

Penerapan Metode K-Means untuk Klasifikasi Rekomendasi Tugas Akhir Mahasiswa

Dani Ramdani¹, Kurnia Naradinata², Firza Aditiya Ardiansah³, Muhamad Sahril⁴,
Ardiansyah Maulana⁵, Devi Yunita^{6*}

¹⁻⁶Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No. 46,
Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia

Email: ¹hi.daani1322@gmail.com, ²dennykukur@gmail.com, ³firzaaditiya22@gmail.com,
⁴mhmdsahril275@gmail.com, ⁵ardimaulana1232@gmail.com, ^{6*}dosen00846@unpam.ac.id

(* : coresponding author)

Abstrak–Tugas Akhir (TA) adalah komponen utama untuk menyelesaikan studi di perguruan tinggi. Namun, banyak mahasiswa mengalami kendala dalam memilih area penelitian TA yang sesuai dengan kemampuan. Penelitian ini bertujuan memberikan rekomendasi area penelitian TA dengan menggunakan data nilai A, B, dan C dari 10 Mata Kuliah Wajib (MKW) selama enam semester oleh mahasiswa Program Studi Informatika di Universitas Pamulang. Analisis data dilakukan menggunakan metode K-Means dengan tiga cluster: C1 (rendah), C2 (menengah), dan C3 (tinggi). Data yang digunakan adalah simulasi dari 5 mahasiswa dengan nilai A, B, dan C untuk setiap MKW. Hasil analisis menunjukkan bahwa pengelompokan dapat memberikan gambaran kesesuaian mahasiswa dengan topik TA. Evaluasi akurasi clustering menggunakan Sum of Squared Errors (SSE) menghasilkan nilai 1.875, mengindikasikan tingkat kesalahan rendah. Perhitungan Silhouette Coefficient (SC) menunjukkan nilai 0.1505, mencerminkan kualitas cluster cukup baik meskipun masih bisa ditingkatkan.

Kata Kunci: *Silhouette Coefficient*, SSE, Rekomendasi Tugas Akhir, *K-Means*

Abstract–*The Final Project (FP) is a key component to completing studies in higher education. However, many students face challenges in selecting research areas that match their abilities. This study aims to provide recommendations for FP research areas by utilizing grades of A, B, and C from 10 mandatory courses completed over six semesters by students of the Informatics Study Program at Universitas Pamulang. The data analysis was carried out using the K-Means method with three clusters: C1 (low), C2 (medium), and C3 (high). The data used is a simulation of 5 students with grades of A, B, and C for each mandatory course. The analysis results show that clustering can provide a clear depiction of students' suitability with their FP topics. The clustering accuracy evaluation using the Sum of Squared Errors (SSE) method resulted in a value of 1.875, indicating a low error rate. The Silhouette Coefficient (SC) calculation showed a value of 0.1505, reflecting fairly good cluster quality, although there is still room for improvement.*

Keywords: *Silhouette Coefficient, SSE, Recommendation, Final Project, K-Means*

1. PENDAHULUAN

Tugas Akhir (TA) merupakan salah satu komponen penting yang harus diselesaikan oleh mahasiswa sebagai syarat kelulusan di jenjang pendidikan tinggi. Namun, banyak mahasiswa menghadapi tantangan dalam menentukan topik atau bidang penelitian yang sesuai dengan kemampuan dan minat mereka. Kesulitan ini sering kali menyebabkan pemilihan topik TA yang kurang tepat atau tidak sesuai dengan pemahaman mahasiswa terhadap materi yang telah dipelajari selama perkuliahan. Oleh sebab itu, pemilihan topik TA yang menarik dan relevan dengan kompetensi yang telah diperoleh selama masa studi menjadi hal yang sangat penting bagi mahasiswa..

Pemilihan topik Tugas Akhir yang tepat berperan penting dalam menentukan keberhasilan serta kelancaran proses penyelesaiannya. Dalam konteks ini, nilai yang didapat oleh mahasiswa pada mata kuliah wajib selama masa studi dapat digunakan sebagai indikator dalam menentukan area penelitian yang sesuai. Nilai akademik tersebut mencerminkan tingkat pemahaman mahasiswa terhadap materi perkuliahan, sehingga dapat menjadi acuan dalam memilih topik penelitian yang relevan dan memungkinkan mereka menyelesaikannya dengan lebih baik.

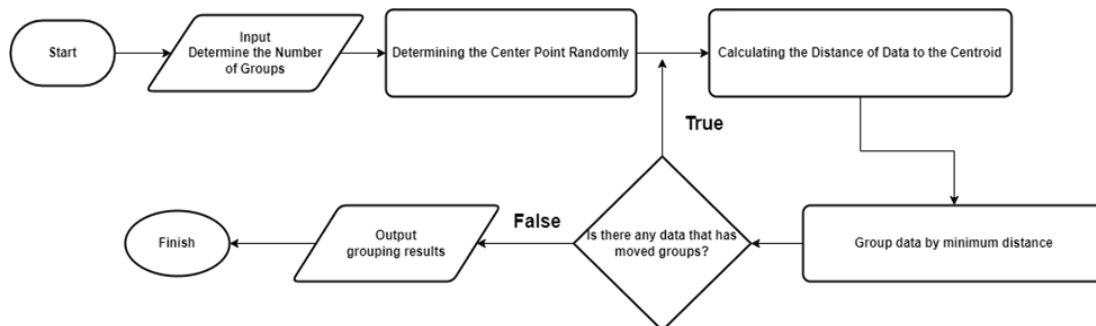
Namun, meskipun nilai akademik memiliki potensi untuk memberikan gambaran yang jelas tentang kemampuan mahasiswa, menentukan area penelitian yang tepat tetap menjadi tantangan. Maka dari itu, Penelitian ini dibuat bertujuan untuk memberikan rekomendasi area penelitian Tugas Akhir (TA) dengan memanfaatkan data nilai A, B, dan C dari 10 Mata Kuliah Wajib (MKW) yang telah ditempuh mahasiswa selama enam semester. Dengan menggunakan metode *K-Means clustering*, penelitian ini akan mengelompokkan mahasiswa berdasarkan nilai yang diperoleh pada MKW, sehingga dapat ditemukan pola-pola tertentu yang mengarah pada pemilihan topik TA yang sesuai.

Dalam penelitian ini, data yang digunakan bersifat simulasi untuk 5 mahasiswa, dengan setiap mahasiswa memiliki nilai A, B, dan C pada masing-masing mata kuliah. Hasil dari analisis clustering ini diharapkan dapat memberikan gambaran mengenai bagaimana mahasiswa dengan pola nilai tertentu dapat dipetakan ke dalam area penelitian yang relevan. Proses clustering dilakukan dengan membagi data ke dalam tiga cluster, yaitu cluster dengan nilai sedikit (C1), sedang (C2), dan banyak (C3). Pengujian akurasi clustering ini menggunakan Silhouette Coefficient (SC) dan Sum of Squared Errors (SSE) untuk memastikan kualitas dan kecocokan cluster yang terbentuk. Dengan demikian, penelitian ini dibuat untuk tujuan mengoptimalkan pemilihan topik tugas akhir bagi mahasiswa, sehingga diharapkan mereka dapat memilih topik yang sesuai tidak hanya dengan kemampuan mereka, tetapi juga bisa meningkatkan kualitas penelitian yang dilakukan.

2. TINJAUAN PUSTAKA

2.1 Algoritma K-Means

Algoritma *K-Means* merupakan metode clustering yang paling sering digunakan dalam data mining dan analisis statistik. Metode ini digunakan untuk mengelompokkan data ke dalam sejumlah cluster yang sudah ditentukan sebelumnya. Dalam setiap cluster, objek yang dikelompokkan memiliki kemiripan yang tinggi, sementara perbedaan antar cluster relatif besar. *K-Means* bekerja dengan cara iteratif, dimulai dengan pemilihan acak titik pusat *cluster (centroid)* dan kemudian mengelompokkan data berdasarkan kedekatannya dengan centroid tersebut. Setelah setiap data dikelompokkan, centroid baru dihitung berdasarkan average posisi data dalam cluster, dan proses ini diulang hingga penempatan data tidak berubah lagi. Penerapan *K-Means* dalam berbagai bidang telah terbukti efektif. Contohnya, dalam penelitian yang dilakukan oleh Asroni (2015), *K-Means* diterapkan untuk mengelompokkan mahasiswa berdasarkan nilai akademik yang mereka dapatkan, mahasiswa dengan IPK tinggi dikelompokkan ke dalam *cluster* yang dapat mewakili potensi mereka untuk mengikuti lomba atau program beasiswa. Penelitian ini menunjukkan bahwa *K-Means* dapat diterapkan untuk berbagai tujuan pengelompokan yang berkaitan dengan data akademik. Selain itu, Purnawansyah & Haviluddin (2017) juga menggunakan *K-Means* dalam analisis trafik jaringan untuk mengelompokkan data dan menemukan pola dalam penggunaan jaringan. Penelitian-penelitian ini menunjukkan bahwa *K-Means* tidak hanya efektif untuk data akademik, tetapi juga untuk data yang lebih kompleks dan bervariasi, termasuk data yang berkaitan dengan perilaku dan kinerja.



Gambar 1. Diagram Alur *K-Means Clustering*

Gambar pertama menunjukkan proses analisis dalam penerapan algoritma K-Means untuk rekomendasi Tugas Akhir. Langkah pertama adalah menentukan jumlah cluster nilai A, B, dan C. Pada penelitian ini, digunakan tiga cluster, C1 mewakili kategori sedikit, C2 untuk kategori sedang, dan C3 untuk kategori banyak. Selanjutnya, centroid awal ditentukan berdasarkan nilai minimum, rata-rata, dan maksimum dari data yang dianalisis. Setelah itu, jarak antara setiap data dengan centroid dihitung menggunakan metode *Euclidean Distance* (ED), seperti yang dirumuskan pada Persamaan 1.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Yang dimana i menunjukkan sebuah objek dan x, y merupakan koordinat dari objek tersebut.

Keempat, mengelompokkan data ke centroid terdekat. Akhirnya, lakukan perhitungan ulang nilai centroid hingga data tidak mengalami perubahan lagi.

2.2 Perhitungan Akurasi

Pengujian model dilakukan untuk menilai sejauh mana hubungan antar objek dalam satu cluster dan sejauh mana pemisahan antar cluster. Dalam penelitian ini, pengujian cluster menggunakan metode Silhouette Coefficient (SC) dan Sum of Squared Errors (SSE). dan SSE digunakan untuk menentukan jumlah cluster yang optimal, Sementara itu, SC berfungsi sebagai metode evaluasi untuk mengukur kekuatan dan ketepatan hasil clustering (Anggara, Sujiani, & Helfi, 2016).

1. SSE atau Sum of Squared Errors

Penelitian menggunakan Metode Sum of Squared Errors untuk mengukur Tingkat ketepatan dengan menghitung selisih antara total dan nilai yang sebenarnya. Nilai Sum of Squared yang mendekati 0 (Nol) menunjukkan jika model memiliki kesalahan acak yang minimal dan dapat digunakan untuk peramalan model yang diamati. Perhitungan SSE menggunakan Persamaan 2.

$$SSE = \sum_{i=1}^n (X_i - Y_i)^2$$

Yang dimana, Y merupakan nilai yang mendekati tercapai, sedangkan X merupakan nilai yang sebenarnya.

2. SC atau Silhouette Coefficient

Metode SC dipakai pada penelitian ini untuk mengevaluasi kualitas clustering dengan 3 atau 4 cluster. Nilai Silhouette Coefficient yang optimal dipakai untuk mengukur sejauh mana kualitas cluster yang dihasilkan (Sugiyama, 2015; Kaufman, Gentle, & Rousseuw, 1991). Menghitung SC menggunakan Persamaan 3.

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Dimana, $S(i)$ adalah nilai SC untuk objek i , $b(i)$ adalah nilai minimum rata-rata jarak ke objek lain, dan $a(i)$ adalah perbedaan average jarak objek i ke semua objek lainnya dalam cluster.

2.3 Pengambilan Sampel Data

Pada penelitian dilaksanakan, data simulasi digunakan untuk menggambarkan nilai mahasiswa pada sepuluh Mata Kuliah Wajib (MKW) yang relevan untuk mendukung penentuan Tugas Akhir (TA). Data ini terdiri dari nilai A, B, dan C dihasilkan untuk 5 sampel mahasiswa. Setiap mahasiswa memiliki nilai pada 10 MKW, yang merupakan mata kuliah (MK) yang dianggap penting dalam menentukan area

penelitian TA. Normalisasi dilakukan menggunakan formula pada Persamaan 4. Proses ini penting untuk memastikan bahwa perbedaan rentang nilai antar atribut tidak mempengaruhi hasil clustering. Selanjutnya, data simulasi nilai dapat dilihat pada Tabel Pertama.

$$X' = \frac{0,8 (X - b)}{(a - b)} + 0,1$$

Yang dimana, X' merujuk pada data yang baru dan X adalah data yang lama dan b menunjukkan nilai terkecil sedangkan a menunjukkan nilai terbesar.

Tabel 1. Nilai A, B, dan C pada 10 Mata Kuliah Wajib (MKW)

NO.	Nama Mata Kuliah	Mahasiswa 1	Mahasiswa 2	Mahasiswa 3	Mahasiswa 4	Mahasiswa 5
1	Basis Data	A (4)	B(3)	C(2)	B(3)	A(4)
2	Sistem Operasi	B (3)	B(3)	C(2)	A(4)	B(3)
3	Struktur Data	C (2)	A(4)	B(3)	A(4)	C(2)
4	Pemrograman Visual	B (3)	C (2)	A (4)	B (3)	C (2)
5	Algoritma dan Pemrograman	A (4)	C (2)	B (3)	A (4)	B (3)
6	Pemrograman Berbasis Objek	B (3)	A (4)	C (2)	B (3)	A (4)
7	Pemrograman Web	C (2)	B (3)	A (4)	B (3)	A (4)
8	Kecerdasan Buatan	A (4)	B (3)	C (2)	A (4)	A (4)
9	Sistem Keamanan Buatan	B (3)	C (2)	A (4)	B (3)	C (2)
10	Sistem Jaringan	C (2)	B (3)	A (4)	C (2)	B (3)

3. ANALISA DAN PEMBAHASAN

Pada percobaan kali ini, metode K-Means diimplementasikan untuk mengelompokkan data nilai lima mahasiswa pada sepuluh Mata Kuliah Wajib (MKW) berdasarkan nilai A, B, dan C. Analisa dibuat untuk mengelompokkan mahasiswa sesuai pola nilai mereka guna memberikan rekomendasi area penelitian Tugas Akhir (TA). Tiga cluster digunakan dalam analisis ini, dengan kriteria C1 (rendah) kelompok mahasiswa yang memiliki nilai rendah, C2 (menengah) kelompok mahasiswa dengan nilai sedang, dan C3 (tinggi) kelompok mahasiswa dengan nilai tinggi.. Nilai centroid awal untuk masing-masing cluster dihitung berdasarkan kategori performa: rata-rata nilai mahasiswa berperforma rendah untuk C1, nilai mahasiswa berperforma menengah untuk C2, dan rata-rata nilai mahasiswa berperforma

tinggi untuk C3. Nilai centroid hasil analisis dapat dilihat pada Tabel 2 yang akan memberi Gambaran data nya.

Tabel 2. Nilai Centroid Awal untuk Masing-Masing Cluster

No	Mata Kuliah	C1	C2	C3
1	Basis Data	0.25	0.50	0.75
2	Sistem Operasi	0.20	0.50	0.83
3	Struktur Data	0.20	0.67	0.83
4	Pemrograman Visual	0.40	0.33	0.50
5	Algoritma & Pemrograman	0.40	0.33	0.75
6	Pemrograman Berbasis Objek	0.30	0.67	0.67
7	Pemrograman Web	0.60	0.50	0.33
8	Kecerdasan Buatan	0.30	0.50	0.83
9	Sistem Keamanan Komputer	0.60	0.33	0.50
10	Sistem Jaringan	0.50	0.67	0.50

Tabel 3. Hasil Pengelompokkan pada Iterasi Pertama

No	Mata Kuliah	Cluster
1	Basis data	C3
2	Sistem Operasi	C3
3	Struktur Data	C2
4	Pemrograman Visual	C1
5	Algoritma & Pemrograman	C1
6	Pemrograman Berbasis Objek	C2
7	Pemrograman Web	C1
8	Kecerdasan Buatan	C3
9	Sistem Keamanan Komputer	C1
10	Sistem Jaringan	C2

Tabel 4. Nilai Centroid yang Diperbarui pada Iterasi Kedua

Cluster	Centroid Baru (x1)	Centroid Baru (x2)	Centroid Baru (x3)
C1	0.50	0.3725	0.52
C2	0.33	0.67	0.67
C3	0.25	0.50	0.80

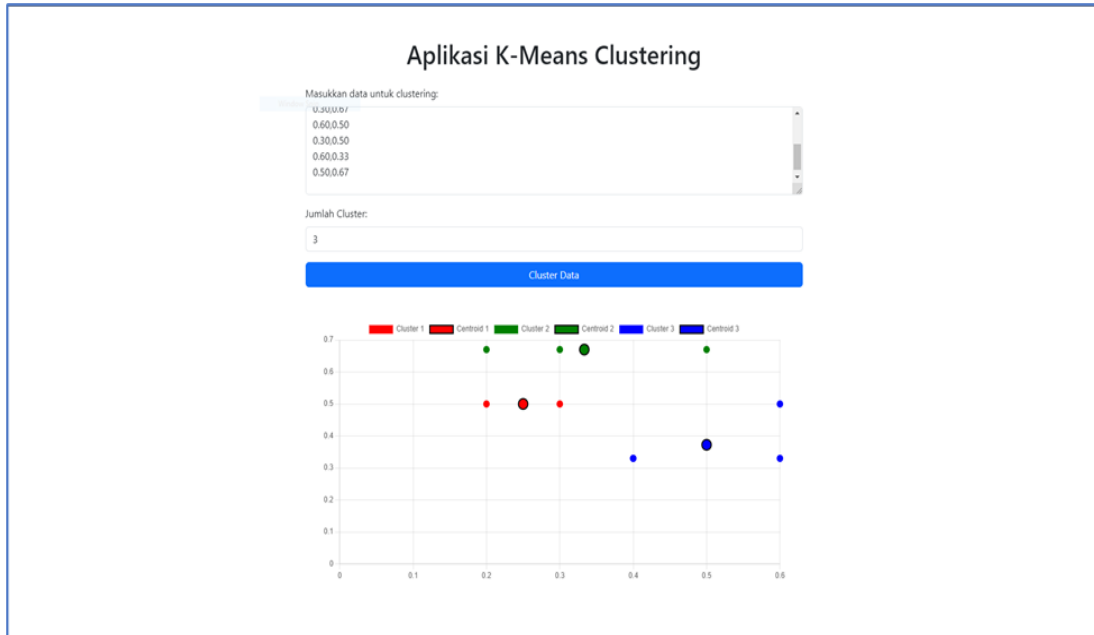
Tabel 5. Hasil Pengelompokan pada Iterasi Terakhir

No	Mata Kuliah	Cluster
1	Basis Data	C3
2	Sistem Operasi	C3
3	Struktur Data	C3
4	Pemrograman Visual	C1
5	Algoritma & Pemrograman	C3
6	Pemrograman Berbasis Objek	C2
7	Pemrograman Web	C1
8	Kecerdasan Buatan	C3
9	Sistem Keamanan Komputer	C1
10	Sistem Jaringan	C2

Tabel 6. Hasil Pengujian Menggunakan *Silhouette Coefficient* Mengindikasikan

Data	3 Cluster (ai)	3 Cluster (bi)	3 Cluster (Si)	4 Cluster (ai)	4 Cluster (bi)	4 Cluster (Si)
1	0.098	0.210	0.532	0.071	0.115	0.096
2	0.138	0.187	0.260	0.189	0.178	0.166
3	0.115	0.383	0.412	0.616	0.182	0.220
4	0.118	0.161	0.267	0.073	0.179	0.132
5	0.143	0.299	0.520	0.213	0.518	0.520
6	0.105	0.221	0.525	0.074	0.183	0.515
7	0.047	0.204	0.768	0.066	0.139	0.520
8	0.046	0.222	0.789	0.083	0.187	0.555
9	0.047	0.196	0.757	0.073	0.121	0.397
10	0.631	0.081	0.386	0.132	0.386	0.366
Global SC	0.585	0.459				

Penelitian ini mencakup aplikasi sederhana berbasis web menggunakan *JavaScript* dan *Chart.js* buat menampilkan hasil algoritma K-Means. Pada gambar ketiga menunjukkan cluster dari 10 data Mata Kuliah Wajib (MKW) dalam grafik *scatter plot*, dengan simbol 'o' sebagai data dan 'x' sebagai *centroid*. Visualisasi ini menggambarkan pembagian cluster secara jelas.



Gambar 2. Tampilan Aplikasi

4. KESIMPULAN

Pengaplikasian algoritma K-Means dalam menganalisis keterhubungan nilai yang didapat pada Mata Kuliah Wajib (MKW) mahasiswa telah dilakukan dengan sukses. Hasil percobaan menunjukkan bahwa dengan menggunakan 3 *cluster*, data MKW dapat dikelompokkan dengan baik. *Cluster* pertama (C1) berhubungan dengan MKW yang lebih sedikit, sementara cluster kedua (C2) berisi MKW dengan jumlah lebih banyak, dan *cluster* ketiga (C3) mengelompokkan MKW yang memiliki hubungan lebih kuat. Pengukuran hasil *cluster* dilakukan menggunakan metode *Sum of Squared Errors* (SSE) dan evaluasi kualitas *cluster* menggunakan metode *Silhouette Coefficient* (SC). Untuk 3 *cluster*, nilai SSE yang diperoleh adalah 0.4506%, sementara untuk 4 *cluster* nilainya meningkat menjadi 1.1072%. Ini menunjukkan bahwa model dengan 3 *cluster* memiliki error yang lebih kecil dan lebih baik dalam membagi data. Evaluasi dengan SC menunjukkan hasil yang lebih baik pada 3 *cluster*, dengan rata-rata SC sebesar 0.5852, sedangkan untuk 4 *cluster*, rata-rata SC hanya mencapai 0.4591. Nilai SC yang lebih tinggi pada 3 *cluster* menunjukkan bahwa pembagian data pada 3 *cluster* lebih baik dalam mengelompokkan objek, sementara nilai SC yang lebih rendah pada 4 *cluster* mengindikasikan bahwa pengelompokan dengan 4 *cluster* kurang optimal dalam menempatkan objek pada *cluster* yang sesuai. Hasil penelitian ini mengungkapkan bahwa algoritma K-Means efektif digunakan untuk menganalisis hubungan antara nilai Mata Kuliah Wajib (MKW) dan topik Tugas Akhir (TA). Diharapkan, temuan ini dapat mendukung mahasiswa, dosen, serta program studi dalam menentukan arah penelitian TA dengan lebih tepat. Penelitian di masa mendatang dapat difokuskan pada perbandingan serta optimalisasi algoritma K-Means guna meningkatkan akurasi dalam proses clustering.

REFERENCES

Adrianto, R. (n.d.). Penerapan algoritma K-Means dalam metode clustering untuk memberikan rekomendasi pemilihan jalur peminatan sesuai kemampuan pada Program Studi Teknik Informatika-S1 Universitas Dian Nuswantoro. *Journal JOINS Udinus*, 101–116.

- Asroni, R. A. (n.d.). Penerapan K-Means untuk clustering mahasiswa berdasarkan nilai akademik dengan Weka Interface: Studi kasus pada Jurusan Teknik Informatika UMM Magelang. *Ilmiah Semesta Teknika*, 18(1), 76–82.
- Farokhah, L. &. (n.d.). Penerapan K-Means untuk rekomendasi topik tugas akhir di STMIK Asia Malang. *Journal of Information Technology and Management*, 3(2), 142–148.
- Kusuma, T. (n.d.). Pengelompokan jenis kupu-kupu menggunakan algoritma K-Means dan fitur ekstraksi GLCM. *Information Technology Journal*, 1–4.
- Muzakir, A. (n.d.). Analisa dan Pemanfaatan K-Means clustering untuk menganalisis data nilai siswa dalam menentukan penerima beasiswa. *Proceedings of the National Seminar on Science & Technology Applications (SNAST)*, 211–216.
- Purnawansyah, &. H. (n.d.). Implementasi K-Means clustering dalam analisis aktivitas lalu lintas jaringan. *Proceedings of CYBERNETICSCOM 2016: International Conference on Computational Intelligence and Cybernetics*.
- Puspitasari, N. &. (n.d.). (2016). Penggunaan algoritma K-Means untuk pengelompokan data curah hujan di wilayah Kalimantan Timur. *Proceedings of SNRIK 2016*.
- Sharma, N. (. (n.d.). Pemahaman teknik clustering dalam data mining. *International Journal of Emerging Technology and Advanced Engineering*, 2(8), 313–317.