal Ilmiah SISFOTENIKA, Vol. 5, No. 1, Januari 2015
- [11]. Nurjanah, dkk., 2014, *Implementasi Metode Fuzzy C-Means Pada Sistem Clustering Data Varietas Padi*, Kumpulan Jurnal Ilmu Komputer (KLIK), ISSN: 2406-7857, Vol. 01, No. 01, September 2014
- [12]. Purwanto, dkk., 2008, *Perancangan dan Analisis Algoritma*, Graha Ilmu, Yogyakarta
- [13]. Sari, H, L., Suranti, D., 2016, *Perbandingan Algoritma Fuzzy C-Means dan Algoritma Mixture Dalam Penclusteringan Data Curah Hujan Kota Bengkulu*, Seminar Nasional Aplikasi Teknologi Informasi (SNATi) Yogyakarta, ISSN: 1907-5022, 6 Agustus 2016
- [14]. Wijaya, K, A., *Implementasi Data Mining dengan Algoritma Fuzzy C-Means Studi Kasus Penjualan di UD Subur Baru*, Paper Fasilkom Udinus, Semarang

