

ALGORITMA FUZZY C-MEANS PADA APLIKASI MATLAB DALAM MENENTUKAN DOSEN PEMBIMBING TUGAS AKHIR

Yunita Sari Siregar¹, Putri Harliana²

^{1,2}Universitas Harapan Medan

yunie.boreg@yahoo.com, cimoputri@gmail.com

ABSTRAK

Algoritma *fuzzy c-means* merupakan suatu teknik pengelompokan data dimana keberadaan tiap-tiap titik data dalam suatu cluster ditentukan oleh derajat keanggotaan. Matlab merupakan sebuah bahasa pemrograman untuk komputasi masalah teknik dengan kinerja yang tinggi. Dalam penelitian ini, model *fuzzy c-means* akan digunakan untuk menentukan dosen pembimbing tugas akhir sesuai dengan bidang keahliannya. Penelitian ini pengujiannya dilakukan menggunakan aplikasi matlab. Hasil pengujian yang dilakukan dapat diketahui beberapa syarat dalam menentukan dosen pembimbing.

Kata Kunci : *Fuzzy C-Means*, Dosen Pembimbing, *Data Mining*, *Matlab*

1. PENDAHULUAN

Algoritma *fuzzy c-means* merupakan suatu teknik pengelompokan data dimana keberadaan tiap-tiap titik data dalam suatu cluster ditentukan oleh derajat keanggotaan. *fuzzy c-means* adalah algoritma pengelompokan yang terawasi, karena pada algoritma *fuzzy c-means* jumlah cluster yang akan dibentuk perlu diketahui terlebih dahulu. Konsep dasar algoritma *fuzzy c-means* adalah menentukan pusat kelompok yang akan menandai lokasi rata-rata untuk tiap-tiap *cluster*. Pada kondisi awal, pusat *cluster* ini masih belum akurat. Tiap-tiap titik data memiliki derajat keanggotaan pada tiap-tiap *cluster*. Dengan cara memperbaiki pusat *cluster* dan derajat keanggotaan tiap-tiap titik data secara berulang-ulang, maka didapat lokasi pusat *cluster* optimal. Perulangan ini didasarkan pada minimasi fungsi obyektif yang menggambarkan jarak dari titik data yang diberikan ke pusat *cluster* yang terbobot oleh derajat keanggotaan titik data tersebut. [6]

Penelitian berjudul Clustering Dengan Algoritma Fuzzy C Means Untuk Rekomendasi Pemilihan Bidang Keahlian Pada Program Studi Teknik Informatika yang dilakukan oleh Muhammad Faisal Mirza dari jurusan Teknik Informatika Universitas Dian Nuswantoro Semarang pada tahun 2009. Penelitian ini membahas tentang penggunaan algoritma FCM untuk mengelompokkan mahasiswa berdasarkan transkrip nilai mata kuliah prasyarat untuk rekomendasi penjurusan. [7]

Berdasarkan penelitian yang sudah pernah dilakukan oleh peneliti lain dan dari uraian di atas penulis tertarik untuk melakukan penelitian perancangan analisis algoritma *fuzzy c-means* dalam penentuan dosen pembimbing. Untuk studi kasus dilaksanakan di Universitas Harapan Medan.

Dari latar belakang yang telah dijelaskan sebelumnya maka masalah dari penelitian ini adalah seberapa banyak syarat yang diperlukan untuk menentukan dosen pembimbing menggunakan algoritma *fuzzy c-means*. Dan tujuan yang ingin dicapai pada penelitian ini adalah untuk menentukan dosen pembimbing sesuai dengan minat dan bidang keahlian dosen dengan metode clustering algoritma *fuzzy c-means*. Manfaat yang diharapkan dari penelitian ini adalah memberikan motivasi kepada penulis lain untuk melanjutkan penelitian ini dengan metode yang sama namun dengan objek yang berbeda atau juga dengan metode yang berbeda pada objek yang sama.

a. Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang besar [4]. *Data mining* didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semiotomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Data yang dibutuhkan dalam jumlah besar [6].

Karakteristik data *mining* sebagai berikut :

- Data *mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- Data *mining* biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
- Data *mining* berguna untuk membuat keputusan yang kritis, terutama dalam strategi.

Berdasarkan beberapa pengertian yang telah disebutkan diatas, maka dapat ditarik kesimpulan bahwa data *mining* adalah suatu teknik untuk menggali informasi yang tersembunyi pada gunung data. Kata *mining* sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar [6]

Data *mining* juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Data *mining* berkaitan dengan bidang ilmu-ilmu lain, seperti *database system*, data *warehousing*, statistik, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi. Selain itu, data *mining* didukung oleh ilmu lain seperti *neural network*, pengenalan pola, *spatial data analysis*, *image database*, *signal processing*. Istilah data *mining* kadang disebut juga *knowledge discovery* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar [5].

b. Fuzzy C-Means

Clustering merupakan proses pengelompokan data dalam kelas-kelas atau *cluster-cluster* sehingga data dalam suatu cluster memiliki tingkat kesamaan yang tinggi antara data satu dengan yang lainnya tetapi sangat berbeda dengan data pada *cluster* lain. *Clustering* juga dapat dianggap sebagai bentuk kompresi data, di mana sejumlah besar sampel diubah menjadi sejumlah kecil perwakilan prototipe atau *cluster*. Tergantung pada data dan aplikasi, berbagai jenis ukuran kesamaan dapat digunakan untuk mengidentifikasi kelas, di mana ukuran kesamaan mengontrol bagaimana *cluster* terbentuk. Beberapa contoh nilai-nilai yang dapat digunakan sebagai parameter kesamaan termasuk jarak, konektivitas dan intensitas (Mantho, 2015).

Clustering dapat diterapkan ke dalam data yang kuantitatif (numerik), kualitatif (kategorikal), atau kombinasi dari keduanya. Data dapat merupakan hasil pengamatan dari suatu proses. Setiap pengamatan dapat memiliki n variabel pengukuran dan dikelompokkan dalam n dimensi vektor $z_k = [z_{1k}, \dots, z_{nk}]^T$, $z_k \in R^n$. Sebuah himpunan dari N pengamatan dinotasikan $Z = \{ z_k \mid k = 1, 2, \dots, N \}$ dan direpresentasikan sebagai matriks $n \times N$. *Cluster* secara umum merupakan wujud himpunan bagian dari suatu himpunan data dan metode *clustering* dapat diklasifikasikan berdasarkan himpunan bagian yang dihasilkan: apakah *fuzzy* atau *crisp* (*hard*) [4].

Fuzzy C-Means adalah suatu teknik *clustering* (pengelompokan) data di mana keberadaan titik-titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan. Penentuan titik *cluster*

dilakukan secara berulang-ulang hingga diperoleh data yang akurat berdasarkan derajat keanggotaannya. Perulangan ini didasarkan pada minimalisasi fungsi obyektif yang menggambarkan jarak dari titik data ke pusat *cluster* yang terbobot oleh derajat keanggotaan. Akibat adanya derajat keanggotaan tersebut, maka suatu titik data bisa dimiliki lebih dari satu kelompok. Metode ini merupakan minimasi dari fungsi obyektif [5].

c. Matlab

Matlab merupakan sebuah bahasa pemrograman untuk komputasi masalah teknik dengan kinerja yang tinggi. Matlab dikembangkan oleh *the mathworks*, memungkinkan matlab melakukan manipulasi matriks, pemlotan fungsi dan data, implementasi algoritma, pembuatan antar muka pengguna dalam bahasa lainnya. Meskipun hanya bernuansa numerik, matlab memungkinkan akses terhadap kemampuan aljabar komputer. [8]

Penggunaan matlab meliputi matematika dan komputasi, pembentukan algoritma, akuisisi data, pemodelan, simulasi, dan pembuatan prototype, analisa data, eksplorasi dan visualisasi, dan grafik keilmuan dan bidang rekayasa. Fitur-fitur matlab sudah banyak dikembangkan, dan lebih kita kenal dengan nama *toolbox*. Seorang pengguna matlab harus tahu *toolbox* mana yang mendukung untuk *learn* dan *apply technology*. *Toolbox* ini merupakan kumpulan dari fungsi-fungsi matlab (*M-files*) yang telah dikembangkan ke suatu lingkungan kerja matlab untuk memecahkan masalah dalam kelas particular. Area-area yang sudah bisa dipecahkan dengan *toolbox* saat ini meliputi pengolahan sinyal, system kontrol, *neural networks*, *fuzzy logic*, *wavelets*, dan lain-lain. [8]

Sebagai sebuah system, matlab tersusun dari 5 bagian utama:

- **Development Environment**, merupakan sekumpulan perangkat dan fasilitas yang membantu kita untuk menggunakan fungsi-fungsi dan file-file matlab. Beberapa perangkat ini merupakan sebuah Graphical User Interfaces (GUI). Termasuk didalamnya adalah *matlab desktop* dan *Command Window*, *Command History*, sebuah editor dan *debugger*, dan *browsers* untuk melihat *help*, *workspace*, *files*, dan *search path*.
- **Matlab Mathematical Function Library**, merupakan sekumpulan algoritma komputasi mulai dari fungsi-fungsi dasar seperti: *sum*, *sin*, *cos*, dan *complex arithmetic*, sampai dengan fungsi-fungsi yang lebih kompek seperti *matrix inverse*, *matrix eigenvalues*, *Bessel functions*, dan *fast Fourier transforms*.

- Matlab Language, merupakan suatu *high-level matrix/array language* dengan *control flow statements, functions, data structures, input/output, dan fitur-fitur object-oriented programming*. Ini memungkinkan bagi kita untuk melakukan kedua hal baik “pemrograman dalam lingkup sederhana” untuk mendapatkan hasil yang cepat, dan “pemrograman dalam lingkup yang lebih besar” untuk memperoleh hasil-hasil dan aplikasi yang kompleks.
- Graphics, Matlab memiliki fasilitas untuk menampilkan vector dan matrices sebagai suatu grafik. Didalamnya melibatkan *high-level functions* (fungsi-fungsi level tinggi) untuk visualisasi data dua dimensi dan data tiga dimensi, *image processing, animation, dan presentation graphics*. Ini juga melibatkan fungsi level rendah yang memungkinkan bagi kita untuk membiasakan diri untuk memunculkan grafik mulai dari bentuk yang sederhana sampai dengan tingkatan graphical user interfaces pada aplikasi Matlab.
- Matlab Application Program Interface (API), merupakan suatu *library* yang memungkinkan program yang telah kita tulis dalam bahasa *C* dan *Fortran* mampu berinteraksi dengan matlab. Ini melibatkan fasilitas untuk pemanggilan *routines* dari matlab (*dynamic linking*), pemanggilan matlab sebagai sebuah *computational engine*, dan untuk membaca dan menuliskan *MAT-files*.

2. METODE

Metodologi penelitian menggunakan algoritma *fuzzy c means* dalam menentukan dosen pembimbing tugas akhir, dalam hal ini penulis melakukan beberapa tahapan sebagai berikut:

1. Penelitian Awal
Pada tahapan ini dikumpulkan bahan penelitian dari berbagai sumber pustaka, seperti buku, jurnal (baik cetak maupun *online*), prosiding, majalah, artikel dan sumber lain yang relevan dalam ilmu pengetahuan.
2. Pengumpulan Data
Sumber data dari penelitian ini dari Sekolah Tinggi Teknik Harapan program studi Teknik Informatika.
3. Inisialisasi Data
Data yang terkumpul diidentifikasi dan di klasifikasikan sesuai dengan kelompoknya. Selain itu juga menentukan validitas data dan variabel yang akan dipakai. Pada tahap inisialisasi data dilakukan pembuatan *interval* variabel *fuzzy c means*.
4. Proses Penentuan Himpunan Data
Dari beberapa variabel *fuzzy c means* yang sudah ditentukan maka selanjutnya dilakukan

penentuan himpunan data dimana jumlah cluster yang diharapkan $1 < c < N$, nilai pembobot $m > 1$, toleransi penghentian $\epsilon > 0$.

5. Inisialisasi Matriks
Hasil dari penentuan himpunan data dilakukanlah inisialisasi matriks partisi secara acak, $U(0) \in M_{fc}$.
6. Proses Penghitungan *Cluster Center (Means)*.
Dari inisialisasi matriks yang dihasilkan, maka akan dilakukan proses penghitungan *cluster center (means)* dimana, $v_i^{(1)} = \frac{\sum_{k=1}^N (\mu_{ik}^{(1-1)})^m z_k}{\sum_{k=1}^N (\mu_{ik}^{(1-1)})^m}$, $1 \leq i \leq c$
7. Proses Penghitungan Jarak
Hasil dari penghitungan cluster center, akan dilakukan penghitungan jarak dimana, $D_{ikA}^2 = (z_k - v_i^{(1)})^T A (z_k - v_i^{(1)})$, $1 \leq i \leq c$, $1 \leq k \leq N$
8. Perbaharui Matriks Partisi
Dari nilai yang sudah dihasilkan maka dilakukanlah pembaharuan matriks partisi untuk $1 \leq k \leq N$, Jika $D_{ikA} > 0$ untuk semua $i = 1, 2, \dots, c$

$$\mu_{ik}^{(i)} = \frac{1}{\sum_j^c \left(\frac{D_{ikA}}{D_{jkA}} \right)^{\frac{2}{m-1}}}$$

Atau dengan kata lain:

$$\mu_{ik}^{(i)} = 0 \text{ jika } D_{ikA} > 0 \text{ dan } \mu_{ik}^{(i)} \in [0,1] \text{ dengan } \sum_{i=1}^c \mu_{ik}^{(i)} = 1$$

Ulangi sampai $\|U^{(l)} - U^{(l-1)}\| < \epsilon$.

9. Kesimpulan
Dari beberapa proses yang dilakukan, maka dapat diambil kesimpulan bahwa akan terus dilakukan proses iterasi sampai hasil yang diperoleh maksimal dan sesuai dengan yang penulis harapkan.

Dalam uji kasus pada sistem *clustering* diperlukan data yang akan menjadi sumber acuan dan pembuatan model tersebut, adapun prosedur dari pemodelan data dapat dijelaskan dari beberapa langkah sebagai berikut:

1. Data masukan diambil dari data Universitas Harapan Medan rogram studi Teknik Informatika.
2. Data merupakan data asli yang kemudian dilanjutkan dengan proses himpunan data. Proses himpunan data bertugas mengubah data asli dari masukan menjadi data *fuzzy* berdasarkan himpunan *fuzzy* yang telah ditetapkan.
3. Setelah menjadi data *fuzzy* kemudian dilanjutkan ke penghitungan matriks partisi, dimana terlebih dahulu dengan ditentukan

jumlah *cluster* yang diharapkan dengan nilai bobot dan nilai toleransi yang selanjutnya akan dilakukan penghitungan secara berulang sampai diperoleh hasil yang diharapkan.

4. Keluaran yang dihasilkan dari proses penghitungan *fuzzy c means* yang telah dilakukan adalah hasil pembagian pembimbing.

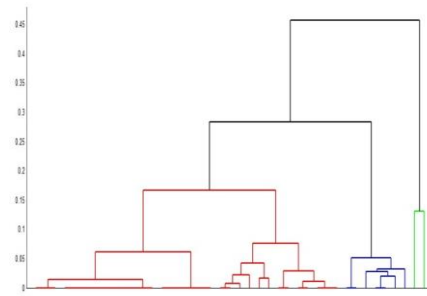
3. PEMBAHASAN

Sesuai dengan tujuan penelitian ini yaitu menentukan dosen pembimbing tugas akhir maka data yang diperlukan adalah data sekunder dari Universitas Harapan Medan program studi Teknik Informatika. Data tersebut merupakan data dosen tetap program studi Teknik Informatika dari tahun 2015-2017 dimana data tersebut terdapat nama dosen beserta dengan NIDN, pendidikan dan jabatan fungsional. Jumlah data hanya 41, karena dosen tetap program studi Teknik Informatika berjumlah 41 orang.

Untuk menentukan pembagian dosen pembimbing tugas akhir maka data akan dibagi menjadi tiga bagian sesuai dengan bidang keahlian yang ada di program studi Teknik Informatika Universitas Harapan Medan, yaitu bidang Robotika, Multimedia dan Jaringan.

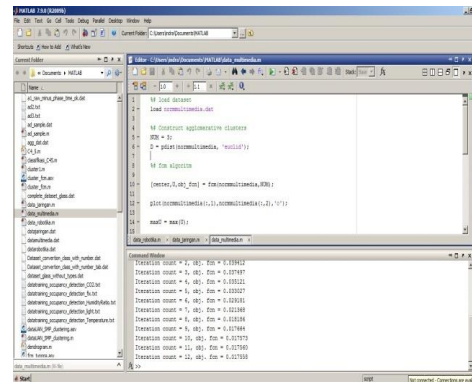
Data yang telah didapatkan akan diuji menggunakan aplikasi Matlab. Adapun hasil pengujian matlab yang telah dilakukan dapat dilihat pada gambar 3.1 – 3.6.

jaringan. Hasil dari pengujian tersebut juga dapat dilihat pada gambar 3.2.

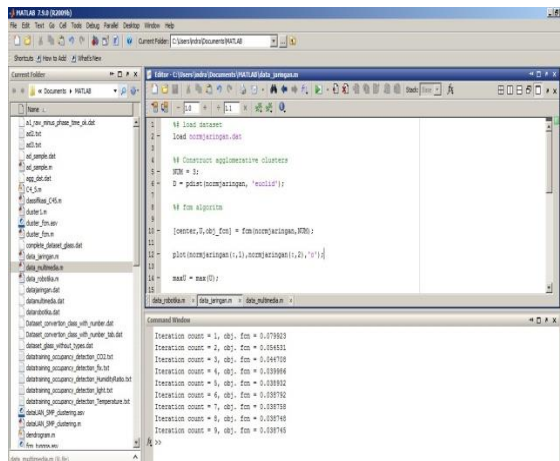


Gambar 2 Grafik Pengujian Matlab Pada Bidang Jaringan

Hasil pengujian matlab dalam bidang matlab dapat dilihat pada gambar 3.3 dan 3.4.



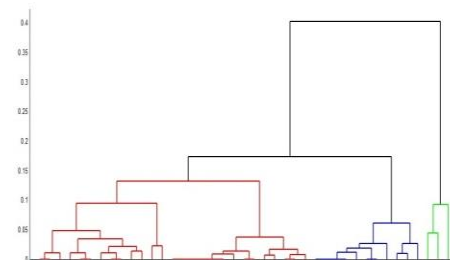
Gambar 3 Pengujian Matlab Pada Bidang Multimedia



Gambar 1 Pengujian Matlab Pada Bidang Jaringan

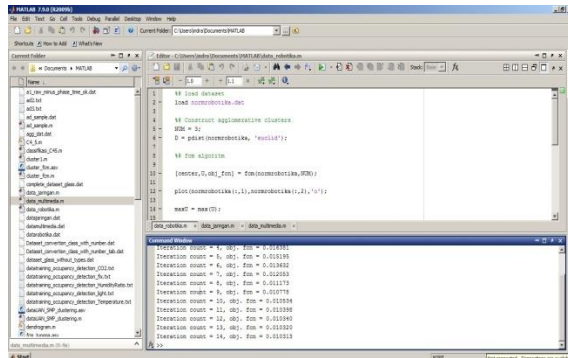
Dari gambar 1 terlihat bahwa hasil normalisasi, derajat keanggotaan dan hasil cluster pada bidang jaringan menggunakan aplikasi matlab. Hasil cluster menunjukkan bahwa untuk dosen yang tidak boleh membimbing pada jaringan terletak pada cluster C1, sementara untuk cluster C2 dan C3 boleh membimbing pada bidang jaringan. Dan sangat jelas bahwa lebih banyak dosen yang tidak boleh membimbing pada bidang

Dari gambar 3.3 terlihat bahwa hasil normalisasi, derajat keanggotaan dan hasil cluster pada bidang multimedia menggunakan aplikasi matlab. Hasil cluster menunjukkan bahwa untuk dosen yang tidak boleh membimbing pada multimedia terletak pada cluster C1, sementara untuk cluster C2 dan C3 boleh membimbing pada bidang jaringan. Dan sangat jelas bahwa lebih banyak dosen yang tidak boleh membimbing pada bidang multimedia. Hasil dari pengujian tersebut juga dapat dilihat pada gambar 3.4.



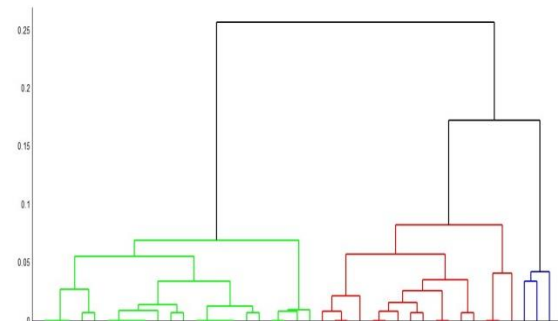
Gambar 4 Grafik Pengujian Matlab Pada Bidang Multimedia

Hasil pengujian matlab pada bidang robotika dapat dilihat pada gambar 3.5 dan 3.6.



Gambar 5 Pengujian Matlab Pada Bidang Robotika

Dari gambar 3.5 terlihat bahwa hasil normalisasi, derajat keanggotaan dan hasil cluster pada bidang multimedia menggunakan aplikasi matlab. Hasil cluster menunjukkan bahwa untuk dosen yang tidak boleh membimbing pada multimedia terletak pada cluster C1, sementara untuk cluster C2 dan C3 boleh membimbing pada bidang jaringan. Dan sangat jelas bahwa lebih banyak dosen yang tidak boleh membimbing pada bidang multimedia. Hasil dari pengujian tersebut juga dapat dilihat pada gambar 3.6.



Gambar 6 Grafik Pengujian Matlab Pada Bidang Robotika

Dari semua proses yang telah dilakukan, maka dapat disimpulkan bahwa bidang keahlian dosen dan penelitian dosen harus sesuai dengan topik tugas akhir dari mahasiswa sehingga pembagian dosen pembimbing akan adil dan merata.

4. KESIMPULAN

Kesimpulan yang dapat diambil dari penelitian ini adalah sebagai berikut :

1. Penentuan dosen pembimbing berdasarkan kepangkatan, jumlah penelitian dan mata kuliah pada bidang keahliannya.
2. Penelitian ini menggunakan aplikasi matlab sebagai pengujian.
3. Algoritma *fuzzy c-means* adalah salah satu algoritma dalam metode clustering yang sangat baik dalam melakukan clusterisasi terhadap data dalam kondisi high level, untuk itu harus memperbanyak jumlah variabel data.

REFERENSI

- [1] Hafisah., Rustamaji, H.J., & Inayanti, Y. 2008. Sistem Pendukung Keputusan Pemilihan Jurusan Di Smu Dengan Logika Fuzzy. *Seminar Nasional Informatika*.pp. 213-218.
- [2] Haryati, S., Sudarsono, A., & Suryana, E. 2015. Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu). *Jurnal Media Infotama***11**(2): 30-138.
- [3] Jayanti, S. & Hartati, S. 2012. Sistem Pendukung Keputusan Seleksi Anggota Paduan Suara Dewasa Menggunakan Metode Fuzzy Mamdani. *Seminar Nasional Matematika dan Pendidikan Matematika*.pp. 55-66.
- [4] Kusriani & Luthfi, E.T. 2009. Algoritma Data Mining, Penerbit Andi, Yogyakarta.
- [5] Mantho, C. J. 2015. Analisis Segmentasi Citra USG Hati Menggunakan Metode Fuzzy C Means. *Citec Journal***2**(3): 256-264.
- [6] Mirza, M. F. 2009. Metode Clustering Dengan Algoritma Fuzzy C Means Untuk Rekomendasi Pemilihan Bidang Keahlian Pada Program Studi Teknik Informatika. *Jurnal Nasional***11**: 1-12.
- [7] Susanto, H., & Sudiyatno. 2014. Data Mining Untuk Memprediksi Prestasi Siswa Berdasarkan Sosial Ekonomi, Motivasi, Kedisiplinan, dan Prestasi Masa Lalu. *Jurnal Pendidikan Vokasi***4**(2): 222-231
- [8] Arifah, T, Murnomo, A., &Suryanto, A. 2017. Implementasi *Neural Network* Pada Matlab Untuk Prakiraan Konsumsi Beban Listrik Kabupaten Ponorogo Jawa Timur. *Jurnal Teknik Elektro* **9**(1): 7-12