

PENENTUAN KOMPETENSI MAHASISWA DENGAN ALGORITMA GENETIK DAN METODE *FUZZY C-MEANS*

Rosalia Hadi¹, I Ketut Gede Darma Putra², I Nyoman Satya Kumara³

Abstract—The Student's competence determination is a very important thing because it will ease the students in entering the work world which is appropriate with their competence in which competence has positive effect to someone's performance on his work. Data clustering technique has been used widely nowadays. However on the real world, the problems about the features used in data clustering process is still very much neglected, thus there are often features used are not quite relevant in the data clustering process. Therefore this research uses the combination between Genetic Algorithm and Fuzzy C-Means in the student's competence determination. The data used is the data of STIKOM Bali students in Information System programme. The genetic algorithm is used to determine the relevant features and Fuzzy C-Means is used in data clustering optimally. The combination between these two algorithms result in 77 features had by the dataset, but only 61 features are relevant and valid to use in clustering process. From the experiment result, the combination of Genetic Algorithm and Fuzzy C-Means gives better result than using only Fuzzy C-Means with the average percentage on the test conducted is 88.89%.

Intisari—Penentuan kompetensi mahasiswa merupakan hal yang penting karena akan memudahkan mahasiswa untuk memasuki dunia kerja yang sesuai dengan kompetensinya dimana kompetensi akan berpengaruh positif terhadap kinerja seseorang dalam pekerjaannya. Teknik klusterisasi data telah dipakai dengan sangat luas sekarang ini. Namun pada kenyataannya, masalah-masalah tentang fitur yang digunakan dalam proses klusterisasi data masih sangat diabaikan, sehingga kerap kali terjadi penggunaan fitur yang kurang relevan dalam proses klusterisasi data tersebut. Untuk itu penelitian ini menggunakan penggabungan antara Algoritma Genetik dan *Fuzzy C-Means* dalam penentuan kompetensi mahasiswa. Data uji yang digunakan adalah data mahasiswa program studi Sistem Informasi STIKOM Bali. Algoritma Genetik digunakan dalam menentukan fitur yang relevan dan *Fuzzy C-Means* digunakan dalam melakukan klusterisasi data secara optimal. Kombinasi antara dua buah algoritma ini menghasilkan dari 77 fitur yang dimiliki oleh dataset, hanya 61 fitur yang relevan dan valid untuk digunakan dalam proses clustering. Hasil percobaan yang dilakukan, penggabungan Algoritma Genetik dan *Fuzzy C-Means* memberikan hasil yang lebih baik dibandingkan *Fuzzy C-Means* saja dengan persentase rata-rata kesesuaian pada pengujian yang dilakukan adalah sebesar 88.89%.

Kata Kunci—kompetensi mahasiswa, pemilihan fitur, klusterisasi data, algoritma genetik, *fuzzy c-means*

I. PENDAHULUAN

Kompetensi adalah suatu kemampuan untuk melaksanakan atau melakukan suatu pekerjaan yang dilandasi atas keterampilan dan pengetahuan serta didukung oleh sikap kerja yang dituntut oleh pekerjaan tersebut [1]. Penentuan kompetensi mahasiswa merupakan hal yang penting karena akan memudahkan mahasiswa untuk memasuki dunia kerja yang sesuai dengan kompetensinya. Mahasiswa akan memiliki landasan untuk bekerja baik sebagai karyawan ataupun sebagai pengusaha berdasarkan kompetensi yang dimilikinya. Dengan adanya kompetensi akan berpengaruh positif terhadap kinerja karyawan pada suatu perusahaan [1].

Klusterisasi dapat diterapkan pada berbagai bidang, misalnya pemasaran. Klusterisasi dapat digunakan sebagai metode untuk mengelompokkan pelanggan yang memiliki kesamaan dalam perilaku keseharian belanja [2]. Penerapan klusterisasi dalam bidang biologi dapat dilakukan salah satunya dalam pengelompokan tumbuhan ataupun hewan. Dalam bidang informatika, teknik klusterisasi sangat banyak digunakan antara lain pada *data mining*, *information retrieval*, pengolahan citra, dan lain sebagainya. Contoh lainnya juga dalam hal penentuan kompetensi mahasiswa. Metode yang dapat digunakan untuk pemecahan studi kasus tersebut adalah dengan *Fuzzy C-Means*.

Metode yang digunakan untuk klusterisasi data pada penelitian ini adalah metode *Fuzzy C-Means* yang memiliki batasan *fuzzy*. Pemodelan *fuzzy* berarti pemodelan yang lebih fleksibel yang keanggotaan berada dalam interval 0 sampai 1. Alasan lainnya untuk menggunakan model *fuzzy* adalah bahwa lebih mudah untuk memecahkan komputasi, karena pada kenyataannya, model non-*fuzzy* sering menghasilkan pencarian yang melelahkan di ruang besar, karena variabel kunci hanya dapat mengambil nilai-nilai 0 dan 1. Sebagai contoh, pada kasus kompetensi mahasiswa. Seorang mahasiswa mungkin saja berada/memiliki lebih dari 1 (satu) buah kompetensi.

Penelitian terdahulu membahas proses klusterisasi data menggunakan algoritma *K-Means* dan Algoritma Genetik [3] dimana proses klusterisasi data pada metode *K-Means*, pembentukan partisi dilakukan sedemikian rupa sehingga setiap objek berada tepat pada satu partisi. Akan tetapi pada suatu saat secara alami hal ini sering tidak dapat dilakukan untuk menempatkan suatu objek tepat pada satu partisi, karena sebenarnya objek tersebut bisa saja terletak di antara dua atau lebih partisi yang lain. Sehingga perlu dilakukan klusterisasi data dengan menggunakan *fuzzy clustering*.

¹Mahasiswa Magister Teknik Elektro, Pascasarjana Universitas Udayana, Jl. PB. Sudirman, Denpasar, Bali INDONESIA (e-mail: rosaliahadi87@gmail.com)

²Staf Pengajar Teknologi Informasi, Fakultas Teknik Universitas Udayana, Jl. Jalan Kampus Bukit Jimbaran 80361 INDONESIA (e-mail: ikgdarmaputra@gmail.com)

³Staf Pengajar Magister Teknik Elektro, Pascasarjana Universitas Udayana, Jl. PB. Sudirman, Denpasar, Bali



Penelitian yang dilakukan oleh Laetitia membahas penyelesaian masalah seleksi fitur dengan menggunakan algoritma genetik (GA) dan menggunakan algoritma *K-Means* untuk *clustering* [4]. Penelitian yang dilakukan oleh Kyoung menerangkan Algoritma genetik (GA) yang digunakan untuk pemilihan variabel dan untuk menentukan jumlah cluster pada kasus nyata segmentasi pemasaran [3]. Penelitian yang dilakukan oleh Bashar membahas tentang algoritma clustering Algoritma Genetik dan *K-Means* pada kasus dunia nyata untuk segmentasi pasar dalam perdagangan elektronik dan menemukan bahwa Algoritma Genetik dan *K-Means* mengakibatkan segmentasi yang lebih baik dari algoritma clustering tradisional *K-Means* dan SOM (*Self Organizing Map*) [5]. Penelitian oleh Prabha membahas tentang peningkatan kualitas *cluster* dari *K-Means clustering* menggunakan algoritma genetika. Algoritma yang diusulkan diuji dalam domain medis dan hasil penelitian didapatkan bahwa algoritma yang diusulkan mencapai hasil yang lebih baik daripada algoritma *K-Means* konvensional [6].

Teknik klusterisasi data telah dipakai dengan sangat luas sekarang ini. Namun ternyata pada kenyataannya, masalah-masalah tentang fitur yang digunakan dalam proses *clustering* masih sangat diabaikan, sehingga kerap kali terjadi penggunaan fitur-fitur yang kurang relevan dalam proses klusterisasi data tersebut. Hal ini dapat mengakibatkan hasil klusterisasi juga akan menjadi kurang optimal. Oleh karena itu, dibutuhkan proses pemilihan fitur sebelum klusterisasi data.

Pemilihan fitur merupakan kegiatan yang termasuk ke dalam *preprocessing* yang bertujuan untuk memilih fitur yang berpengaruh dan mengesampingkan fitur yang tidak berpengaruh dalam suatu kegiatan pemodelan atau penganalisaan data. Terdapat banyak alternatif yang dapat digunakan dan juga dilakukan proses mencoba-coba untuk mencari fitur yang cocok [4]. Metode *K-Means* digunakan pada proses clustering dimana setiap objek berada tepat pada satu partisi. Akan tetapi pada studi kasus pemilihan kompetensi, seorang mahasiswa sangat memungkinkan berada pada beberapa kompetensi, sehingga metode ini kurang cocok untuk digunakan. Sedangkan *fuzzy clustering* mampu menangani permasalahan yang ada, yaitu menggunakan metode *Fuzzy C-Means* dimana setiap object dapat berada pada beberapa partisi berbeda.

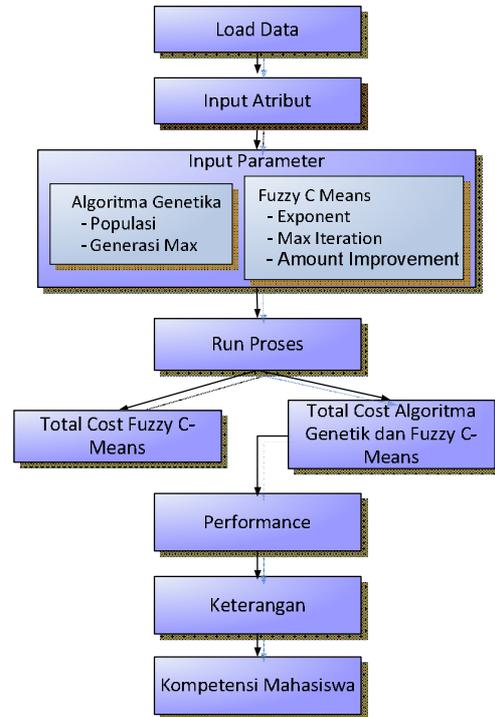
Algoritma Genetik telah banyak digunakan dalam penelitian serupa. Algoritma genetik berhasil dalam memilih dan menemukan fitur genetik dan faktor lingkungan yang terlibat dalam penyakit multifaktorial seperti obesitas dan diabetes [4]. Metode Algoritma Genetik juga digunakan dalam menyelesaikan masalah variabel yang tidak relevan dalam hal segmentasi pemasaran [2]. Penelitian [7] menyimpulkan bahwa algoritma genetik dapat digunakan dalam menentukan nilai awal dari *centroid cluster*. Kinerja *centroid cluster* diawali dengan menggunakan algoritma genetika memberikan hasil yang lebih baik daripada menggunakan nomor acak.

Berdasarkan latar belakang tersebut, maka pada penelitian ini dilakukan penggabungan antara Algoritma Genetik dan *Fuzzy C-Means* dalam penentuan kompetensi mahasiswa. Algoritma Genetik digunakan dalam menentukan fitur yang

valid dan metode *Fuzzy C-Means* digunakan dalam melakukan pengelompokkan kompetensi secara optimal. Kombinasi antara dua buah algoritma ini menghasilkan nilai *fitness* dan tingkat konvergensi yang lebih baik serta membentuk hasil cluster yang semakin akurat dalam proses klusterisasi data. Batasan yang digunakan adalah batasan *soft*, yaitu dengan menggunakan *fuzzy* dimana memungkinkan sebuah data berada pada beberapa *cluster* yang berbeda.

II. METODE PENELITIAN

A. Gambaran Umum Sistem



Gambar 1: Gambaran umum sistem

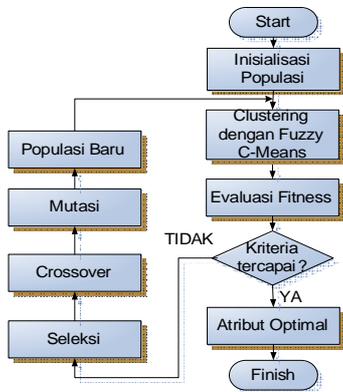
Gambar 1 menunjukkan gambaran umum sistem yang dibuat meliputi beberapa proses yaitu input data uji, input 3 atribut untuk sumbu x, y dan z, input parameter Algoritma Genetik dan parameter *Fuzzy C-Means*. Selanjutnya ada eksekusi program untuk metode yang digunakan yaitu hasil penggabungan Algoritma Genetika dan *Fuzzy C-Means* serta hasil dengan menggunakan *Fuzzy C-Means* saja. Evaluasi fitness menggunakan *Objective Function*.

B. Data Uji

Penelitian ini akan menggunakan data uji berupa data mahasiswa program studi STIKOM Bali yang telah menempuh delapan semester. Data yang digunakan berjumlah 300 dengan jumlah variabel sebanyak 77 variabel. Studi kasus yang diangkat dalam penelitian ini adalah mengenai penentuan kompetensi mahasiswa, dimana seorang mahasiswa mungkin saja memiliki lebih dari satu kompetensi.

C. Penggabungan Algoritma Genetik dan *Fuzzy C-Means*

Penggabungan antara Algoritma Genetik dan *Fuzzy C-Means* dalam menentukan fitur yang valid dan melakukan pengelompokan secara optimal dalam hal penentuan kompetensi mahasiswa. Dengan adanya kombinasi antara dua buah algoritma ini akan dihasilkan nilai *fitness* dan tingkat konvergensi yang lebih baik serta membentuk hasil cluster yang semakin akurat dalam proses klusterisasi (*clustering*).



Gambar 2: Clustering K-Prototype dengan Algoritma Genetik

1. Inisialisasi Populasi Awal

Langkah awal yang dilakukan pada penelitian ini merupakan inisialisasi populasi. Inisialisasi populasi merupakan pembuatan kromosom awal yang dilakukan secara acak sejumlah populasi yang dimasukkan oleh user. *Populasi awal P(0) dipilih secara acak menggunakan kromosom biner*. Kromosom biner merupakan kromosom yang disusun dari gen-gen yang bernilai 0 dan 1.

2. Clustering menggunakan *Fuzzy C-Means*

Setiap kromosom yang terbentuk pada fase inisialisasi populasi akan melewati proses *Clustering*. *Fuzzy C-Means* digunakan untuk menentukan pengelompokan yang akan terbentuk. Proses ini dimulai dari melakukan inisialisasi jumlah *cluster* awal. Selanjutnya dilakukan alokasi data sesuai jumlah *cluster*. Dilakukan proses perhitungan centroid dari masing-masing *cluster*. Apabila perubahan dari *membership function* melebihi nilai *thresholdnya*, maka akan dilakukan proses perhitungan ulang. *Cluster* ditentukan atas kedekatan jarak tiap-tiap data dengan centroidnya.

3. Evaluasi Fitness Menggunakan *Objective Funtion*

Dalam penelitian ini, *Clustering Criterion* yang digunakan adalah *objective function*, *Objective function* digunakan untuk mencari lokasi nilai terbaik pada tiap cluster terhadap masing-masing fitur. Adapun persamaan *objective function* sebagai berikut:

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m |x_i - c_j|^2, \quad 1 \leq m < \infty \tag{1}$$

dimana *m* adalah bilangan *real* yang lebih besar dari 1, *u_{ij}* adalah derajat keanggotaan dari *x_i* di kluster *j*, *x_i* adalah

dimensi data yang diukur, *c_j* adalah pusat dimensi cluster, dan *||*||* adalah setiap aturan yang mengungkapkan kesamaan antara pusat dan data diukur. Partisi *Fuzzy* dilakukan melalui optimasi berulang dari *objective function* yang ditunjukkan di atas, dengan *update* keanggotaan *u_{ij}* dan pusat *cluster c_j* oleh:

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}, \quad c_j = \frac{\sum_{i=1}^N u_{ij}^m \cdot x_i}{\sum_{i=1}^N u_{ij}^m} \tag{2}$$

Iterasi ini akan berhenti ketika $\max_j \{ |u_{ij}^{k+1} - u_{ij}^{(k)}| \} < \epsilon$ di mana kriteria terminasi antara 0 dan 1, sedangkan *k* adalah langkah-langkah iterasi. Prosedur ini konvergen ke minimum lokal atau titik *J_m*.

Algoritma ini terdiri dari langkah-langkah berikut:

- 1) Inisialisasi *U*=[*u_{ij}*] matriks, *U*⁽⁰⁾
- 2) Pada *k*-langkah: menghitung pusat vektor *C*^(*k*)=[*c_j*] dengan *U*^(*k*)

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m \cdot x_i}{\sum_{i=1}^N u_{ij}^m} \tag{3}$$

- 3) *Update U*^(*k*), *U*^(*k*+1)

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \tag{4}$$

- 4) Jika $\|U^{(k+1)} - U^{(k)}\| < \epsilon$ kemudian berhenti; jika tidak kembali ke langkah 2.

4. Seleksi

Proses seleksi, setelah melalui evaluasi *fitness* maka proses selanjutnya memilih individu yang akan digunakan dalam proses *crossover* dan selanjutnya akan dimutasikan. Proses seleksi disini menggunakan teknik *Roulette-Wheel*. Pada proses ini menghasilkan dua buah *parent* yang disebut *P1* dan *P2*. *Parent* ini menghasilkan keturunan baru untuk diproses ke tahap *crossover*.

5. *Crossover*

Operator *crossover* yang digunakan dalam penelitian ini menggunakan satu titik potong, Pada *crossover* satu titik (*one point crossover*), satu titik sepanjang kromosom dipilih secara *random*. Segmen induk dari titik point ke kiri atau kekanan ditukar untuk menghasilkan individu baru. sehingga nantinya setiap anaknya akan terdiri dari sebagian pasangan *parent*-nya.

6. Mutasi

Tahapan mutasi ini diperlukan dalam hal mengubah nilai yang ada pada gen didalam kromosom dimana pada proses mutasi tersebut menghasilkan individu baru dengan merubah satu atau lebih gen pada satu individu. Individu yang terpilih untuk proses mutasi dapat

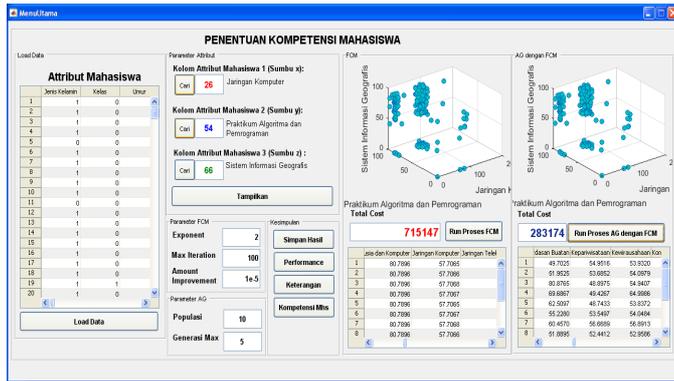


dilakukan dengan membandingkan nilai probabilitas mutasinya dengan probabilitas mutasi yang telah ditentukan atau dapat dipilih secara acak (*random*).

III. HASIL DAN PEMBAHASAN

A. Hasil

Untuk mempermudah dalam proses pengelompokan data maka dibuat sebuah aplikasi yang dapat digunakan dalam pengujian terhadap metode yang digunakan. Aplikasi pengujian tersebut dapat dilihat pada Gambar 3.



Gambar 3: Aplikasi utama

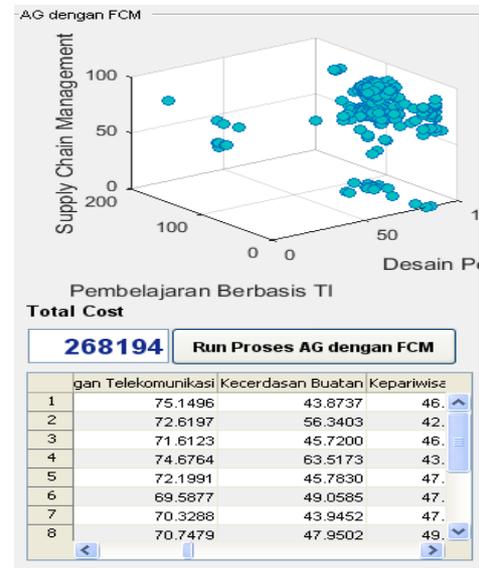
Aplikasi penentuan kompetensi mahasiswa dengan algoritma genetik dan *fuzzy c-means* pada Gambar 3, terdapat beberapa fitur yang diantaranya:

1. *Load Data*, digunakan untuk memanggil data yang akan diolah, kemudian ditampilkan dalam bentuk Grid.
2. Penginputan atribut, yaitu pemilihan tiga buah atribut bebas dari total 77 jumlah atribut yang ada. Parameter atribut ini digunakan sebagai tiga buah sumbu yang ada yaitu sumbu x, y dan z untuk penggambaran grafik tiga dimensi.
3. Penginputan parameter, yaitu pengisian parameter algoritma genetik dan parameter *fuzzy c-means*. Total Populasi dapat mempengaruhi hasil dan waktu yang dibutuhkan untuk melakukan proses, apabila total populasi yang digunakan semakin banyak dalam satu generasi, maka akan menghasilkan solusi yang lebih baik. Generasi Max mewakili jumlah populasi yang telah ditentukan pada tahap pengisian total populasi, sehingga menghasilkan kromosom yang lebih variatif dalam proses fitness. *Exponent* digunakan untuk menyatakan pangkat dari *matrix* yang dihasilkan oleh membership dari tiap *cluster*, yang bertujuan mengetahui besar kecilnya hubungan antar *membership* dengan masing-masing pusat *cluster*. *Max iteration* merupakan nilai maksimum yang ditentukan pada saat melakukan perulangan dalam mencari pusat *cluster* pada tiap member. *Amount improvement* adalah nilai minimum peningkatan dalam pencarian pusat *cluster* untuk masing-masing *member*.
4. Eksekusi program

Hasil penggabungan algoritma genetik dan *fuzzy c-means* dibandingkan dengan hasil dari *fuzzy c-means* saja dan dihasilkan total jarak antara *cluster* yang terbentuk dilihat dari total *cost* yang digunakan sebagai nilai pengukurannya. Jika nilai total *cost* yang diperoleh semakin kecil berarti jarak antara semua obyek dengan *cluster*nya semakin dekat, Sehingga bisa dikatakan bahwa *cluster* yang terbentuk semakin kompak dan bagus.

B. Pembahasan

Untuk mengukur hasil dari metode yang digunakan yaitu algoritma genetik dan *fuzzy c-means*, maka dilakukan perbandingan dengan hanya menggunakan *fuzzy c-means* saja, tanpa penggabungan algoritma genetik. Pengujian dilakukan dengan metode pengukuran hasil *Clustering* disebut dengan *Clustering Criterion*. *Clustering Criterion* inilah yang akan dijadikan nilai *fitness* dari setiap kromosom yang telah dievaluasi. *Clustering Criterion* yang digunakan adalah *cost function* untuk mengukur jarak setiap kelompok data yang terbentuk. Total jarak tiap *cluster* yang nantinya dihasilkan disebut Total *Cost*. Semakin kecil nilai total *cost* yang dihasilkan maka semakin bagus *cluster* yang terbentuk.



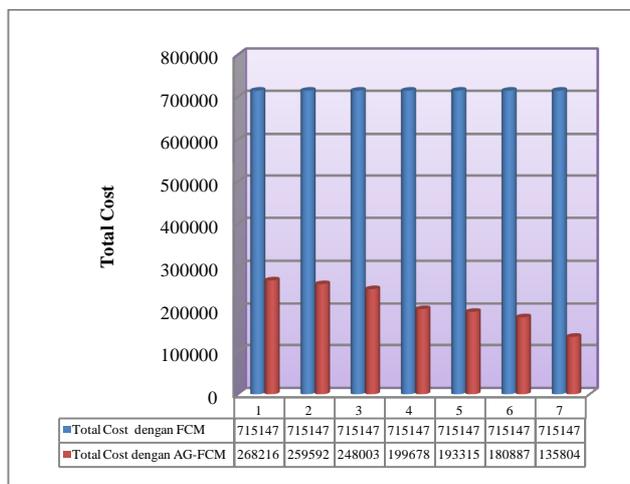
Gambar 4: Hasil penggabungan AG dan Fuzzy C-Means

Pengujian dilakukan sebanyak 7 kali, dimana pengujian dilakukan dengan menaikkan nilai total populasi dan generasi max. Hasil pengujian dapat dilihat pada Tabel 1.

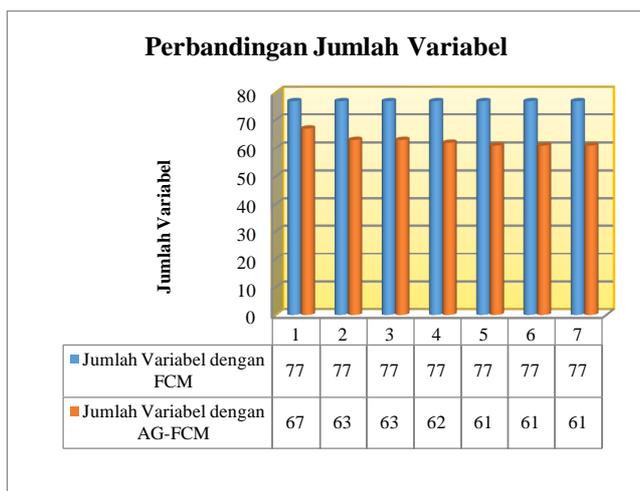
Percobaan dilakukan dengan menaikkan nilai total populasi dan generasi max. Hal ini akan mengakibatkan jumlah komputasi yang bertambah. Dengan menaikkan nilai total populasi dan generasi max akan memberikan solusi yang lebih baik, namun waktu yang dibutuhkan untuk pengolahan data juga semakin bertambah. Hasil perbandingan yang diperoleh dapat ditabulasikan dalam bentuk diagram batang, seperti yang terlihat pada Gambar 5.

TABEL I
HASIL PENGUJIAN

Percobaan	Total Populasi	Generasi Max	FCM		Algoritma Genetik - FCM	
			Total Cost	Jumlah Variabel	Total Cost	Jumlah Variabel
1	5	5	715.147	77	268.216	67
2	10	10	715.147	77	259.592	63
3	15	15	715.147	77	248.003	62
4	50	25	715.147	77	199.678	62
5	100	50	715.147	77	193.315	61
6	150	50	715.147	77	180.887	61
7	150	100	715.147	77	135.804	61



Gambar 5: Grafik pengujian perbandingan total cost

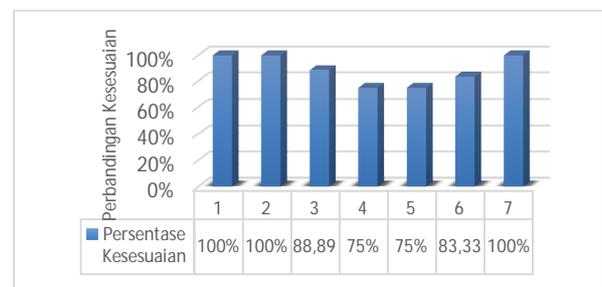


Gambar 6: Grafik pengujian perbandingan jumlah variabel

Gambar 5 menunjukkan bahwa hasil penggabungan Algoritma Genetik dan *Fuzzy C-Means* lebih baik dibandingkan dengan hanya *Fuzzy C-Means* saja. Hal ini terlihat dari total *cost* penempatan *object* pada *cluster* yang lebih rendah, yang artinya biaya yang dibutuhkan semakin kecil atau jarak yang semakin pendek. Hal ini berarti tingkat kesamaan ciri atau karakteristik dari setiap kelompok yang terbentuk menjadi lebih tinggi.

Gambar 6 menunjukkan hasil penggabungan Algoritma Genetik dan *Fuzzy C-Means* yang memperlihatkan terjadi penurunan jumlah variabel. Hal ini berarti pemilihan fitur berhasil dilakukan, dimana memilih fitur yang berpengaruh dan mengesampingkan fitur yang tidak berpengaruh.

Selanjutnya dilakukan pengujian data dengan kuesioner terhadap mahasiswa Program Studi Sistem Informasi STIKOM Bali. Metode kuesioner yang digunakan adalah kuesioner terstruktur yang terbuka melalui pengambilan *sample (sampling)*. *Sample* yang diambil berupa mahasiswa program studi Sistem Informasi yang memiliki status aktif dan sedang menempuh semester 8 (delapan). Data diinputkan ke sistem dan dibandingkan dengan hasil kuisisioner mahasiswa.



Gambar 7: Grafik pengujian kuesioner

Gambar 7 menunjukkan hasil uji coba yang dilakukan dengan membandingkan hasil kuesioner terhadap 7 orang responden dengan hasil sistem. Hasil memperlihatkan bahwa tingkat kesesuaian minimal sebesar 75% dengan rata-rata kesesuaian sebesar 88.89%.

IV. KESIMPULAN

Berdasarkan hasil beberapa percobaan yang dilakukan dapat diambil kesimpulan bahwa penggabungan Algoritma Genetik dan *Fuzzy C-Means* yang diusulkan mampu menentukan kompetensi mahasiswa yang lebih baik dibandingkan *Fuzzy C-Means* saja. Penggabungan Algoritma Genetik dan *Fuzzy C-Means* berhasil memperkecil total *cost* dan jumlah variabel. Dari 77 variabel yang dimiliki oleh dataset, hanya 61 variabel yang berpengaruh dan valid yang digunakan dalam proses clustering. Persentase rata-rata kesesuaian pada pengujian yang dilakukan dengan menggunakan data kuesioner terhadap mahasiswa program studi Sistem Informasi yang menempuh semester 8 (delapan) adalah sebesar 88.89%.



REFERENSI

- [1] Posuma, Christilia O. 2013. *Kompetensi, Kompensasi, Dan Kepemimpinan Pengaruhnya Terhadap Kinerja Karyawan Pada Rumah Sakit Ratumbuang Manado*. Jurnal EMBA Vol.1 No.4. 646-656.
- [2] Hsiang-Hsi Liu a, Chorng-Shyong Ong. 2008. *Variable selection in clustering for marketing segmentation using genetic algorithms*, Expert Systems with Applications 34. 502–510.
- [3] Kyoung-jae Kim, Hyunchul Ahn. 2008. *A recommender system using GA K-Means clustering in an online shopping market*. Expert Systems with Applications 34. 1200–1209.
- [4] Laetitia Jourdan, Clarisse Dhaenens, El-Ghazali Talbi. 2001. *A Genetic Algorithm for Feature Selection in Data-Mining for Genetics*. Metaheuristics International Conference. 4th. 29-33.
- [5] Bashar Al-Shboul, and Sung-Hyon Myaeng. 2009. *Initializing K-Means using Genetic Algorithms*, World Academy of Science, Engineering and Technology 54.114-118.
- [6] Prabha, K.Arun, R.Saranya. 2011. *Refinement of K-Means Clustering Using Genetic Algorithm*. Journal of Computer Applications (JCA) ISSN: 0974-1925, Volume IV, Issue 2.
- [7] Hartono, Erianto Ongko dan Dahlan Abdullah. 2015. *Determining a Cluster Centroid of K-Means Clustering Using Genetic Algorithm*. International Journal of Computer Science and Software Engineering (IJCSSE), Volume 4, Issue 6.