

Perbandingan Metode *Fuzzy C-Means* dan *K-Means* Dalam Klasifikasi Kelulusan Mahasiswa (Studi Kasus : Jurusan Manajemen, Universitas Maritim Raja Ali Haji)

Alena Uperiati¹, Martaleli Bettiza^{2,*}, Atika Puspasari³
^{1,2,3}Jurusan Teknik Informatika, Fakultas Teknik, Universitas Maritim Raja Ali Haji
^{1,2,3}Jl. Politeknik Senggarang, Tanjungpinang 29100
*Corresponding Author: mbettiza@umrah.ac.id

Abstract— Imbalance between the number of students entering and leaving causes a buildup of students, where students enter in large numbers but the number that passed on time is much smaller. Therefore the research was conducted to classify student graduation using the fuzzy c-means and k-means. Comparison method fuzzy c-means and k-means is used to get accurate classification document. Based on the result of the testing that has been done on fuzzy c-means, the best parameter that produces a near-accurate classification result is the rank 3, maximum iteration of 50 and the error value 0.01 obtained an accuracy value of 81.91%. Where in the test the accuracy of the k-means is 63 %. This proves that fuzzy c-means method is more accurate compared to method k-means in the students graduation classification.

Keywords— Classification, Fuzzy C-means, K-means, Accuracy

Intisari— Ketidakseimbangan antara jumlah mahasiswa yang masuk dan keluar menyebabkan penumpukan jumlah mahasiswa, dimana mahasiswa masuk dalam jumlah besar namun jumlah yang lulus tepat waktu jumlahnya jauh lebih kecil. Oleh karena itu, pada penelitian ini dilakukan guna untuk mengklasifikasikan kelulusan mahasiswa menggunakan metode *fuzzy c-means* dan *k-means*. Perbandingan metode *fuzzy c-means* dan *k-means* digunakan untuk mendapatkan hasil klasifikasi yang tepat dan akurat. Berdasarkan hasil pengujian yang telah dilakukan pada *fuzzy c-means*, parameter terbaik yang menghasilkan hasil klasifikasi yang mendekati akurat adalah nilai pangkat (w) 3, maksimum iterasi 50 dan nilai error 0.01 didapatkan nilai akurasi sebesar 81.91%. Sedangkan pada pengujian tingkat akurasi metode *k-means* didapatkan nilai tingkat akurasi sebesar 63 %. Hal ini membuktikan bahwa metode *fuzzy c-means* lebih akurat dibandingkan dengan metode *k-means* dalam klasifikasi kelulusan mahasiswa

Kata kunci— Klasifikasi, *Fuzzy C-means*, *K-means*, Tingkat Akurasi.

I. PENDAHULUAN

Berdasarkan hasil pengamatan di Universitas Maritim Raja Ali Haji, ditemukan ketidakseimbangan antara jumlah mahasiswa yang masuk dan keluar. Mahasiswa masuk dalam jumlah besar, namun jumlah yang lulus

tepat waktu sesuai dengan ketentuan jauh lebih kecil dibandingkan jumlah mahasiswa yang masuk ke Universitas Maritim Raja Ali Haji Tanjungpinang. Pada tahun 2014 jumlah mahasiswa masuk tidak sebanding dengan mahasiswa yang lulus tepat pada waktunya.

Dalam jangka panjang, hal tersebut dapat menyebabkan terjadinya penumpukan jumlah mahasiswa di Universitas Maritim Raja Ali Haji. Masalah ini dapat diatasi dengan meningkatkan kualitas pembelajaran dan layanan akademik untuk mahasiswa. Selain itu, jika waktu penyelesaian studi mahasiswa dapat diprediksi maka penanganan mahasiswa akan lebih efektif.

Salah satu teknik melakukan prediksi yang dapat digunakan adalah dengan teknik penggalian data atau data mining. Penggalian data atau data mining berdasarkan data pendidikan di perguruan tinggi dapat meningkatkan kualitas pembelajaran mahasiswa di perguruan tinggi. Algoritma data mining yang digunakan dalam penelitian ini adalah algoritma *Fuzzy C-Means* dan *K-Means*.

Algoritma *Fuzzy C-Means* pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981. Metode *Fuzzy C-Means* ini cara kerjanya adalah dengan mencari nilai pusat dari acuan ketentuan cluster tersebut dikategorikan sebagai cluster jenis apa. Langkah selanjutnya adalah menghitung nilai derajat keanggotaan pada setiap data. Derajat keanggotaan dari setiap data dipergunakan untuk mengetahui data tersebut cenderung ke cluster jenis apa [1].

Sedangkan berdasarkan algoritma dasar *k-means*, adalah dimana cluster sepenuhnya tergantung kepada pemilihan centroid cluster awal. Elemen data *K* dipilih sebagai pusat awal. Lalu jarak semua data elemen di hitung dengan rumus jarak Euclidean. Elemen data yang memiliki jarak kurang ke centroid di pindahkan ke cluster yang sesuai. Proses dilanjutkan hingga tidak ada lagi perubahan yang terjadi dalam pengelompokan [2].

II. LANDASAN TEORI

A. Normalisasi

Normalisasi data sangat diperlukan sebelum proses data mining supaya tidak ada parameter yang mendominasi dalam perhitungan jarak antar data. Normalisasi data yang digunakan yaitu normalisasi skor *z* (*zero-mean normalization*) yang merupakan salah satu

metode atau teknik yang diterapkan untuk normalisasi data, dimana nilai dari atribut *X* dinormalisasikan berdasarkan rata-rata dan standar deviasi dari atribut *X* dengan rumus [3]:

$$x_k = \frac{1}{N} \sum_{i=1}^N x_{ik} \quad (1)$$

$$\sigma_k^2 = \frac{1}{N-1} \sum_{i=1}^N (x_{ik} - x_k)^2 \quad (2)$$

Data hasil normalisasi dapat dihitung menggunakan cara pertama :

$$x_{ik}' = \frac{x_{ik} - x_k}{\sigma_k} \quad (3)$$

B. *Fuzzy C-Means*

Fuzzy C-Means adalah teknik pengelompokan data yang keberadaan tiap-tiap data dalam suatu kluster ditentukan oleh nilai derajat keanggotaan. Dengan cara memperbaiki pusat kluster dan nilai keanggotaan tiap-tiap data secara berulang, maka akan terlihat bahwa pusat kluster yang bergerak menuju lokasi yang tepat [1].

Algoritma *Fuzzy C-Means* seperti berikut ini:

- 1) Membentuk matriks partisi awal (U^0). Data dengan matriks $m \times n$.

$$\begin{bmatrix} (x_1) & (x_2) & (x_n) \\ (x_1) & (x_2) & (x_n) \end{bmatrix} \quad (4)$$

- 2) Menentukan beberapa aturan seperti:
 - a. Menentukan jumlah data yang akan dicluster (n) dan jumlah variabel (m)
 - b. Jumlah cluster yang akan dibentuk (C)
 - c. Pangkat (w)
 - d. Maksimal iterasi
 - e. Kriteria penghentian (ϵ)
 - f. Iterasi awal, $t = 1$
- 3) Setiap data mempunyai derajat keanggotaan pada setiap kelompok, dinyatakan dengan U_{ij} , dengan nilai antara 0 dan 1. Jumlah nilai derajat keanggotaan setiap data selalu sama dengan 1

$$\sum_{k=1}^c \mu_{ik} = 1 \quad (5)$$

- 4) Menghitung pusat cluster sebagai berikut:

$$V_{kj} = \frac{\sum_{i=1}^n (\mu_{ik})^w \cdot x_{ij}}{\sum_{i=1}^n (\mu_{ik})^w} \quad (6)$$

5) Menghitung fungsi objektif (P_t) pada iterasi ke-t dibawah ini:

$$P_t = \sum_{i=1}^n \sum_{k=1}^c [\sum_{j=1}^m (x_{ij} - v_{kj})^2 (\mu_{ik})^w] \quad (7)$$

6) Menghitung perubahan matriks partisi baru (μ'_{ik}):

$$\mu'_{ik} = \frac{((\sum_{j=1}^m (x_{ij} - v_{kj})^2)^{\frac{-1}{w-1}})}{\sum_{k=1}^c ((\sum_{j=1}^m (x_{ij} - v_{kj})^2)^{\frac{-1}{w-1}}} \quad (8)$$

7) Memeriksa kondisi berhenti

- Jika $|P_t - P_{t-1}| < \epsilon$ atau ($t > \text{Maxiter}$) maka berhenti.
- Jika tidak $t = t + 1$, mengulang ke langkah 3.

C. K-Means

Algoritma clustering yang digunakan untuk mempartisi data kedalam beberapa cluster, dimana data yang memiliki tingkat kemiripan yang tinggi dikelompokkan dalam satu cluster sedangkan data yang memiliki karakteristik yang berbeda dikelompokkan ke dalam cluster yang berbeda [4].

Langkah-langkah penyelesaian algoritma *k-means* sebagai berikut:

- Identifikasi data yang akan dicluster dan juga tentukan jumlah cluster data, X_{ij} ($i = 1, 2, \dots, n; j = 1, 2, \dots, m$) dimana n adalah jumlah data yang akan di cluster dan m adalah jumlah variabel data.
- Pada awal iterasi pusat setiap cluster ditentukan secara acak (sembarang), C_{kj} ($k = 1, \dots, k; j = 1, \dots, m$).
- Tentukan jarak (d) setiap data terhadap pusat cluster dengan menggunakan formula Euclidean sebagai berikut:

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \quad (9)$$

- Kelompokkan data berdasarkan jarak minimum data terhadap pusat cluster.
- Cek kondisi apakah ada data yang masih pindah pada kelompok lain, jika ia lakukan iterasi berikutnya dengan menghitung pusat cluster (C_{kj}) berdasarkan nilai rata-rata (p)

dari data yang menjadi anggota pada cluster yang terbentuk hasil interaksi sebelumnya.

$$c_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p} \quad (10)$$

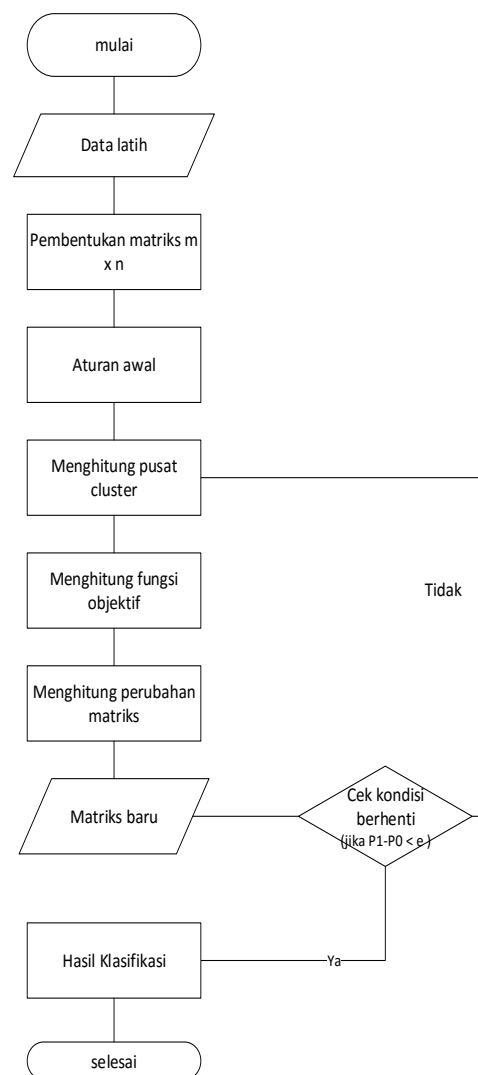
Dimana $a X_{ij} \in \text{cluster ke-}k$ dan adalah jumlah anggota cluster

6) Jika tidak ada lagi data yang pindah pada cluster lain, maka proses interaksi berhenti.

D. Tingkat Akurasi

Untuk mencari akurasi klasifikasi adalah dengan membagi hasil yang berhasil atau benar dengan jumlah data percobaan yang selanjutnya dikali dengan 100%. Berikut adalah rumus yang digunakan untuk menghitung akurasi [5]:

$$\text{Akurasi} = \frac{\text{Hasil Benar}}{\text{Jumlah Total Data}} \times 100\% \quad (11)$$

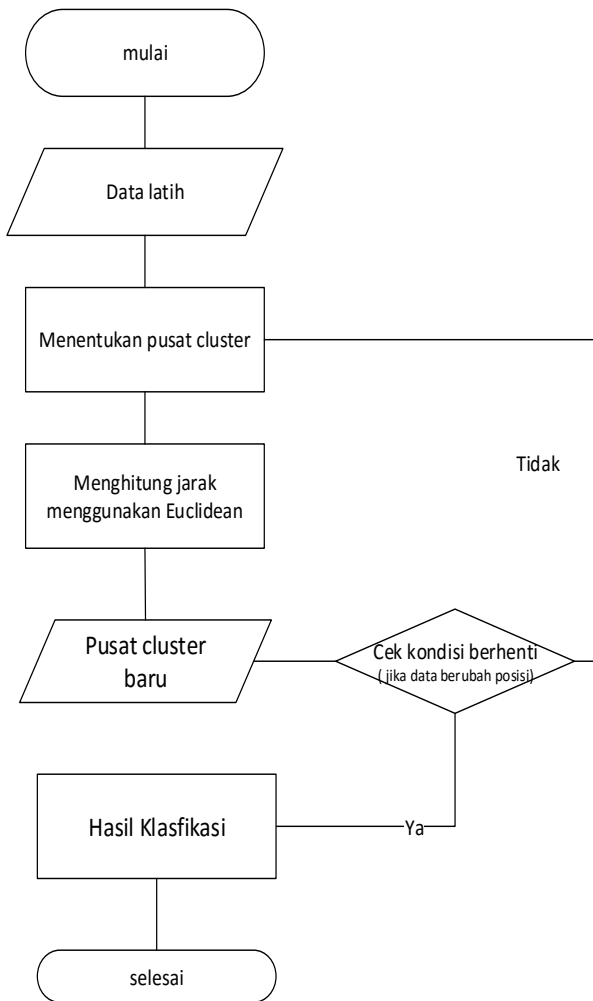


Gambar 1. Flowchart fuzzy c-means

E. Perancangan Sistem

Pada alur data kerja sistem ini nantinya akan menghitung variabel yang telah dilakukan normalisasi dengan menggunakan metode *fuzzy c-means* dan *k-means* kemudian dibandingkan berdasarkan tingkat akurasi dari kedua metode tersebut (Gambar 1).

Pada gambaran umum sistem untuk metode *fuzzy c-means*, sistem dimulai dengan membentuk matriks awal kemudian menginput nilai pangkat (w), maksimum iterasi (maksit) dan nilai error (e) ke dalam sistem kemudian akan diproses dengan menghitung pusat cluster dan fungsi objektif yang akan digunakan untuk mendapatkan matriks baru untuk hasil klasifikasi nantinya.



Gambar 2. Flowchart k-means

Pada alur kerja sistem untuk metode *k-means* seperti yang terlihat pada Gambar 2,

sistem dimulai dengan menentukan pusat cluster yang diambil acak berdasarkan data yang telah dilakukan normalisasi. Kemudian sistem melakukan proses perhitungan dengan menghitung jarak menggunakan Euclidean untuk menentukan pusat cluster baru. Setelah tidak terjadi perpindahan pada pusat cluster maka didapatkan hasil klasifikasinya.

III. PENGUJIAN DAN ANALISIS

Pada penelitian perbandingan metode *fuzzy c-means* dan *k-means* yang dilakukan untuk klasifikasi kelulusan mahasiswa manajemen angkatan tahun 2014 Universitas Maritim Raja Ali Haji. Penelitian ini menggunakan 94 data dengan 2 klasifikasi yaitu tepat waktu dan tidak tepat waktu.

A. Pengujian Tingkat Akurasi Fuzzy C-Means

Pengujian ini bertujuan untuk mengetahui pada jumlah nilai pangkat (w), jumlah maksimum iterasi (maksit), dan nilai error mana yang menghasilkan klasifikasi dengan tingkat akurasi terbaik yang ditunjukkan pada Tabel 1.

Tabel 1. Hasil klasifikasi menggunakan parameter terbaik

no	ips	sks	nilai kp	skripsi	Data aktual	Hasil klasifikasi
1	3.57	23	86	74	Tepat Waktu	Tidak Tepat Waktu
2	3.08	18	77	81	Tepat Waktu	Tepat Waktu
3	3.71	21	79	81	Tepat Waktu	Tepat Waktu
4	3.64	21	85	78	Tepat Waktu	Tidak Tepat Waktu
5	3.14	21	81	80	Tepat Waktu	Tepat Waktu
6	3.43	21	86	84	Tepat Waktu	Tepat Waktu
7	3.56	24	85	85	Tepat Waktu	Tidak Tepat Waktu
8	3.25	24	79	81	Tepat Waktu	Tepat Waktu
9	3.43	21	83	79	Tepat Waktu	Tepat Waktu
10	2.93	21	79	80	Tepat Waktu	Tepat Waktu
11	3.42	18	100	81	Tepat Waktu	Tepat Waktu
12	3.25	24	90	76	Tepat Waktu	Tepat Waktu

13	3.21	21	85	83	Waktu Tepat Waktu	Waktu Tepat Waktu
14	3.57	23	83	84	Tepat Waktu	Tidak Tepat Waktu
...
94	3.81	24	86	0	Tidak Tepat Waktu	Tidak Tepat Waktu

Setelah dilakukan beberapa kali pengujian dengan parameter yang berbeda maka didapatkan hasil klasifikasi dengan parameter terbaik pada percobaan kedua. Berikut hasil klasifikasi berdasarkan tingkat akurasi yang dapat dilihat pada Tabel 2.

Tabel 2. Tingkat Akurasi *Fuzzy C-Means*

no	w (nilai pangkat)	Maksimum iterasi	e (nilai error)	Tingkat akurasi
1	2	50	0.01	71.33 %
2	3	50	0.01	81.91 %
3	4	50	0.01	37.23 %
4	5	50	0.01	44.53 %
5	2	100	0.02	63.66 %
6	3	100	0.02	70.63 %
7	4	100	0.02	50,69 %
8	5	100	0.02	30,76 %
9	2	50	0.02	43.98 %
10	3	50	0.02	75.21 %
11	4	50	0.02	28.98 %
12	5	50	0.02	51.77 %

B. Pengujian Tingkat Akurasi *K-Means*

Pengujian ini bertujuan untuk mengetahui hasil klasifikasi metode *k-means* dengan tingkat akurasi terbaik setelah dilakukan beberapa kali percobaan. Setelah dilakukan beberapa kali pengujian maka didapatkan hasil klasifikasi. Berikut hasil klasifikasi berdasarkan tingkat akurasi yang dapat dilihat pada Tabel 3.

Tabel 3. Hasil Klasifikasi *K-Means*

Iterasi ke-	Tingkat Akurasi
-------------	-----------------

no		
1	3	58 %
2	4	63 %
3	3	58 %
4	2	26.66 %
5	4	63 %
6	5	42.76 %
7	2	26.66 %
8	4	63 %
9	4	63 %
10	5	42.76 %
11	3	58 %
12	2	26.66 %

C. Analisis

Dalam studi kasus ini perhitungan nilai tingkat akurasi digunakan untuk mengetahui seberapa akuratnya metode yang digunakan dalam penelitian ini. Dari hasil pengujian didapatkan bahwa tingkat akurasi *k-means* dan *c-means* dapat dilihat pada Tabel 4 berikut.

Tabel 4. Perbandingan Tingkat Akurasi

no	<i>Fuzzy C-Means</i>	<i>K-Means</i>
1	71.33 %	58 %
2	63.66 %	63 %
3	81.91 %	58 %
4	37.23 %	26.66 %
5	63.66 %	63 %
6	70.63 %	42.76 %
7	50,69 %	26.66 %
8	30,76 %	63 %
9	43.98 %	63 %
10	75.21 %	42.76 %
11	28.98 %	58 %
12	51.77 %	26.66 %

Dari hasil pegujian tingkat akurasi didapatkan perbandingan tingkat akurasi antara fuzzy c-means dan k-means. Dimana pada metode fuzzy c-means tingkat akurasi dipengaruhi berdasarkan inputan nilai pangkat, maksimum iterasi, nilai error. Sedangkan pada metode k-means dimana iterasi keberapa baru didapatkan tingkat akurasinya pada pengujian ini iterasi yang sering muncul adalah iterasi ke-4.

IV. KESIMPULAN

Kesimpulan yang didapatkan dalam penelitian ini adalah :

1. Penggunaan metode fuzzy c-means dan k-means sama-sama dapat digunakan dalam melakukan klasifikasi kelulusan mahasiswa.
2. Pada pengujian klasifikasi menggunakan fuzzy c-means mendapatkan parameter terbaik pada percobaan kedua dimana parameter w (nilai pangkat) = 3 , maksit (maksimum iterasi) = 50, dan ϵ (nilai error) = 0.01 dengan tingkat akurasi sebesar 81.91 %.
3. Pada pengujian tingkat akurasi menggunakan k-means didapatkan tingkat akurasi terbaik pada percobaan pertama sebesar 63 %.
4. Dari pengujian metode fuzzy c-means dan metode k-means didapatkan kesimpulan bahwa tingkat akurasi klasifikasi kelulusan mahasiswa lebih baik hasil akurasinya menggunakan metode fuzzy c-means sebesar 81.91 % dibandingkan metode k-means yang tingkat akurasinya hanya sebesar 63 %.

V. SARAN

Pada penelitian ini untuk mengklasifikasikan atau mengelompokkan kelulusan mahasiswa yang menggunakan metode fuzzy c-means dan k-means. Untuk penelitian selanjutnya dapat membandingkan metode klasifikasi lainnya atau menambah variabel yang akan digunakan.

UCAPAN TERIMA KASIH

Terima kasih kami ucapkan kepada Universitas Maritim Raja Ali Haji yang telah memberikan dukungan dan wadah untuk mempublikasikan penelitian ini.

REFERENSI

- [1] W. Anggraeni, "Penentuan Nilai Pangkat Pada Algoritma *Fuzzy C-Means*". *Factor Exacta*, vol. 8, no. 3, hlm. 266–278, 2015.
- [2] U. R. Raval, dan C. Jani "Implementing & Improvisation of K-means Clustering Algorithm". *International Journal of Computer Science and Mobile Computing*, vol. 5, no. 5, hlm. 191–203, Mei 2016.
- [3] B. E. Adiana, I. Soesanti, A. E. Permanasari, "Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering," *JUTEI.*, vol. 2, no. 1, hlm. 25–32, April 2018.
- [4] N. Butarbutar, A. P. Windarto, P. Hartama, dan Solikhun, "Komparasi Kinerja Algoritma *Fuzzy C-Means* dan *K-Means* Dalam Pengelompokan Data Siswa Berdasarkan Prestasi Nilai Akademik Siswa (studi kasus : SMP Negeri 2 Pemantangsiantar)," *Jurnal Riset Sistem Informasi dan Teknik Informatika (JURASIK)*, vol. 1, no. 1, hlm. 46–55, 2016.
- [5] H. Simaremare, dan A. Kurniawan, "Perbandingan Akurasi Pengenalan Wajah Menggunakan Metode LBPH dan *Eigenface* dalam Mengenali Tiga Wajah Sekaligus secara *Real-Time*," *Jurnal Sains, Teknologi dan Industri.*, vol. 14, no. 1, 2014.
- [5] A. Aziz, "Aplikasi Algoritma Fuzzy C-Means Clustering Untuk Pengelompokan Lulusan," *Jurnal Ilmiah Teknologi dan Informasi ASIA.*, vol. 3, no. 2, hlm. 35-43, 2009.
- [6] M. R. Faiz, "Pengklasifikasian Lulusan Jurusan Teknik Elektro Berdasarkan Nilai IPK dengan Metode *Fuzzy Clustering*," vol. 3, hlm. 55-60, 2010.
- [7] A. Indriani, "Klasifikasi Data Forum dengan Menggunakan Metode *Naïve Bayes Classifier*," *J Seminar Nasional Aplikasi Teknologi Informasi (SNATI).*, 5-10, 2014.
- [8] A. Praja, C. Lubis, dan D. E. Herdiwindiati, "Deteksi Penyakit Diabetes dengan Metode Fuzzy C-Means Clustering dan K-Means Clustering," *Journal of*

- Computer Science and Information Systems*, vol. 1, hlm. 11–24.
- [9] S. F. Rudiensyah, “*K-Means* dan *Fuzzy C-Means* Pada Analisis Data Polusi Udara di Kota X.” *Seminar Nasional Teknologi Informasi dan Multimedia*, vol. 2, no. 1, hlm. 25–30, 2017.
- [10] C. L. Simbolon, N. Kusumastuti, dan B. Irawan, “Clustering lulusan mahasiswa matematika fmipa untan pontianak menggunakan algoritma fuzzy c - means,” *Buletin Ilmiah Mat. Stat. Dan Terapan(Bimaster.)*, vol. 2, no. 1, hlm. 21–36, 2013.