

3

Penerapan Fuzzy C-Means Clustering Untuk Mengoptimalkan Penentuan Media Promosi

Oleh : R. Joko Sarjanako

Abstrak

Media promosi merupakan salah satu faktor penting dalam pemasaran suatu perguruan tinggi untuk menarik minat calon mahasiswa agar melanjutkan pendidikannya ke tingkat yang lebih tinggi. Pemerataan pendistribusian media promosi pada STIKOM Binaniaga yang sama ke semua sekolah menengah atas dan pada semua program studi membuat kurang efektifnya promosi. Pengelompokan data mahasiswa berdasarkan kriteria asal sekolah, jurusan dan daerah asal mahasiswa dapat membantu dalam mengoptimalkan penentuan media promosi. Metode yang bisa digunakan dalam mengelompokkan mahasiswa ini adalah Fuzzy C-Means Clustering. Metode Fuzzy C-Means Clustering berusaha mengelompokkan data yang ada kedalam beberapa kelompok, dimana dalam penelitian ini data mahasiswa dikelompokkan menjadi 3 cluster. Cluster yang pertama menghasilkan sebanyak 42,006 % dari seluruh total mahasiswa sedangkan cluster kedua menghasilkan sebanyak 33,229% dari seluruh total mahasiswa dan cluster ketiga menghasilkan 24,764 % dari seluruh total mahasiswa. Dari ketiga cluster tersebut kemudian dihitung nilai validitas internalnya menggunakan Dunn Index, dunn index menghitung validitas cluster menggunakan diameter cluster (kohesi) dan jarak antara dua cluster (separasi) dimana pada penelitian kali ini nilai DI sebesar 8.6674. Nilai DI yang semakin besar menandakan hasil clustering yang semakin baik.

Kata Kunci : FCM, Fuzzy Clustering, Matlab, Dunn Index

Pendahuluan

Latar Belakang

Perguruan tinggi merupakan elemen utama yang dipertimbangkan oleh pelajar sekolah menengah atas. Pelajar yang akan melanjutkan studi tentunya akan mempertimbangkan beberapa hal sebagai acuan untuk memilih perguruan tinggi mana yang akan dijadikan sebagai sarana menuntut ilmu ke tingkat yang lebih tinggi. Lembaga pendidikan sudah menjadi komoditas industri yang diperlakukan seperti suatu perusahaan. Pertumbuhan perguruan tinggi swasta membuat persaingan memperebutkan calon siswa berlangsung ketat. Persaingan dimulai bahkan ketika calon mahasiswa masih menempuh pendidikan SMA-nya. Hal serupa dialami oleh STIKOM Binaniaga Bogor. Penelitian ini bertujuan untuk memberikan usulan kajian strategi promosi STIKOM Binaniaga perekrutan mahasiswa baru mendatang.

Adapun yang akan dijadikan sebagai objek dalam penelitian ini adalah sebuah perguruan tinggi swasta yang berlokasi di daerah Bogor dengan nama STIKOM Binaniaga Bogor. Dalam prosesnya STIKOM Binaniaga melakukan promosi dengan beberapa media publikasi seperti brosur, spanduk, surat kabar, website dan melakukan pameran di sekolah-sekolah. Pemilihan media promosi yang sesuai dengan keadaan pendidikan tinggi sangat diperlukan. Hal ini disebabkan karena media promosi banyak digunakan untuk memperkenalkan pendidikan tinggi kepada masyarakat luas terutama dalam membangun dan mengembangkan opini dan persepsi masyarakat tentang eksistensi pendidikan tinggi tersebut. Promosi merupakan suatu kegiatan penyampaian informasi dengan tujuan mendapatkan peminat calon mahasiswa lebih banyak.

Teknologi data mining merupakan salah satu alat bantu untuk penggalian data pada basis data berukuran besar dengan spesifikasi kerumitan tinggi dan telah banyak digunakan pada

lingkungan aplikasi bisnis seperti perbankan, provider telekomunikasi, perusahaan pertambangan, perminyakan, dan lain-lain. Clustering merupakan salah satu teknik data mining yang berfungsi melakukan pengelompokan sejumlah data atau objek ke dalam cluster (group) sehingga setiap cluster akan berisi data yang semirip mungkin dan berbeda dengan objek dalam cluster yang lainnya.

Analisis cluster merupakan salah satu teknik data mining yang bertujuan untuk mengidentifikasi sekelompok objek yang mempunyai kemiripan karakteristik tertentu yang dapat dipisahkan dengan kelompok objek lainnya, sehingga objek yang berada dalam kelompok yang sama relatif lebih homogen daripada objek yang berada pada kelompok yang berbeda. Jumlah kelompok yang dapat diidentifikasi tergantung pada banyak dan variasi data objek. Tujuan dari pengelompokan sekumpulan data objek ke dalam beberapa kelompok yang mempunyai karakteristik tertentu dan dapat dibedakan satu sama lainnya adalah untuk analisis dan interpretasi lebih lanjut sesuai dengan tujuan penelitian yang dilakukan.

Algoritma Fuzzy C-Means adalah suatu teknik *clustering* yang banyak digunakan dalam aplikasi *clustering*. Fuzzy C-Means menerapkan pengelompokan fuzzy, dimana setiap data dapat menjadi anggota dari beberapa *cluster* dengan derajat keanggotaan yang berbeda-beda pada setiap cluster. Fuzzy C-Means merupakan algoritma iteratif, yang menerapkan iterasi pada proses *clustering* data. Tujuan dari Fuzzy C-Means adalah untuk mendapatkan pusat cluster yang nantinya akan digunakan untuk mengetahui data yang masuk ke dalam sebuah cluster (Kusumadewi & Purnomo, 2010).

Berdasarkan penelitian diatas, Fuzzy C-Means pada penelitian ini akan diterapkan pada data mahasiswa di STIKOM Binaniaga berdasarkan asal sekolah sehingga bisa didapatkan kelompok-kelompok sekolah yang siswanya banyak terdaftar sebagai mahasiswa di STIKOM Binaniaga atau berdasarkan kriteria –kriteria yang ditentukan. Dengan metode tersebut disertai dengan penerapannya menggunakan tools Matlab diharapkan mampu memberikan kemudahan dalam mengidentifikasi sekolah yang paling potensial untuk perekrutan mahasiswa baru mendatang dengan rekomendasi penerapan media promosi yang optimal.

Permasalahan Penelitian

1. Identifikasi Masalah :
 - a. Penentuan media promosi yang belum optimal.
 - b. Belum diketahui sekolah yang paling potensial untuk perekrutan mahasiswa baru.
2. Rumusan Masalah :
 - a. Problem Statement
Belum optimalnya penentuan media promosi yang dilakukan oleh STIKOM Binaniaga
 - b. Research Question
Bagaimana penerapan *Fuzzy C-Means clustering* untuk mengoptimalkan penentuan media promosi yang dilakukan oleh STIKOM Binaniaga.

Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah untuk :

1. Mengoptimalkan penentuan media promosi di STIKOM Binaniaga.
2. Mengetahui sekolah yang paling potensial untuk perekrutan mahasiswa baru.

Landasan Teori

Fuzzy Clustering

Pada logika fuzzy, ada beberapa metode yang dapat digunakan untuk melakukan pengelompokan sejumlah data yang sering kita kenal dengan *fuzzy clustering*. Pada kebanyakan situasi, *fuzzy clustering* lebih alami jika dibandingkan dengan pengclusteran secara klasik. Suatu algoritma clustering dikatakan sebagai algoritma *fuzzy clustering* jika dan hanya jika algoritma tersebut menggunakan parameter strategi adaptasi secara *soft competitive* (non-crisp) (Baraldi, 1998). Sebagian besar algoritma *fuzzy clustering* didasarkan atas optimasi fungsi objektif atau modifikasi dari fungsi obyektif tersebut.

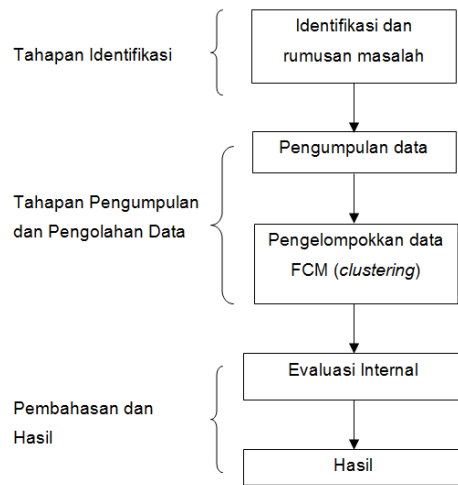
Pemilihan algoritma clustering yang tepat sangatlah penting demi suksesnya proses clustering. Secara umum, algoritma pengclusteran dicirikan berdasarkan ukuran kedekatan dan kriteria pengclusteran (Vazirgiannis, 2003). Ukuran kedekatan menunjukkan seberapa dekat kedekatan fitur antara 2 data; sedangkan kriteria pengclusteran bisaanya diekspresikan dengan menggunakan fungsi biaya atau tipe aturan yang lainnya. (Sri Kusumadewi, 2006, p. 294)

Fuzzy C-Means

Clustering dengan metode *Fuzzy C-Means* (FCM) didasarkan pada teori logika fuzzy. Teori ini pertama kali diperkenalkan oleh Lotfi Zadeh (1965) dengan nama himpunan *fuzzy* (*fuzzy set*). Dalam teori *fuzzy*, keanggotaan sebuah data tidak diberikan nilai secara tegas dengan 1 (menjadi anggota) dan 0 (tidak menjadi anggota), melainkan dengan suatu nilai derajat keanggotaan yang jangkauan nilainya 0 sampai 1. Nilai keanggotaan suatu data dalam sebuah himpunan menjadi 0 ketika sama sekali tidak menjadi anggota, dan menjadi 1 ketika menjadi anggota secara penuh dalam suatu himpunan. Umumnya nilai keanggotaan antara 0 dan 1. Semakin tinggi nilai keanggotaannya maka semakin tinggi derajat keanggotaannya, dan semakin kecil maka semakin rendah derajat keanggotaannya. Kaitannya dengan K-Means, sebenarnya FCM merupakan versi *fuzzy* dari K-Means dengan beberapa modifikasi yang membedakannya dengan K-Means. (Prasetyo, 2014, p. 218)

Konsep dasar FCM, pertama kali adalah menentukan pusat cluster yang akan menandai lokasi rata-rata untuk tiap-tiap *cluster*. Pada kondisi awal, pusat cluster ini masih belum akurat. Tiap-tiap data memiliki derajat keanggotaan untuk tiap-tiap cluster. Dengan cara memperbaiki pusat *cluster* dan nilai keanggotaan tiap-tiap data secara berulang, maka dapat dilihat bahwa pusat cluster akan menuju lokasi yang tepat. Perulangan ini didasarkan pada minimasi fungsi obyektif (Gelley, 2000).

Desain Penelitian



Gambar 1. Desain Penelitian

Sumber Data

Data merupakan bagian terpenting dalam penelitian. Sumber data dari penelitian ini adalah data pendaftaran mahasiswa berbentuk *spreadsheet* yang diperoleh dari Bagian Administrasi Akademik (BAAK)

Instrument Penelitian

Instrumen yang digunakan dalam penelitian ini berupa data mahasiswa tahun 2006 sampai dengan tahun 2015 dengan menggunakan tool berupa MATLAB untuk proses pengelompokan mahasiswa (*clustering*).

Analisa Data

Setelah melalui proses perhitungan data melalui clustering dan hasil dari prediksi data. Proses akhir yaitu melakukan perhitungan evaluasi internal dengan menggunakan evaluasi Dunn Index (DI). Dunn Index (DI) menghitung validitas *cluster* menggunakan diameter *cluster* (kohesi) dan jarak antara dua *cluster* (separasi). Untuk mendapatkan diameter sebuah cluster ke-*i* dilakukan dengan menghitung jarak pasangan dua data dalam sebuah cluster, kemudian diambil yang terbesar, seperti dinyatakan oleh persamaan Dunn Index (DI) berikut :

$$\Delta_i = \max_{x,y \in C_i} d(x,y)$$

Δ_i Dunn Index

X dan y adalah data yang berada dalam cluster C_i .

Konsep formula persamaan diatas juga serupa dengan konsep pengelompokan berbasis hierarki sehingga untuk mendapatkan diameter dalam DI, selain bisa menggunakan max, boleh juga menggunakan min, average, bahkan jarak dua centroid (Bezdek, 1998). Dunn Index didapatkan dari persamaan berikut :

$$\text{Dunn Index (C)} = \min_{1 \leq i \leq N} \left\{ \min_{1+1 \leq j \leq N} \left(\frac{\text{dist}(c_i, c_j)}{\max_{1 \leq i \leq N} \text{diam}(c_i)} \right) \right\}$$

N= jumlah kluster

Nilai DI yang semakin besar menandakan hasil *clustering* yang semakin baik. (Prasetyo, 2014, p. 290)

Hasil Penelitian

1. Pengelempokan Menggunakan Fuzzy C-Means

Proses pengelompokan yang dilakukan pada penelitian ini dilakukan dengan mengelompokkan calon mahasiswa berdasarkan asal sekolah, jurusan, daerah asal tempat tinggal mahasiswa dan informasi media promosi. Proses *clustering* dilakukan dengan menggunakan software Matlab R2012b (8.0.0.783). Sebelum melakukan clustering dengan FCM, ditentukan terlebih dahulu parameter-parameter FCM yang dibutuhkan untuk *clustering* data.

Untuk data yang akan diproses pada penelitian ini, terlebih dahulu data di convert ke dalam file CSV (Comma Separated Value), kemudian data dibuka menggunakan notepad dan disimpan kembali ke dalam bentuk format dengan ekstension (.dat) agar bisa dibaca oleh software tersebut untuk proses *clustering* data.

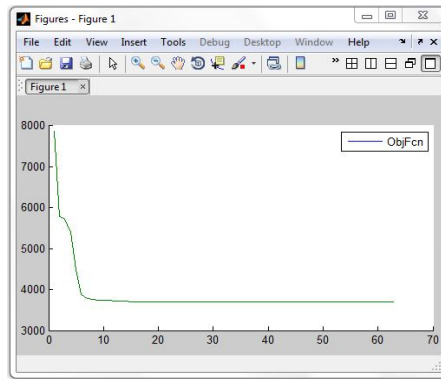
Jumlah cluster pengguna parameter fuzzy seperti pembobot (w), iterasi maksimum, nilai kriteria penghentian yang digunakan dalam FCM adalah nilai default yang bisa digunakan dalam penelitian atau yang ada di dalam matlab.

2. Hasil Clustering

Hasil dari iterasi pertama dengan pengujian *clustering* berjumlah 3 *cluster* mendapatkan total 63 iterasi yang artinya pada iterasi ke 63 sudah optimal.

Proses iterasi berhenti di pengulangan ke-63 dari maksimal iterasi yang ditentukan yaitu iterasi ke-100, proses iterasi ini berhenti karena sudah mendapatkan hasil iterasi sebelum mencapai batas maksimum iterasi.

Didapatnya fungsi objektif objfcn yang dihitung dari iterasi pertama hingga iterasi selesai dihentikan yaitu pada iterasi ke 63 dari percobaan pada perintah 100 iterasi, proses berhentinya iterasi pada matlab dapat dilihat pada gambar 2 dibawah ini.



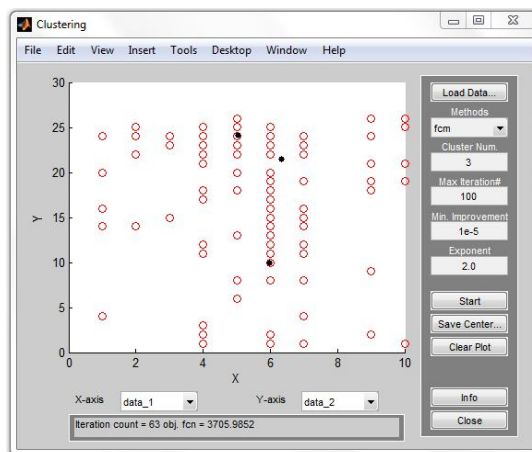
Gambar 2. Grafik Fungsi Objektif

Grafik dari fungsi objektif yang dihasilkan dari *clustering* berdasarkan data mahasiswa dapat dilihat pada gambar 2. Pada grafik tersebut terlihat nilai fungsi objektif sudah mulai stabil pada iterasi ke-63, artinya *cluster* yang terbentuk pada iterasi tersebut sudah mulai optimal.

Dari hasil V sebagai titik pusat *centroid* maka didapat urutan hasil yang bisa diperoleh adalah :

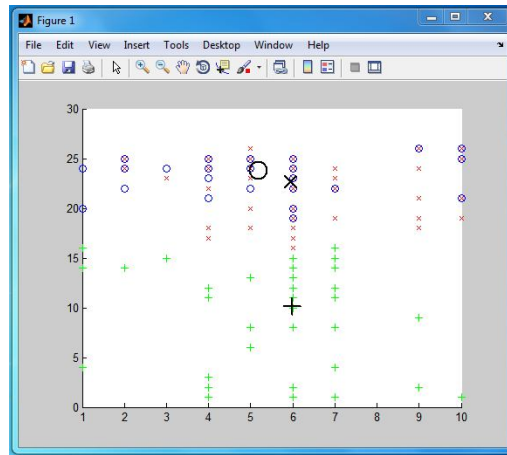
- a. Sekolah yang paling banyak menarik mahasiswa yaitu SMA Negeri
- b. Selain dari SMA negeri, yang berpotensi menarik banyak peminat yaitu sekolah menengah kejuruan yang memiliki program jurusan Teknik Komputer Jaringan, Multimedia dan Penjualan
- c. Daerah yang potensial adalah daerah tanah sereal, bogor utara dan bogor selatan
- d. Media promosi yang paling baik digunakan yaitu menggunakan spanduk

Dari hasil *clustering* dapat juga divisualisasi hasil tersebut kedalam plot 2 dimensi seperti yang digambarkan dalam gambar 3 berikut ini.



Gambar 3. Kumpulan Anggota Cluster dan Pusat Cluster

Pada gambar 3 dapat dilihat titik centroid yang terbagi menjadi 3 cluster dapat dilihat pada gambar dengan gambar titik hitam besar, dan titik-titik yang bersarna merah merupakan anggota dari data. Penjelasan gambar yang lebih jelas ditampilkan dalam gambar 4 berikut ini :



Gambar 4 Plot Clustering

Seperti yang dapat dilihat pada gambar 6. Cluster pertama digambarkan tanda O berwarna biru dan titik pusat cluster bertanda O lebih besar berwarna hitam. Cluster kedua digambarkan tanda + berwarna hijau dan titik pusat cluster bertanda + lebih besar berwarna hitam. Cluster ketiga digambarkan tanda X berwarna merah dan titik pusat cluster bertanda X lebih besar berwarna hitam.

Pembahasan

Dari hasil perhitungan *clustering* diatas, terdapat 134 mahasiswa yang terbentuk kedalam cluster pertama atau 42,006 %, cluster kedua dengan jumlah mahasiswa 106 atau 33,229% dan cluster terakhir dengan jumlah mahasiswa sebesar 79 atau 24,764 % .

Dari hasil clustering akan dilakukan validitas dengan dunn index. Parameter jarak yang digunakan untuk mengukur jarak antar-data adalah 'Euclidean', sedangkan tipe jarak antar cluster yang digunakan adalah single linkage.

Dunn Index (DI) menghitung validitas cluster menggunakan diameter cluster (kohesi) dan jarak antara dua cluster (separasi). Untuk mendapatkan diameter sebuah cluster ke-i dilakukan dengan menghitung jarak pasangan dua data dalam sebuah cluster, kemudian diambil yang terbesar, seperti dinyatakan oleh persamaan berikut :

$$\Delta_i = \max_{x,y \in C_i} d(x,y)$$

x,y adalah data yang berada dalam cluster C_i

Jarak setiap data ke data yang lain (separasi) dalam cluster 1, 2, dan 3 masing-masing diambil nilai terbesarnya . Untuk nilai Δ_i pada hasil cluster didapatkan dari matriks partisi dari

nilai U yang kemudian didapatkan setiap cluster indeksnya sebagai berikut :

$$\Delta_1 = 0.9803$$

$$\Delta_2 = 0.8913$$

$$\Delta_3 = 0.9880$$

Sehingga :

$$\Delta_i = \max (0.9803, 0.8913, 0.9880) = 0.9880$$

Hasil dunn index atau metode pengelompokan yang dilakukan berbasis partisi yang didasarkan pada gabungan nilai kohesi (jumlah kedekatan data terhadap centroid dan cluster yang diikutinya Δ_i) dan separasi (kedekatan antara dua cluster $\delta_{i,j}$) pada penelitian diatas didapatkan hasil sebesar **8.6674** yang artinya nilai DI yang semakin besar menandakan hasil *clustering* yang semakin baik. (Prasetyo, 2014, p. 290)

Kesimpulan

Berdasarkan hasil pengujian metode Fuzzy C-Means dalam pengelompokan mahasiswa dengan menggunakan software Matlab R2012b (8.0.0.783), maka dapat disimpulkan bahwa :

1. Proses pengelompokan berdasarkan sekolah, jurusan , daerah asal mahasiswa serta informasi media promosinya dapat dilakukan dengan menggunakan algoritma Fuzzy C-Means.
2. Dari nilai center diperoleh hasil 3 cluster, cluster pertama dengan titik center (5.1695, 23.7855, 6.8057, 2.3064) , cluster kedua dengan titik center (5.9552, 22.6939, 4.1826, 2.5299) dan cluster ketiga dengan titik center (5.9881, 10.1303, 5.9437, 2.0341).
3. Berdasarkan hasil tersebut dapat diketahui terdapat 134 mahasiswa yang terbentuk kedalam cluster pertama atau 42,006 % , cluster kedua dengan jumlah mahasiswa 106 atau 33,229% dan cluster terakhir dengan jumlah mahasiswa sebesar 79 atau 24,764 %
4. Dari nilai center juga diperoleh hasil :
 - a. Sekolah yang paling banyak menarik mahasiswa yaitu SMA Negeri.
 - b. Selain dari SMA negeri, yang berpotensi menarik banyak peminat yaitu Sekolah kejuruan yang memiliki program jurusan Teknik Komputer Jaringan, Multimedia dan Penjualan.
 - c. Daerah yang potensial adalah daerah tanah sereal, bogor utara dan bogor selatan.
 - d. Media promosi yang paling baik digunakan yaitu menggunakan spanduk.

Dari ketiga cluster tersebut kemudian dihitung nilai validitas internal menggunakan Dunn Index, dunn index menghitung validitas *cluster* menggunakan gabungan dari diameter *cluster* (kohesi) dan jarak antara dua *cluster* (separasi) dimana pada penelitian kali ini nilai DI sebesar **8.6674**. Nilai DI yang semakin besar menandakan hasil *clustering* yang semakin baik.

Saran

Adapun saran yang dapat disampaikan untuk pengembangan lebih lanjut dari penelitian ini adalah:

1. Untuk pengembangan selanjutnya dapat melakukan clustering dengan menambah variabel lain dari data mahasiswa yang tidak digunakan dalam penelitian ini misalnya dengan menambahkan indeks prestasi kumulatif.
2. Data yang digunakan dalam penelitian selanjutnya menggunakan data terbaru

Daftar Pustaka

- Danang Sunyoto, S. S. (2014). *Dasar-Dasar Manajemen Pemasaran (Konsep, Strategis Dan Kasus)*. Jakarta, Jawa Barat: Center Of Academic Publishing Service.
- Elvira Asril, F. W. (2015). *Analisis Data Lulusan Dengan Data Mining Untuk Mendukung Strategi Promosi Universitas Lancang Kuning*. 6(2).
- Lindawati. (2008, Mei 24). *Data Mining Dengan Teknik Clustering Dalam Pengklasifikasian Data Mahasiswa Studi Kasus Prediksi Lama Studi Mahasiswa Universitas Bina Nusantara*.
- Mudini, D. (2015). *Penerapan Metode Fuzzy Mamdani Untuk Optimalisasi Jumlah Pemasangan Media Promosi Spanduk (Studi Kasus Stikom Binaniaga Bogor)*.
- Ningrum, L. T. (2012). *Penerapan Fuzzy-Ahp Untuk Menentukan Media Promosi Pada Penerimaan Mahasiswa Baru*.
- Ong, J. O. (2013, Juni). *Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University*. 12(1).
- Prasetyo, E. (2014). *Data Mining - Mengolah Data Menjadi Informasi Menggunakan Matlab*. (A. Sahala, Ed.) Yogyakarta: Andi Offset.
- Ramadhani, R. D. (N.D.). *Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro*.
- Ri Handayani, E. B. (2011, November 12). *Implementasi Algoritma Clustering Ismc Dan Fcm (Studi Kasus: Jalur Pmb Di It Telkom Bandung)*.
- Sri Kusumadewi, S. H. (2006). *Fuzzy Multi-Attribute Decission Making (Fuzzy Madm)* (Vol. 1). Yogyakarta: Graha Ilmu.
- Suprawoto, T. (2016, 2). *Klasifikasi Data Mahasiswa Menggunakan Metode K-Means Untuk Menunjang Pemilihan Strategi Pemasaran*, 1(1).
- Www.Stikombinaniaga.Ac.Id. (N.D.). *About Us*. Retrieved 03 15, 2016, From Www.Stikombinaniaga.Ac.Id