

Perbandingan Penerapan Algoritme *K-Means* dan *Fuzzy C-Means* untuk Mengelompokkan Data Kinerja Dosen Universitas Islam Bandung

Shavira Siti Rachmasari*, Abdul Kudus

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Islam Bandung, Indonesia.

*shavirasr66@gmail.com, akudus69@unisba.ac.id

Abstract. In Data mining, the clustering method is divided into two, namely cluster hierarchies and non-hierarchical clusters. In a hierarchical cluster start grouping with two or more objects that have the closest similarity to, a non-hierarchical cluster starting with the number of clusters first. Among the non-hierarchical cluster methods are the K-Means method and the Fuzzy C-Means (FCM) method. The K-Means method is hard clustering in contrast to FCM which includes fuzzy clustering. Judging from previous studies, these two methods produced the best clusters. This thesis explores the optimal k value using principal component analysis, the final result is 4 clusters. Clustering on the performance data of Bandung Islamic University lecturers for the February 2021 period using the second method resulted in a different number of cluster members. The second comparison is based on the validation of the SI value of the K-Means which is 0.55 while the SI value of the FCM is 0.05. From these results, it shows that the K-Means algorithm is superior to the Fuzzy C-Means algorithm in clustering data on the performance of lecturers at the Islamic University of Bandung. Based on the validation of the cluster that a value closer to 1 has a better cluster quality. The results of the K-Means clustering are in cluster 1 consisting of lecturers who do not have a Sinta accredited journal, cluster 2 consists of lecturers who have a minimum of 5 conference articles, lecturers with Q3 class journals, and do not have a Sinta accredited journal, cluster 3 consists of lecturers who have at least 3 articles, journals / books that are lecturers with Q2 class journals and do not have Sinta accredited journals, and cluster 4 consists of lecturers with the number of articles in S6 class.

Keywords: Clustering, K-Means, Fuzzy C-Means, Silhoutte Index.

Abstrak. Dalam Data Mining metode pengklasteran dibagi menjadi dua yaitu *cluster* hirarki dan non-hirarki. Pada *cluster* hirarki memulai pengelompokkan dengan dua atau lebih objek yang mempunyai kesamaan paling dekat sedangkan pada *cluster* non hirarki justru dimulai dengan penentuan jumlah *cluster* terlebih dahulu. Diantaranya yang termasuk metode *cluster* non hirarki adalah metode *K-Means* dan metode *Fuzzy C-Means* (FCM). Metode *K-Means* bersifat *hard clustering* berbeda dengan *FCM* yang termasuk *fuzzy clustering*. Dilihat dari penelitian-penelitian sebelum nya kedua metode ini menghasilkan *cluster* terbaik. Pada skripsi ini mengeksplorasi nilai k optimal menggunakan analisis komponen utama, didapat hasil akhirnya yaitu 4 *cluster*. Pengklasteran pada data kinerja dosen Universitas Islam Bandung pada periode Februari 2021 menggunakan kedua metode tersebut menghasilkan jumlah anggota *cluster* yang berbeda. Perbandingan kedua metode berdasarkan validasi nilai SI dari *K-Means* yaitu 0.55 sedangkan nilai SI dari FCM yaitu 0.05. Dari hasil tersebut menunjukkan algoritma *K-Means* lebih unggul dibandingkan algoritma *Fuzzy C-Means* dalam pengklasteran data kinerja dosen Universitas Islam Bandung. Berdasarkan ketentuan dari validasi *cluster* bahwa nilai yang lebih mendekati 1 mempunyai kualitas *cluster* yang semakin baik. Hasil pengklasteran *K-Means* yaitu pada *cluster* 1 beranggotakan dosen yang belum memiliki jurnal terakreditasi Sinta, *cluster* 2 beranggotakan dosen yang memiliki

minimal 5 jumlah artikel dikonferensi, dosen dengan jurnal kelas Q3, dan belum memiliki jurnal terakreditasi Sinta, *cluster* 3 beranggotakan dosen yang memiliki minimal 3 artikel, jurnal, /buku yang dipublikasikan, dosen dengan jurnal kelas Q2 dan belum memiliki jurnal terakreditasi Sinta, dan *cluster* 4 beranggotakan dosen dengan jumlah artikel di jurnal kelas S6.

Kata Kunci: Pengkasteran, *K-Means*, *Fuzzy C-Means*, *Silhouette Index*.

1. Pendahuluan

Universitas Islam Bandung (UNISBA) merupakan salah satu perguruan tinggi yang turut berupaya untuk memajukan pendidikan di Indonesia. Misi yang diemban adalah menyelenggarakan pendidikan atau pengajaran untuk menghasilkan sumber daya manusia tidak hanya unggul di bidang akademik juga unggul dalam agama yaitu beriman, bertakwa dan berakhlakul karimah sebagai mujahid, mujtahid, dan mujaddid. Peran dosen selain mengajar dosen juga membuat penelitian seperti membuat jurnal, buku, artikel dan karya ilmiah lainnya untuk mengembangkan dan menyebarkan ilmu. SINTA (*Science and Technology Index*) merupakan *website* sebagai wadah hasil penelitian para peneliti maupun dosen untuk dipublikasikan secara *online* menciptakan *big data* yang tersimpan di database.

Data Mining adalah suatu proses untuk pengumpulan atau mengeksplorasi informasi penting dari suatu *big data*. Salah satu metode dalam *data mining* adalah menggunakan metode pengklasteran data. Menurut Tan (2006) pengklasteran adalah sebuah proses untuk mengelompokkan data ke dalam beberapa *cluster* sehingga data dalam satu *cluster* memiliki tingkat homogenitas (kesamaan) yang maksimum dan data antar *cluster* memiliki heterogenitas (perbedaan) yang tinggi. Metode pengklasteran dapat dibagi menjadi dua yaitu *cluster* hirarki dan non-hirarki. Perbedaan pada *cluster* hirarki dan non-hirarki yaitu terletak pada penentuan jumlah *cluster*, pada metode hirarki memulai pengelompokkan dengan dua atau lebih objek yang mempunyai kesamaan paling dekat sedangkan pada metode non-hirarki justru dimulai dengan terlebih dahulu menentukan jumlah *cluster* yang diinginkan. Diantaranya yang termasuk metode *cluster* non hirarki adalah metode *K-means* dan metode *Fuzzy C-means* (FCM). *K-means* bersifat *hard clustering* yaitu data secara tegas dibagi ke dalam beberapa *cluster*, sedangkan FCM bersifat *fuzzy clustering* yaitu untuk mengklasteran data dilihat dari nilai derajat keanggotaan yang jangkauan nilainya yaitu 0 sampai 1.

Pada penelitian-penelitian sebelumnya menerapkan algoritma *K-means* dan FCM sebagai penelitian mengelompokkan data, diantaranya: Nur Wakhidah dalam penelitiannya yang berjudul “*Clustering* menggunakan *K-means Algorithm*” menyimpulkan bahwa *K-means* algoritma yang sederhana dan mampu menyelesaikan permasalahan yang ada. Siti Mujilmahwati, Retno Wardhani (2021) dalam penelitiannya yang berjudul “Implementasi *Fuzzy C-means* untuk *clustering* mahasiswa berdasarkan nilai masuk perguruan tinggi” menyimpulkan bahwa dengan metode FCM penentuan kelompok menjadi lebih tepat dan cepat ditentukan. Pada hasil kedua penelitian tersebut menyebutkan bahwa algoritma *K-means* dan FCM baik untuk mengelompokkan data. Pengujian validasi *cluster* menggunakan nilai *Silhouette Index* adalah salah satu cara untuk melihat kualitas dan kekuatan *cluster*, dan menentukan metode mana yang lebih baik digunakan untuk *clustering*. Berdasarkan penjelasan latar belakang di atas, maka identifikasi masalah dalam penelitian ini yaitu “Bagaimana hasil pengklasteran algoritma *K-means* dan *Fuzzy C-means*?” serta “Bagaimana perbandingan penerapan algoritma *K-means* dan *Fuzzy C-means* dalam pengklasteran pada data kinerja dosen Universitas Islam Bandung?”. Selanjutnya, tujuan dalam penelitian ini yaitu untuk mengetahui hasil pengklasteran pada algoritme *K-means* dan *Fuzzy C-means* serta mengetahui perbandingan algoritme *K-means* dan *Fuzzy C-means* dalam pengklasteran pada data kinerja dosen Universitas Islam Bandung.

2. Metodologi

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari hasil publikasi *website* bernama SINTA (*Science and Technology Index*). *Website* SINTA yang berisi hasil publikasi jurnal maupun kinerja peneliti dosen Universitas Islam Bandung sebagai data pengamatan. Data diambil dari seluruh dosen Universitas Islam Bandung dari berbagai Progam Studi yang berjumlah 512 dosen pada bulan Februari 2021. Hasil publikasi penelitian dosen SINTA dilihat dari hasil Penelitian dosen yang telah dipublikasikan (X_1), *Quartile Scopus* yaitu kelas kualitas jurnal Q1-Q4 versi Scopus (X_2), dan Sinta 1 hingga Sinta 6 yaitu kelas jurnal terakreditasi nasional (X_3) dan terdapat indikator – indikator di setiap variabelnya.

Pengklasteran K-Means

Secara umum dalam metode pengklasteran dapat dibagi menjadi dua yaitu cluster hirarki dan non-hirarki. Metode *K-Means* merupakan salah satu algoritme *clustering* non hirarki yang paling sederhana dan umum digunakan. *K-means* merupakan salah satu algoritme pengklasteran yang termasuk *hard clustering* dimana di setiap klasternya memiliki nilai *centroid* (titik pusat kluster). Proses pengklasteran dengan algoritme *K-Means* adalah sebagai berikut :

1. Tentukan banyaknya *cluster* yang diinginkan dari dataset.
2. Alokasikan objek sesuai dengan jumlah kluster yang ditentukan.
3. Tentukan nilai *centroid* pada tiap-tiap kluster dengan menetapkan secara acak k buah objek sebagai pusat awal kluster (*centroid*).
4. Hitung jarak terdekat data dengan *centroid* menggunakan rumus Euclidean.
5. Tampilkan hasil berdasarkan jarak terendah dan hasil perhitungan poin 4.
6. Jika belum didapatkan hasil yang sesuai, iterasi kembali dilanjutkan dengan menggunakan poin 3. Iterasi dihentikan jika hasil pengklasteran sudah sama dengan iterasi sebelumnya.

Untuk menentukan nilai *centroid* yaitu secara acak menempatkan sebanyak k buah objek sebagai *centroid* awal.

Untuk menentukan jarak digunakan rumus Euclidean sebagai berikut :

$$d_{ij} = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2} \quad \dots(2.1)$$

Keterangan :

$d(i, j)$: Jarak antara data ke i ke pusat *cluster* ke j

X_{ik} : Data ke i pada atribut ke k

X_{jk} : Titik pusat ke j , pada atribut ke k

Pengklasteran Fuzzy C-Means

Fuzzy C-Means merupakan salah satu metode pengklasteran dalam data mining, dimana keberadaan tiap-tiap titik data dalam suatu kluster ditentukan oleh derajat keanggotaannya. *Fuzzy C-means* merupakan salah satu jenis *fuzzy clustering* di mana kelebihannya dalam mengelompokkan suatu objek, yaitu jika setiap objek memiliki peluang untuk menjadi anggota lebih dari satu *cluster*. *Fuzzy C-Means* merupakan algoritme iteratif, dimana menerapkan iterasi pada setiap proses kluster.

Algoritme *Fuzzy C-Means* adalah sebagai berikut :

1. Tentukan terlebih dahulu jumlah *cluster* (c), pangkat pembobot ($m > 1$), maksimum iterasi ($MaxIter$), Eps (galat terkecil), fungsi objektif awal ($P_0 = 0$), dan iterasi awal ($t=1$).
2. Bangkitkan bilangan *random* μ_{ik} ; $i = 1, 2, \dots, n$; $k = 1, 2, \dots, c$; sebagai elemen-elemen matriks partisi awal U . Matriks partisi pada pengelompokkan *fuzzy* memenuhi kondisi sebagai berikut:
 $\mu_{ik} \in [0, 1]$; $1 \leq i \leq n$; $1 \leq k \leq c$;
 μ_{ik} merupakan derajat keanggotaan yang merujuk pada seberapa besar kemungkinan objek ke- i bisa menjadi anggota *cluster* ke- k .

3. Menghitung pusat *cluster* ke- k pada variabel ke- j ; v_{kj} dengan $k = 1, 2, \dots, c$; $j = 1, 2, \dots, p$:

$$v_{kj} = \frac{\sum_{i=1}^n (\mu_{ik})^m \cdot x_{ij}}{\sum_{i=1}^n (\mu_{ik})^m} \quad \dots(2.2)$$

4. Hitung fungsi objektif pada iterasi ke- t (P_t):

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^p (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^m \right) \quad \dots(2.3)$$

5. Hitung perubahan matriks keanggotaan μ_{ik} :

$$\mu_{ik} = \frac{\left[\sum_{j=1}^p (X_{ij} - V_{ij})^2 \right]^{-\frac{1}{m-1}}}{\sum_{k=1}^c \left[\sum_{j=1}^p (X_{ij} - V_{ij})^2 \right]^{-\frac{1}{m-1}}} \quad \dots(2.4)$$

6. Cek kondisi berhenti :

- Jika $(|P_t - P_{t-1}| < \text{Eps})$ atau $(t > \text{MaxIter})$ maka berhenti.
- Jika tidak, iterasi dinaikkan $t=t+1$, ulangi langkah ke 4.

Analisis Komponen Utama

Pengklasteran dengan algoritma *K-means* dan *Fuzzy C-Means* termasuk metode non-hirarki dimana dimulai dengan penentuan jumlah *cluster* atau nilai k terlebih dahulu. Untuk mengetahui apakah nilai k yang ditentukan telah optimum atau tidak digunakan metode Analisis Komponen Utama. Komponen utama adalah himpunan variabel baru yang merupakan kombinasi linier dari variabel-variabel yang amati. Untuk menentukan berapa banyak komponen utama adalah menggunakan *scree plot*, banyak komponen yang diambil adalah pada titik kurva tidak lagi menurun tajam atau mulai melandai, atau dengan melihat proporsi kumulatif varians terhadap total varians. Jika nilai proporsi kumulatif $>80\%$ maka komponen tersebut sudah mencukupi, setelah ditentukan berapa banyak komponen utama dilakukan plot diagram pencar untuk melihat ada berapa kelompok pada data pengamatan tersebut.

Silhouette Index

Silhouette Index (SI) atau *silhouette coefficient* digunakan untuk melihat kualitas dan kekuatan *cluster*, seberapa baik suatu objek ditempatkan dalam suatu *cluster*. Metode ini banyak digunakan untuk memvalidasi *cluster*. SI merupakan metode gabungan dari metode *cohesion* dan *separation*. Nilai SI yang didapat dalam rentang -1 hingga $+1$. Dimana nilai *Silhouette Index* semakin mendekati nilai $+1$ maka semakin bagus hasil pengelompokkan sebuah metode tersebut.

Tahapan perhitungan dari *Silhouette Index* adalah sebagai berikut :

- Hitung rata-rata jarak dari suatu data misalkan i dengan semua data lain yang berada dalam satu *cluster*.

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j) \quad \dots(2.5)$$

Keterangan :

A = Konstanta

j = data lain dalam suatu *cluster* A

$d(i, j)$ = adalah jarak i dengan j .

- Hitung rata-rata jarak dari data ke- i tersebut dengan semua data di *cluster* lain, dan diambil nilai terkecil.

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j) \quad \dots(2.6)$$

$$b(i) = \min C \neq A d(i, C) \quad \dots(2.7)$$

Keterangan :

$d(i, C)$ = jarak rata-rata data i dengan semua objek cluster lain C dimana $A \neq C$

3. Nilai SI dihitung dengan

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad \dots(2.8)$$

3. Pembahasan dan Diskusi

Sebelum melakukan analisis *cluster* menggunakan algoritma *K-Means* dan *Fuzzy C-Means* terhadap data kinerja dosen Universitas Islam Bandung. Dilakukan *pre-processing* atau *cleaning* data yang tidak dibutuhkan karena pada data kinerja dosen Universitas Islam Bandung terdapat beberapa dosen yang variabelnya semua bernilai 0, nilai 0 ini tidak akan berarti jika digunakan, maka data penelitian menjadi 333 dosen. Dengan bantuan software R, didapatkan hasil pengklasteran dengan algoritma *K-Means* dan *Fuzzy C-Means*. Berikut klaster yang dihasilkan:

Pengklasteran dengan Algoritme *K-Means*

Dengan menggunakan software R dengan jumlah *cluster* yaitu $k=4$, didapat hasil *cluster* dengan algoritma *K-Means* sebagai berikut:

Tabel 1. Hasil Pengklasteran Data Kinerja Dosen Algoritma *K-Means*

Anggota Cluster			
No	Nama Dosen	Dosen Jurusan	Cluster
1	Asep Ramdan Hidayat	Syariah	1
2	Erwin Hamdani	Matematika	2
3	Dedy Ansari	Manajemen	1
4	A Harits Nu'man	Teknik Industri	2
5	Hendrati Dwi N	Manajemen	3
⋮	⋮	⋮	
333	Agus N	Teknik Industri	1

Penclustering dengan algoritme *K-means* membentuk 4 *cluster*, yaitu untuk *cluster* 1 terdapat 273 dosen, *cluster* 2 terdapat 41 dosen, *cluster* 3 terdapat 16 dosen, dan *cluster* 4 terdapat 3 dosen.

Pengklasteran dengan Algoritme *Fuzzy C-Means*

Setelah ditentukan jumlah *cluster* dengan jumlah *cluster* yaitu $c=4$, pemberian nilai awal dengan besar pangkat pembobot dipilih $m = 2$, iterasi akan dilakukan diperhitungan ini adalah 100. Dengan bantuan software R, didapat hasil *cluster* dengan algoritma *Fuzzy C-Means* sebagai berikut:

Tabel 2. Hasil Pengklasteran Data Kinerja Dosen Algoritma *Fuzzy C-Means*

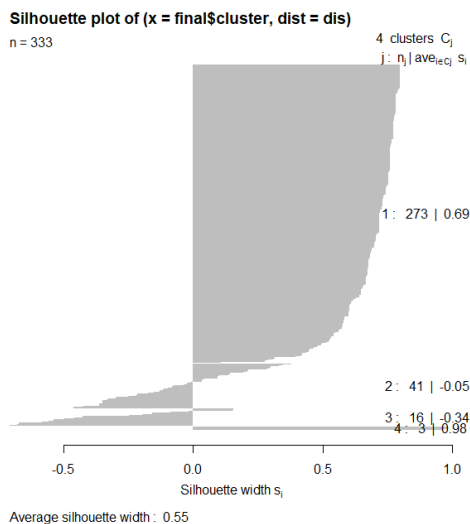
Anggota Cluster						
No	Nama Dosen	μ_{i1}	μ_{i2}	μ_{i3}	μ_{i4}	Cluster
1	Asep Ramdan Hidayat	0.2537	0.2406	0.2553	0.2502	3
2	Erwin Hamdani	0.2200	0.3196	0.2159	0.2444	2
3	Dedy Ansari	0.3283	0.1089	0.2977	0.2650	1
⋮	⋮					
333	Agus N	0.1699	0.0250	0.3854	0.4196	3

Penclustering dengan algoritme *Fuzzy C Means* membentuk 4 *cluster*, yaitu untuk *cluster* 1 terdapat 107 dosen, *cluster* 2 terdapat 33 dosen, *cluster* 3 terdapat 126 dosen, dan *cluster* 4 terdapat 67 dosen.

Uji Validitas clustering

Algoritma K-Means dengan Silhouette Index

Hasil perhitungan *Silhouette Index* (SI) untuk data kinerja dosen menggunakan metode K-Means dengan bantuan *software* R, parameter yang digunakan adalah *Euclidean*. Nilai SI yang didapat dalam rentang [-1,+1].

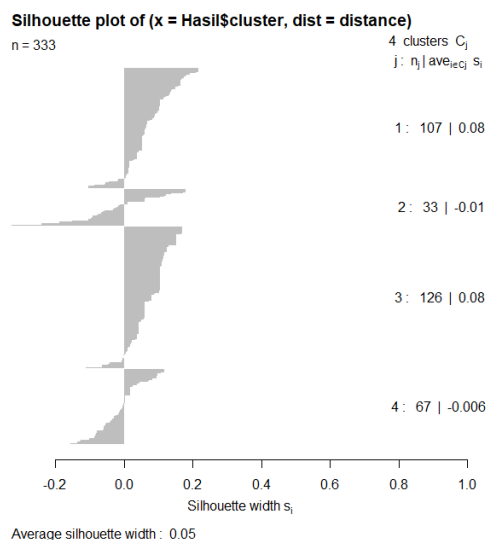


Gambar 1. Plot Nilai *Silhouette Index* pada *K-Means*

Berdasarkan plot nilai *Silhouette Index* (SI) pada *K-Means* diatas diketahui bahwa nilai rata-rata SI yaitu sebesar 0.55.

Algoritma Fuzzy C Means dengan Silhouette Index

Hasil Validasi *Fuzzy C-Means* dengan *Silhouette Index* (SI) untuk data kinerja dosen dengan bantuan *software* R. Nilai SI yang didapat dalam rentang [-1,+1].



Gambar 2. Plot Nilai Silhouette Index Pada *Fuzzy C- Means*

Berdasarkan plot nilai *Silhouette Index* (SI) pada *Fuzzy C-Means* diatas diketahui bahwa nilai rata-rata SI yaitu sebesar 0.05.

Interpretasi Uji Validitas

Berdasarkan hasil validasi menggunakan *Silhouette Index* (SI), dilihat dalam plot bahwa nilai

rata-rata SI pada *K-Means* adalah 0.55 dan nilai rata-rata SI pada *Fuzzy C-Means* adalah 0.05. Berdasarkan ketentuan validasi nilai yang lebih mendekati satu mempunyai arti bahwa kualitas *cluster* yang didapatkan semakin baik. Kesimpulannya bahwa algoritma *K-Means* lebih unggul dibandingkan algoritma *Fuzzy C-Means* dalam pengklasteran data kinerja dosen Universitas Islam Bandung. Dilihat dari hasil nilai SI *K-Means* tertinggi yaitu 0.55 maka tergolong dalam *medium structure*.

4. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut:

1. Pengklasteran menggunakan algoritma *K-Means* membentuk 4 *cluster*, yaitu untuk *cluster* 1 terdapat 273 dosen, *cluster* 2 terdapat 41 dosen, *cluster* 3 terdapat 16 dosen, dan *cluster* 4 terdapat 3 dosen. Sedangkan pengklasteran menggunakan algoritma *Fuzzy C-Means* membentuk 4 *cluster*, yaitu untuk *cluster* 1 terdapat 107 dosen, *cluster* 2 terdapat 33 dosen, *cluster* 3 terdapat 126 dosen, dan *cluster* 4 terdapat 67 dosen.
2. Berdasarkan hasil validasi *cluster* dari algoritma *K-Means* dan *Fuzzy C-Means* menunjukkan bahwa algoritma *K-Means* lebih unggul dibandingkan algoritma *Fuzzy C-Means* dalam pengklasteran data kinerja dosen Universitas Islam Bandung, hal ini dibuktikan hasil validasi menggunakan *Silhouette Index* (SI) *K-Means* 0.55 dan *Fuzzy C-Means* 0.05. Pengklasteran *K-Means* menunjukkan nilai SI tertinggi yaitu 0.55 maka tergolong dalam *medium structure*. Pada *cluster* 1 beranggotakan dosen yang belum memiliki jurnal terakreditasi Sinta, *cluster* 2 beranggotakan dosen yang memiliki minimal 5 jumlah artikel dikonferensi, dosen dengan jurnal kelas Q3, dan belum memiliki jurnal terakreditasi Sinta, *cluster* 3 beranggotakan dosen yang memiliki minimal 3 artikel, jurnal, buku yang dipublikasikan, dosen dengan jurnal kelas Q2 dan belum memiliki jurnal terakreditasi Sinta, dan *cluster* 4 beranggotakan dosen dengan jumlah artikel di jurnal kelas S6.

Acknowledge

Terimakasih banyak kepada seluruh pihak yang telah membantu dalam memberikan saran ataupun masukan sampai terlaksananya penelitian ini hingga selesai.

Daftar Pustaka

- [1] Cinaroglu, Songul. (2020). Comparison Of Fuzzy C-Means Dan K-Means Pengklasteran Performance: An Application On Household Budget Survey Data. Halaman: 54-62.
- [2] Kustini, & Lutfi, E. T. (2009). Algoritme Data Mining. Yogyakarta: Andi Offset. Sani, Asrul. (2018). Penerapan Metode K-Means Pengklasteran pada Perusahaan.
- [3] Aditya, A., Jovian, I., & Sari, B. N. (2020). Implementasi K-Means Pengklasteran Ujian Nasional Sekolah Menengah Pertama Di Indonesia Tahun 2018/2019. Volume 4, No. 1:51-58.
- [4] Simbolon, C. L., Kusumastuti, N., & Irawan Beni. (2013). Pengklasteran Lulusan Mahasiswa Matematika FMIPA UNTAN Pontianak Menggunakan Algoritme Fuzzy C-Means. Volume 01, No. 1:21-26.
- [5] Mirza, M. F. (2009). Metode Pengklasteran Dengan Algoritme Fuzzy C-Means Untuk Rekomendasi Pemilihan Bidang Keahlian Pada Program Studi Teknik Informatika.
- [6] Pradnyana, G. A., Permana, A. A. J. (2018). Sistem Pembagian Kelas Kuliah Mahasiswa Dengan Metode K-Means Dan K-Nearest Neighbors Untuk Meningkatkan Kualitas Pembelajaran. Volume 16, No. 1:59-68.
- [7] Handoyo, R., M, R., R., & Nasution, S., M. (2014). Perbandingan Metode Clustering Menggunakan Metode Single Linkage Dan K-Means Pada Pengelompokan Dokumen. Volume 15, No. 2.
- [8] Merliana, N., P., E., Ernawati, & Santoso, A., J. Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-Means Clustering.

- [9] Shihab, Ardhi. 2015/2016. Analisis Multivariat II dengan R. URL : Analisis Multivariat II dengan R - StuDocu. Diakses tanggal 11 Juni 2021.
- [10] Mujilahwati, Siti., Wardhani, Retno. (2021). Implementasi Fuzzy C-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Masuk Perguruan Tinggi. Volume 6, No. 1.
- [11] Wakhidah, Nur. Clustering Menggunakan K-Means Algorithm Sari, Kartika. (2020). Analisis Evaluasi Perhitungan Jarak Terhadap Nilai Silhoutte Coefficient pada Algotirma K-Means. (Universitas Sumatera Utama, 2020).
- [12] Yulianto Anggi Priliani, Darwis Sutawanir. (2021). *Penerapan Metode K-Nearest Neighbors (kNN) pada Bearing*. Jurnal Riset Statistika, 1(1), 10-18.