

Optimasi Fitur dalam Klasterisasi Mahasiswa Program Studi Sistem Informasi Dengan Algoritma Genetik

Feature Optimization in Clustering Student of Information System Using Genetic Algorithm

Rosalia Hadi

STMIK STIKOM Bali Jl. Raya Puputan No. 86 Renon Denpasar, telp. 0361 244445
Program Studi Sistem Informasi
E-mail: rosa@stikom-bali.ac.id

Abstrak

Klasterisasi merupakan proses mengelompokkan atau menggolongkan obyek dengan memaksimalkan kesamaan antar anggota dalam satu kelas dan meminimalkan kesamaan antar satu kelas dengan kelas lainnya. Implementasi klasterisasi dapat diterapkan pada berbagai bidang, salah satunya dalam hal penentuan kompetensi mahasiswa. 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. Teknik klasterisasi data (*clustering*) telah dipakai dengan sangat luas sekarang ini. Namun ternyata pada kenyataannya, masalah-masalah tentang fitur yang digunakan dalam proses klasterisasi data masih sangat diabaikan, sehingga kerap kali terjadi penggunaan fitur-fitur yang kurang relevan dalam proses klasterisasi data (*clustering*) tersebut. Hal ini dapat mengakibatkan hasil klasterisasi juga akan menjadi kurang optimal. Sehingga dibutuhkan proses optimasi fitur. Oleh karena itu dilakukan optimasi fitur dalam klasterisasi mahasiswa Program Studi Sistem Informasi menggunakan algoritma genetik untuk menentukan fitur yang valid dan mengabaikan fitur yang tidak valid. Sehingga proses klasterisasi nantinya akan dapat menghasilkan nilai *fitness* dan tingkat konvergensi yang lebih baik serta membentuk hasil *cluster* yang semakin akurat terkait dengan penentuan kompetensi mahasiswa.

Kata kunci— klasterisasi, optimasi fitur, mahasiswa

Abstract

Clustering is a process of grouping or classifying objects by maximizing the similarity between members of one class and minimize the similarity between one class with another class. Implementation of clustering can be applied to various fields, one of them in terms of determining the competence of students. Students will have the basis to work either as employees or as entrepreneurs based on its competence. The competencies will have positive influence on the performance of employees in a company. Clustering techniques have been used very widely today. But it turns out in fact, the problems of the features used in the clustering process data is still very neglected, so often happens the use of features that are less relevant in the clustering process. This can lead to clustering results will be less than optimal. So it takes a process of optimization features. So the optimization features in students clustering of Information Systems using genetic algorithms to determine what features a valid and ignore features that are not valid. So the clustering process will be able to generate the value of fitness and better convergence rate and form clusters results are more accurate to determine student competence.

Keywords— *clustering, feature selection, students*

1. PENDAHULUAN

Klasterisasi merupakan proses mengelompokkan atau menggolongkan obyek dengan memaksimalkan kesamaan antar anggota dalam satu kelas dan meminimalkan kesamaan antar satu kelas dengan kelas lainnya.. Implementasi klasterisasi dapat diterapkan pada berbagai bidang sebagai contoh dalam hal pemasaran, klasterisasi dapat digunakan sebagai metode untuk mengelompokkan pelanggan yang memiliki kesamaan dalam perilaku keseharian [1]. Penerapan klasterisasi dalam bidang ilmu biologi dapat dilakukan salah satunya dalam pengelompokan tumbuhan ataupun hewan. Dalam bidang informatika, teknik klasterisasi sangat banyak digunakan antara lain pada *data mining*, *information retrieval*, pengolahan citra, dan lain sebagainya. Contoh lainnya juga dalam hal penentuan kompetensi mahasiswa.

Kompetensi adalah kemampuan melakukan pekerjaan berdasarkan ilmu pengetahuan dan keahlian dari seseorang [2] . Penentuan kompetensi mahasiswa merupakan hal yang penting. 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.

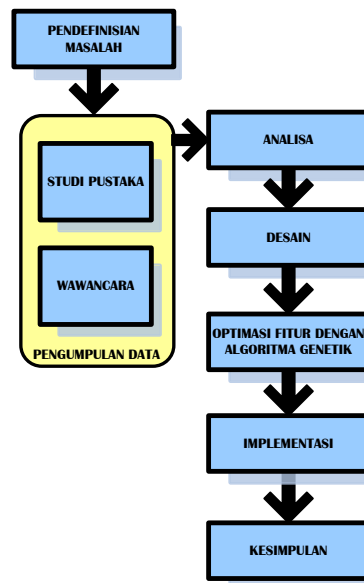
Teknik klasterisasi data (*clustering*) telah dipakai dengan sangat luas sekarang ini. Namun ternyata pada kenyataannya, masalah-masalah tentang fitur yang digunakan dalam proses klasterisasi data masih sangat diabaikan, sehingga kerap kali terjadi penggunaan fitur-fitur yang kurang relevan dalam proses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

Pemilihan fitur (*feature selection*) merupakan kegiatan yang termasuk ke dalam *preprocessing* untuk memilih fitur yang berpengaruh dalam suatu proses analisa data [3]. Terdapat banyak metode yang dapat digunakan dalam mencari fitur yang cocok, salah satunya algoritma genetik. Algoritma Genetik telah banyak digunakan dalam penelitian serupa. Jika Algoritma Genetik didesain secara baik, populasi akan mengalami konvergensi dan akan didapatkan sebuah solusi yang optimum [4]. Algoritma genetik berhasil dalam memilih dan menemukan fitur genetik dan faktor lingkungan yang terlibat dalam penyakit multifaktorial seperti obesitas dan diabetes [5]. Metode Algoritma Genetik juga digunakan dalam menyelesaikan masalah variabel yang tidak relevan dalam hal segmentasi pemasaran [1].

Berdasarkan permasalahan yang ada tersebut, maka pada penelitian ini digunakan Algoritma Genetik untuk optimasi fitur dalam klasterisasi mahasiswa Program Studi Sistem Informasi. Algoritma Genetik akan digunakan untuk menentukan fitur yang valid dan mengabaikan fitur yang tidak valid. Sehingga proses klasterisasi nantinya akan dapat menghasilkan nilai *fitness* dan tingkat konvergensi yang lebih baik serta membentuk hasil *cluster* yang semakin akurat terkait dengan penentuan kompetensi mahasiswa.

2. METODE PENELITIAN

Penelitian ini diawali dengan pendefinisian masalah, pengumpulan data dan dilanjutkan dengan analisa, desain dan implementasi sistem. Gambar .1 menunjukkan metode penelitian yang digunakan.



Gambar 1 Alur Analisis

Penelitian dimulai tahapan pengumpulan data sampai dengan melakukan implementasi sistem. Tahapan kegiatan secara rinci dari penelitian yang dilakukan adalah sebagai berikut:

- a. Pendefinisian permasalahan terkait dengan optimasi fitur dalam penentuan kompetensi mahasiswa.
- b. Studi Pustaka melalui pengumpulan data berupa literatur.
- c. Wawancara, dilakukan proses tanya jawab untuk pengumpulan data.
- d. Analisa, melakukan proses analisa kebutuhan
- e. Desain, dilakukan perancangan sistem berdasarkan hasil analisa yang telah dilakukan.
- f. Optimasi fitur dengan Algoritma Genetik, yaitu penerapan algoritma genetik untuk mengoptimalkan fitur dalam penentuan kompetensi mahasiswa.
- g. Implementasi, dilakukan pembuatan sistem berdasarkan hasil rancangan.
- h. Pengambilan kesimpulan, merupakan tahap penarikan kesimpulan dari hasil penelitian yang telah dilakukan.

3. HASIL DAN PEMBAHASAN

3.1 Variabel Dataset

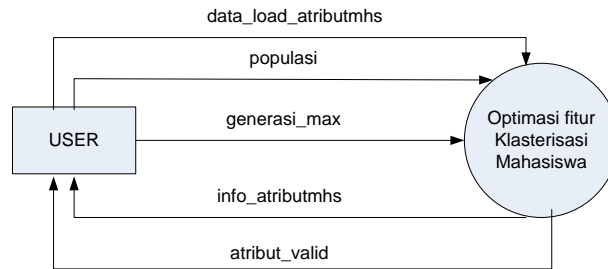
Studi kasus yang digunakan dalam penelitian ini adalah Program Studi Sistem Informasi STIKOM Bali. Data mahasiswa yang didapat memiliki variabel sebanyak 77 buah.

3.2 Perancangan Aplikasi

Berikut merupakan perancangan aplikasi Optimasi Fitur dalam Klasterisasi Mahasiswa Program Studi Sistem Informasi dengan Algoritma Genetik. Perancangan yang dibuat berupa diagram konteks dan DFD level 0.

3.2.1 Diagram Konteks

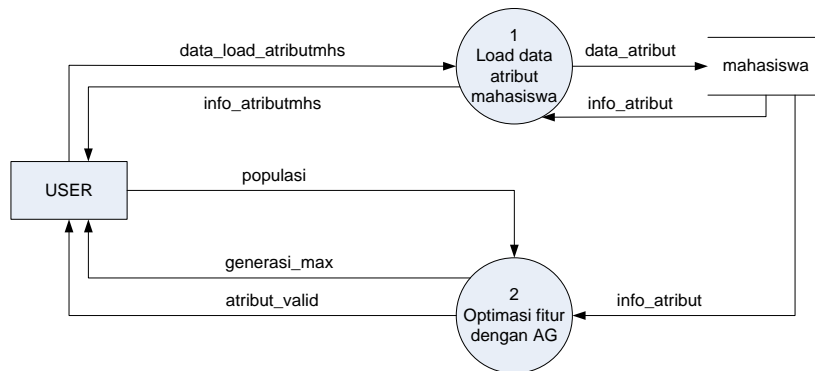
Berikut merupakan diagram konteks dari aplikasi Optimasi Fitur dalam Klasterisasi Mahasiswa Program Studi Sistem Informasi dengan Algoritma Genetik. Diagram konteks dapat dilihat pada Gambar 1



Gambar 2 Diagram Konteks

3.2.2 DFD Level 0

DFD Level 0 merupakan rincian dari diagram konteks. DFD Level 0 dari aplikasi Optimasi Fitur dalam Klasterisasi Mahasiswa Program Studi Sistem Informasi dapat dilihat pada Gambar 2.



Gambar 3 DFD Level 0

3.3 Implementasi Sistem

Implementasi sistem dilakukan setelah tahapan perancangan sistem. Berikut akan dijelaskan lebih rinci mengenai tahapan implementasi sistem.

3.3.1 Load Data Mahasiswa

Load data mahasiswa digunakan untuk memanggil data yang akan diolah, yang kemudian ditampilkan dalam bentuk *Grid* seperti yang terlihat pada Gambar 3. *Dataset* yang digunakan adalah mahasiswa Program Studi Sistem Informasi angkatan 2010 yang telah menempuh perkuliahan selama 8 (delapan) semester. *Dataset* yang ditampilkan berupa *grid* terdiri dari 300 *record* data dan 77 variabel.

	Jenis Kelamin	Kelas	Umur
1	1	0	24
2	1	0	24
3	1	0	22
4	1	0	25
5	0	0	26
6	1	0	23
7	1	0	22
8	1	0	22
9	1	0	21
10	1	0	23
11	0	0	23
12	1	0	22
13	1	0	23
14	1	0	23
15	1	0	24
16	1	0	22
17	1	0	22
18	1	0	22
19	1	1	22
20	1	0	22
21	1	0	22

Gambar 4 Dataset telah di-load

3.3.2 Parameter Algoritma Genetik

Selanjutnya dilakukan penginputan parameter Algoritma Genetik pada *textbox* yang disediakan. Total Populasi diisi dengan angka, sebagai contoh total populasi diisi dengan angka 30. 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. Semakin banyak total populasi yang digunakan maka proses akan membutuhkan waktu semakin lama. Begitupun sebaliknya, semakin sedikit total populasi yang digunakan maka semakin singkat pula waktu proses yang dibutuhkan.

Generasi Max dapat diisi dengan angka, misalnya 5. Dimana 1 generasi dapat mewakili jumlah populasi yang telah ditentukan pada tahap pengisian total populasi, sehingga menghasilkan kromosom yang lebih variatif dalam proses *fitness*, pada generasi selanjutnya akan menghasilkan populasi yang baru.

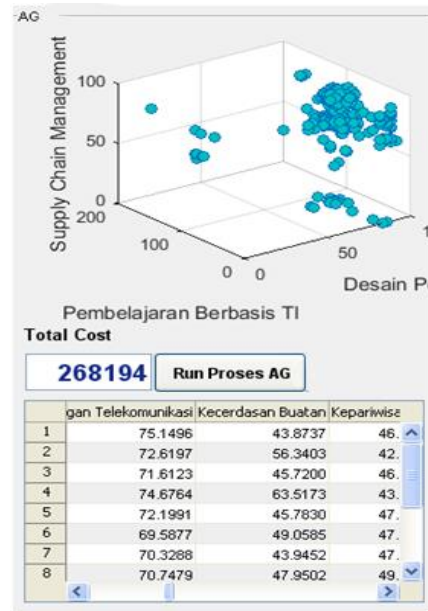
Gambar 4 menggambarkan penginputan dari parameter Algoritma Genetik. Total populasi sebesar 30 dan generasi max sebesar 5, maka jumlah komputasi pada proses pengolahan data adalah sebanyak 150 kali atau sama dengan total populasi dikalikan generasi max.

Gambar 5 Parameter Algoritma Genetik

3.3.3 Hasil Optimasi Fitur

Optimasi fitur bertujuan untuk memilih fitur yang berpengaruh, sehingga tidak terjadi kelimpahan atribut yang berlebihan namun tidak relevan. Tombol *Run* Proses Algoritma Genetik untuk dapat melihat hasil optimasi fitur oleh Algoritma Genetik. Gambar 5 menunjukkan hasil optimasi fitur dengan Algoritma Genetik. Jumlah total atribut dari hasil optimasi fitur adalah sebanyak 61 atribut dari atribut awal yang

berjumlah 77. Sedangkan hasil total cost adalah sebesar 268194. Total cost merupakan nilai *fitness* terbaik dari iterasi. Dengan posisi posisi jarak masing-masing atribut terhadap pusat cluster dapat dilihat pada *Grid*, dan digambarkan pula pada grafik 3 dimensi.



Gambar 6 Hasil Optimasi Fitur dengan Algoritma Genetik

4. KESIMPULAN

Keimpulan yang dapat ditarik adalah sebagai berikut :

1. Dilakukan Optimasi Fitur dalam Klasterisasi Mahasiswa Program Studi Sistem Informasi STIKOM Bali dengan Algoritma Genetik.
2. Jumlah total atribut dari hasil optimasi fitur adalah sebanyak 61 atribut dari atribut awal yang berjumlah 77.

5. SARAN

Saran untuk ke depannya dapat dilakukan proses clustering untuk penentuan kompetensi mahasiswa Program Studi Sistem Informasi menggunakan fitur-fitur hasil optimasi.

UCAPAN TERIMA KASIH

Ucapan terima kasih peneliti haturkan kepada tim reviewer yang telah bersedia meluangkan waktu untuk me-review artikel ini.

DAFTAR PUSTAKA

- [1] Hsiang-Hsi Liu a, Chornng-Shyong Ong. (2008). *Variable selection in clustering for marketing segmentation using genetic algorithms*, Expert Systems with Applications 34. 502–510.

- [2] Posuma, Christilia O. (2013). *Kompetensi, Kompensasi, Dan Kepemimpinan Pengaruhnya Terhadap Kinerja Karyawan Pada Rumah Sakit Ratumbusang Manado*. Jurnal EMBA Vol.1 No.4. 646-656.
- [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] Basuki, Achmad. (2003). *Algoritma Genetika, Suatu Alternatif Penyelesaian Permasalahan Searching, Optimasi dan Machine Learning*, Politeknik Elektronika Negeri Surabaya PENS-ITS Surabaya.
- [5] 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.