

TESIS

**STUDI KOMPARATIF ALGORITMA K-MEANS
CLUSTERING DAN FUZZY C-MEANS CLUSTERING UNTUK
PENGELOMPOKAN DATA ARTIKEL ILMIAH**



Oleh:

Maryama Kurnia Amri

NIM: 22206051001

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA

PROGRAM STUDI INFORMATIKA
PROGRAM MAGISTER FAKULTAS SAINS DAN TEKNOLOGI
UIN SUNAN KALIJAGA

YOGYAKARTA

2025

PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Maryama Kurnia Amri

NIM : 22206051001

Jenjang : Magister

Program Studi : Informatika

menyatakan bahwa naskah tesis ini secara keseluruhan adalah hasil penelitian/karya saya sendiri, kecuali pada bagian-bagian yang dirujuk sumbernya.

Yogyakarta, 13 Januari 2025

Saya yang menyatakan,



Maryama Kurnia Amri

NIM: 22206051001

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

PERNYATAAN BEBAS PLAGIASI

Yang bertanda tangan di bawah ini:

Nama : Maryama Kurnia Amri

NIM : 22206051001

Jenjang : Magister

Program Studi : Informatika

menyatakan bahwa naskah tesis ini secara keseluruhan benar-benar bebas dari plagiasi. Jika di kemudian hari terbukti melakukan plagiasi, maka saya siap ditindak sesuai ketentuan hukum yang berlaku.

Yogyakarta, 13 Januari 2025

Saya yang menyatakan,



Maryama Kurnia Amri

NIM: 22206051001

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-138/Un.02/DST/PP.00.9/01/2025

Tugas Akhir dengan judul : STUDI KOMPARATIF ALGORITMA K-MEANS CLUSTERING DAN FUZZY C-MEANS CLUSTERING UNTUK PENGELOMPOKAN DATA ARTIKEL ILMIAH

yang dipersiapkan dan disusun oleh:

Nama : MARYAMA KURNIA AMRI, S. Kom.
Nomor Induk Mahasiswa : 22206051001
Telah diujikan pada : Rabu, 15 Januari 2025
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR

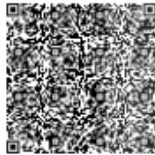


Valid ID: 678eebdc9fdaf

Ketua Sidang

Dr. Agung Fatwanto, S.Si., M.Kom.

SIGNED



Valid ID: 678eeaf6d7189c

Penguji I

Prof. Dr. Ir. Shofwatul Uyun, S.T., M.Kom.,
IPM., ASEAN Eng.

SIGNED

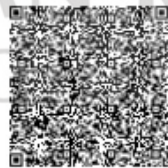


Valid ID: 678eeec756e79f

Penguji II

Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D.

SIGNED



Valid ID: 678f24c14ce4a

Yogyakarta, 15 Januari 2025

UIN Sunan Kalijaga

Dekan Fakultas Sains dan Teknologi

Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.

SIGNED



**KEMENTERIAN AGAMA REPUBLIK INDONESIA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
MAGISTER INFORMATIKA**

Jl. Marsda Adisucipto Tel. (0274) 558254. Fax. (0274) 586117. Yogyakarta 55281.

SURAT PERSETUJUAN TUGAS AKHIR

Hal : Persetujuan Tugas Akhir

Kepada Yth.,
Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga Yogyakarta

Assalamu 'alaikum wr. wb.

Setelah melakukan bimbingan, arahan, dan koreksi terhadap penulisan tesis yang berjudul: “Studi Komparatif Algoritma K-Means Clustering dan Fuzzy C-Means Clustering untuk Pengelompokan Data Artikel Ilmiah”, Yang ditulis oleh:

Nama : Maryama Kurnia Amri
NIM : 22206051001
Jenjang : Magister
Program Studi : Informatika

Sudah dapat diajukan kepada Program Studi Magister Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga sebagai salah satu syarat untuk memperoleh gelar Magister Informatika.

Dengan ini saya mengharap agar tugas akhir tersebut diatas dapat segera dimunaqosahkan. atas perhatiannya saya ucapkan terimakasih.

Wassalamu'alaikum wr. wb.

Yogyakarta, 13 Januari 2025

Pembimbing,

Dr. Agung Fatwanto, S.Si., M.Kom.

NIP: 19770103 200501 1 003

NOTA DINAS PEMBIMBING

Kepada Yth.,
Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga Yogyakarta

Assalamu 'alaikum wr. wb.

Setelah melakukan bimbingan, arahan, dan koreksi terhadap penulisan tesis yang berjudul:

Studi Komparatif Algoritma K-Means Clustering dan Fuzzy C-Means Clustering untuk Pengelompokan Data Artikel Ilmiah

Yang ditulis oleh:

Nama : Maryama Kurnia Amri
NIM : 22206051001
Jenjang : Magister
Program Studi : Informatika

Saya berpendapat bahwa tesis tersebut sudah dapat diajukan kepada Magister Informatika UIN Sunan Kalijaga untuk diujikan dalam rangka memperoleh gelar Magister Informatika.

Wassalamu'alaikum wr. wb.

Yogyakarta, 13 Januari 2025

Pembimbing



Dr. Agung Fatwanto, S.Si., M.Kom.

NIP: 19770103 200501 1 003

ABSTRAK

Belakangan ini, jumlah publikasi artikel ilmiah dapat mencapai angka jutaan setiap tahunnya. Peningkatan yang cukup massif ini sangat menyulitkan bagi para peneliti, terutama saat akan mencoba memahami tren *cluster* perkembangan studi pada bidang tertentu. Untuk itu diperlukan alat yang bisa membantu proses clustering artikel ilmiah secara otomatis. Penelitian ini bermaksud mengkaji algoritma yang berpotensi cocok untuk diterapkan pada proses *clustering* data artikel ilmiah. Secara khusus penelitian ini bermaksud membandingkan kinerja antara dua algoritma *clustering* yaitu K-Means dengan Fuzzy C-Means (FCM), untuk pengelompokan data artikel ilmiah. Penelitian ini dilakukan dengan menerapkan metode eksperimen. Kinerja kedua algoritma dibandingkan menggunakan data judul dan abstrak dari artikel ilmiah yang diambil dari berbagai *digital library* yaitu IEEE, ACM dan Springer Link sejumlah 7136 data. Berdasarkan hasil pengukuran skor *Silhouette Coefficient*, Pada atribut judul dari 4 kata kunci yang diuji, K-Means lebih unggul pada dua kata kunci, sedangkan pada dua kata kunci lainnya K-Means memiliki skor yang sama dengan FCM. Pada atribut abstrak FCM lebih unggul pada tiga kata kunci. Selanjutnya berdasarkan hasil uji Wilcoxon terhadap data survei responden, menunjukkan tidak ada perbedaan signifikan atas hasil *clustering* kedua algoritma. Hasil penelitian menunjukkan bahwa K-Means lebih unggul dalam *clustering* atribut judul, sementara FCM lebih baik dalam clustering atribut abstrak berdasarkan hasil *silhouette score*. Namun berdasarkan opini responden, tidak ada perbedaan hasil *clustering* antara K-Means dan FCM.

Kata Kunci : *Clustering*, K-Means, Fuzzy C-Means (FCM), *Silhouette Score*, Uji Wilcoxon.

ABSTRACT

In recent times, the annual number of published scientific articles has risen to millions. This significant increase poses challenges for researchers, particularly when attempting to understand clustering trends in the development of studies within specific fields. Therefore, tools capable of automating the clustering process for scientific articles are essential. This study aims to evaluate algorithms that are potentially suitable for clustering scientific article data. Specifically, this study aims to compare K-Means and Fuzzy C-Means (FCM) algorithm's performance for clustering scientific article data. An experimental method was employed, using a dataset of 7,136 articles comprising titles and abstracts obtained from various digital libraries, including IEEE, ACM, and Springer Link. Based on the Silhouette Coefficient scores, for title attributes across four tested keywords, K-Means outperformed FCM on two keywords, while the remaining two keywords showed equal scores for both algorithms. For abstract attributes, FCM performed better on three out of four keywords. Furthermore, the Wilcoxon test on survey responses revealed no significant differences in the clustering results between the two algorithms. The findings indicate that K-Means excels in clustering title attributes, while FCM is more effective for abstract attributes based on Silhouette scores. However, according to respondent opinions, the clustering results between K-Means and Fuzzy C-Means (FCM) exhibit no significant differences.

Keywords : *Clustering, K-Means, Fuzzy C-Means, Silhouette Coefficient Score, Wilcoxon Test.*



MOTTO

Jalan Perlahan, tapi Tetap ke Depan.



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

PERSEMBAHAN

Tugas Akhir ini dipersembahkan untuk

Almamater, Universitas Islam Negeri Sunan Kalijaga
Yogyakarta dan rekan-rekan Magister Informatika tahun Angkatan
2022.



Kata Pengantar

Dengan segala kerendahan hati, penulis mengucapkan syukur ke hadirat Allah SWT atas limpahan rahmat dan karunia-Nya, sehingga penelitian yang berjudul “Studi Komparatif Algoritma K-Mean Clustering dan Fuzzy C-Mean Clustering untuk Pengelompokan Data Artikel Ilmiah” dapat diselesaikan. Tesis ini diajukan sebagai salah satu persyaratan untuk meraih gelar Magister di UIN Sunan Kalijaga Yogyakarta. Penulis memahami bahwa proses penyusunan tesis ini tidak terlepas dari dukungan dan bantuan berbagai pihak, baik secara langsung maupun tidak langsung. Oleh karena itu, penulis ingin menyampaikan penghargaan dan terima kasih yang mendalam kepada semua pihak yang telah berkontribusi dalam menyelesaikan tesis ini. Rasa terima kasih secara khusus juga ditujukan kepada:

1. Rektor UIN Sunan Kalijaga Yogyakarta, Bapak Prof. Noorhaidi, S.Ag., M.A., M.Phil., Ph.D.
2. Dekan Fakultas Sains dan Teknologi, Ibu Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.
3. Ketua Program Studi Magister Informatika, Bapak Dr. Ir. Sumarsono, ST, M.Kom., Yang tiada hentinya mengingatkan dan memberikan kemudahan bagi penulis untuk menyelesaikan tugas akhir ini.
4. Pembimbing tesis, Bapak Dr. Agung Fatwanto, S.Si., M.Kom., Yang senantiasa meluangkan waktu untuk memberikan arahan dan bimbingan dalam penulisan tesis ini, sehingga penelitian ini dapat selesai dengan baik.
5. Bapak dan ibu dosen yang mengampu matakuliah, selama penulis menimba ilmu di Magister Informatika UIN Sunan Kalijaga.

6. Kedua orang tua tercinta yang selalu menanamkan pemahaman pentingnya ilmu pengetahuan, sehingga penulis selalu mendapatkan dukungan untuk menempuh Pendidikan sampai ke jenjang magister ini.
7. Untuk suami tercinta Syahrudin Simanjuntak, yang memberikan izin kepada penulis untuk melanjutkan studi dan selalu memberikan dukungan penuh, sehingga penulis dapat menyelesaikan studi magister ini.
8. Untuk keempat saudara saya Wahyu Amri, Ruslana Amri, M. Muizzuddin Amri dan Muzillul Kabir Amri, yang selalu mendukung dan menguatkan penulis dalam kondisi apapun setelah orang tua kita meninggal dunia.
9. Rekan-rekan seperjuangan Magister Informatika terutama Angkatan 2022, yang walau hanya dalam waktu singkat kita berbagi suka duka dalam tugas kuliah dan berbagi ilmu selama menempuh studi.

Tesis ini masih memiliki berbagai kekurangan dan belum mencapai kesempurnaan, sehingga masukan serta saran konstruktif sangat diharapkan dari berbagai pihak. Penelitian ini juga diharapkan dapat memberikan dampak positif dan manfaat bagi para pembacanya.

Yogyakarta, 13 Januari 2025

Peneliti

Maryama Kurnia Amri, S.Kom.

22206051001

DAFTAR ISI

Pernyataan Keaslian	ii
Surat Pernyataan Bebas Plagiasi	iii
Surat Pengesahan Tim Penguji	iv
Surat Pengesahan Tugas Akhir	v
Nota Dinas Pembimbing	vi
Abstrak	vii
Motto.....	ix
Persembahan.....	x
Kata Pengantar	xi
Daftar Isi.....	xiii
Daftar Tabel	xvii
Daftar Gambar	xviii
Bab I Pendahuluan	1
A. Latar Belakang	1
B. Rumusan Masalah	3
C. Batasan Masalah.....	3

D. Tujuan Penelitian	4
E. Manfaat Penelitian.....	4
BAB II Tinjauan Pustaka dan Landasan Teori.....	5
A. Tinjauan Pustaka.....	5
B. Landasan Teori.....	9
1. Clustering	9
a. K-Means	9
b. Fuzzy C-Means (FCM).....	13
2. Silhouette Coefficient Score	15
3. Analisa Statistik	16
I. Statistik Deskriptif.....	17
a. Mean	17
b. Median	17
c. Modus	18
d. Standar Deviasi	18
e. Persentil	19
f. Margin of Error	20

II. Statistik Inferensial	20
a. Uji Validitas	21
b. Uji Reliabilitas	21
c. Uji Normalitas	22
d. Uji Wilcoxon	23
BAB III Metode Penelitian	24
A. Kajian Literatur.....	24
B. Tahapan Penelitian	24
1. Pengumpulan Data.....	25
2. Clustering.....	26
3. Evaluasi.....	27
a) Silhouette Coefficient Score	27
b) Survei	27
C. Alat Penelitian.....	39
BAB IV Hasil dan Pembahasan	31
A. Clustering	31
1. K-Means	35

2. Fuzzy C-Means.....	37
B. Evaluasi	40
a. Silhouette Coefficient Score.....	40
b. Survei.....	42
1. Uji Validitas	44
2. Uji Reliabilitas.....	45
3. Uji Normalitas	45
4. Uji Wilcoxon.....	46
BAB V Kesimpulan dan Saran.....	50
A. Kesimpulan	50
B. Saran	50
Daftar Pustaka	51

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
 YOGYAKARTA

DAFTAR TABEL

Tabel 1. Penelitian Terdahulu.....	7
Tabel 2. Bobot Skala Likert.....	28
Tabel 3. Instrumen Pertanyaan	28
Tabel 4. Silhouette Score.....	40
Tabel 5. Karakteristik Responden	42
Tabel 6. Presentase opini responden terhadap algoritma clustering....	43
Tabel 7. Uji Validitas	44
Tabel 8. Uji Reliabilitas.....	45
Tabel 9. Uji Normalitas	46
Tabel 10. Perhitungan <i>central tendency</i>	47
Tabel 11. Uji Wilcoxon	48

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR GAMBAR

Gambar 1. Tahapan Penelitian	25
Gambar 2. Script program untuk scraping situs ACM	31
Gambar 3. Hasil scraping data.....	31
Gambar 4. Hasil Proses <i>Clustering</i>	39



BAB I

PENDAHULUAN

A. LATAR BELAKANG

Belakangan ini, artikel ilmiah dalam berbagai bidang semakin banyak dipublikasikan setiap tahunnya. Dengan berbagai alasan berbagai bidang profesi melakukan penelitian ilmiah dan membuat artikel ilmiah, salah satunya adalah penelitian yang dilakukan mahasiswa untuk sekedar tugas kuliah atau bahkan tugas akhir yang merupakan syarat kelulusan dari universitasnya. Hal ini membuat semakin banyaknya artikel ilmiah yang dipublikasikan *digital library*. Setiap tahunnya, jutaan artikel ilmiah diterbitkan. Peningkatan jumlah artikel ilmiah yang *massif* ini membuat peneliti kesulitan untuk memahami tren perkembangan studi bidang tertentu. Fakta ini juga menarik perhatian sejumlah peneliti untuk menjadikan artikel ilmiah sebagai objek penelitian. Beragam metode digunakan dalam penelitian tersebut, termasuk melakukan klasterisasi artikel ilmiah berdasarkan kategori tertentu dan algoritma yang spesifik.

Klasterisasi (*clustering*) adalah salah satu fungsi dalam *data mining* untuk mengidentifikasi kelompok-kelompok data berdasarkan karakteristik tertentu yang dimiliki oleh setiap kelompok. Metode ini bertujuan untuk menciptakan pengelompokan di mana variasi antar objek dalam satu cluster relatif kecil, sementara perbedaan antara cluster satu dengan lainnya lebih signifikan (Wicaksono, 2016).

Metode klasterisasi memiliki beberapa algoritma, penelitian ini akan difokuskan pada dua algoritma, yaitu K-Means *clustering* dan Fuzzy C-Means (FCM) *clustering*. K-Means adalah salah satu algoritma untuk klasterisasi paling sederhana dan paling sering digunakan, karena

kemampuannya untuk pengelompokan dan pemetaan sejumlah besar data dalam waktu komputasi yang relatif cepat dan efisien (Malik *et al*, 2018). Sementara itu, Fuzzy C-Means (FCM) adalah teknik *clustering* yang menentukan posisi setiap titik data pada suatu kluster tertentu berdasarkan derajat keanggotaannya. FCM diterapkan dengan pendekatan untuk mengelompokan fuzzy terhadap indeks kekaburan berbasis *Euclidean Distance*. Oleh karenanya, data-data tersebut dapat menjadi anggota kluster dari semua kluster yang terbentuk dari derajat keanggotaannya yang bervariasi antara 0 hingga 1 (Zaenab *et al*, 2022).

Penelitian ini dilakukan dengan membuat aplikasi aggregator artikel ilmiah menggunakan bahasa pemrograman Python untuk mempermudah melakukan proses *clustering*. Data-data yang akan diklusterisasi berasal dari *digital library* yang kemudian dilakukan proses *scraping* untuk menyimpan data artikel ilmiah pada *database local computer*.

Pada penelitian ini, akan dilakukan clustering terhadap data artikel ilmiah dengan algoritma K-Means dan FCM untuk mengetahui perbedaan diantara hasil *clustering* dari keduanya. *Clustering* hanya akan dilakukan pada judul dan abstrak. Hasil clustering kemudian akan dievaluasi menggunakan *silhouette coefficient score* atau yang dikenal dengan sebutan *silhouette score*, untuk mengetahui perbandingan performa dari kedua algoritma *clustering* tersebut. Selain pengujian menggunakan *silhouette score*, penelitian ini juga akan diuji berdasarkan opini responden melalui survei. Survei dilakukan dengan menyebarkan kuesioner melalui media digital dan akan diisi oleh responden yang sebelumnya sudah menggunakan aplikasi aggregator artikel ilmiah untuk melakukan *clustering* berdasarkan kata kunci yang

diinginkan. Survei dilakukan untuk mengetahui hasil clustering algoritma mana antara K-Means dan FCM yang lebih unggul berdasarkan opini responden.

B. RUMUSAN MASALAH

Berdasarkan pada uraian latar belakang yang telah dijabarkan, penelitian ini akan membahas tentang:

1. Perbandingan performa hasil *clustering* antara k-mean dan fuzzy c-means menggunakan skore silhouette coefficient.
2. Penilaian hasil *clustering* dari kedua algoritma tersebut berdasarkan opini responden menggunakan Analisa statistik.

C. BATASAN MASALAH

Adapun batasan masalah yang akan dibahas dalam penelitian ini mencakup:

1. Penelitian dilakukan dengan pendekatan kuantitatif, data-data penelitian diperoleh melalui eksperimen dan survei.
2. Studi komparasi hanya akan dilakukan pada performa hasil klasterisasi K-Means dan Fuzzy C-Means (FCM), serta opini responden terhadap hasil dari kedua algoritma tersebut.
3. Data yang akan diuji merupakan data judul dan data abstrak hasil dari *scraping* data artikel ilmiah pada *digital library* seperti IEEE, ACM dan Springer Link yang sudah disimpan di *database local computer*.

D. TUJUAN PENELITIAN

Sesuai dengan rumusan dan batasan masalah yang telah ditentukan, tujuan dilakukan penelitian adalah melakukan studi komparatif terhadap performa hasil klasterisasi antara algoritma K-Means dengan FCM, serta opini responden terhadap hasil dari kedua algoritma tersebut. Proses klasterisasi dimulai dengan melakukan clustering terhadap data hasil *scraping* dari tiga *digital library* yaitu IEEE, ACM dan Springer Link yang sebelumnya sudah disimpan pada penyimpanan lokal komputer. Dari hasil *clustering* tersebut akan dilakukan survei terhadap opini responden tentang kualitas hasil *clustering* antara kedua algoritma.

E. MANFAAT PENELITIAN

Penelitian ini diharapkan dapat memberikan manfaat dan wawasan tentang algoritma mana, antara K-Means dan FCM, yang memiliki performa dan kinerja yang lebih unggul berdasarkan nilai *silhouette score* serta opini responden.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

BAB V

KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan temuan dari penelitian ini, dua kesimpulan utama dapat disimpulkan, yaitu terkait performa algoritma clustering yang diuji dan opini responden:

- 1) Berdasarkan hasil silhouette score, K-Means lebih unggul dari pada Fuzzy C-Means dalam hal *clustering* judul.
- 2) Berdasarkan hasil silhouette score, Fuzzy C-Means lebih unggul dari pada K-Means untuk clustering abstrak.
- 3) Berdasarkan opini responden, hasil clustering antara kedua algoritma tidak memiliki perbedaan (hasil pengelompokan kedua algoritma dianggap memiliki kualitas hasil yang sama).

B. Saran

Hasil clustering pada beberapa kata kunci menunjukkan skor silhouette yang rendah, bahkan mendekati -1, yang mengindikasikan kualitas clustering yang kurang optimal. Oleh karena itu, disarankan untuk mempertimbangkan penggunaan algoritma clustering alternatif yang dapat memberikan hasil lebih baik sesuai dengan karakteristik data. Selain itu, karena penelitian ini membatasi jumlah cluster hanya pada 5 cluster (jumlah cluster statis), disarankan untuk menerapkan pendekatan clustering yang lebih dinamis. Dengan menggunakan jumlah cluster yang adaptif, algoritma dapat menentukan jumlah cluster yang optimal berdasarkan data, sehingga diharapkan dapat meningkatkan kualitas hasil clustering secara keseluruhan.

DAFTAR PUSTAKA

- Altman, D. G., & Bland, J. M. (2005). Standard deviations and standard errors. *BMJ*, 331(7521), 903.
<https://doi.org/10.1136/bmj.331.7521.903>
- B. Bala & S. Behal. (2024). A Brief Survey of Data Preprocessing in Machine Learning and Deep Learning Techniques. *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 1755–1762.
<https://doi.org/10.1109/I-SMAC61858.2024.10714767>
- Cherradi, M., & El Haddadi, A. (2024). Exploration of Scientific Documents through Unsupervised Learning-Based Segmentation Techniques. *Seminars in Medical Writing and Education*, 3, 68. <https://doi.org/10.56294/mw202468>
- Diksha Bhola, Rachit Yadav, & Vimmi Malhotra. (2024). Introduction to Statistical Data Analysis. *International Journal of Advanced Research in Science, Communication and Technology*, 185–194. <https://doi.org/10.48175/IJAR SCT-17830>
- Efendi, M., & Mustofa, N. H. (2024). Memahami Esensi Metode Penelitian Kuantitatif. *TSAQOFAH*, 4(5), 3745–3755.
<https://doi.org/10.58578/tsaqofah.v4i5.3586>
- Ewens, W. J., & Brumberg, K. (2023). Non-parametric Tests. In W. J. Ewens & K. Brumberg (Eds.), *Introductory Statistics for Data Analysis* (pp. 205–217). Springer Nature Switzerland.
https://doi.org/10.1007/978-3-031-28189-1_14
- Jo, T. (2021). K Means Algorithm. In T. Jo (Ed.), *Machine Learning Foundations: Supervised, Unsupervised, and Advanced Learning* (pp. 217–240). Springer International Publishing.
https://doi.org/10.1007/978-3-030-65900-4_10
- Kitani, M., & Murakami, H. (2022). One-sample location test based on the sign and Wilcoxon signed-rank tests. *Journal of Statistical Computation and Simulation*, 92(3), 610–622.
<https://doi.org/10.1080/00949655.2021.1968399>
- Kusumadewi, S., Rosita, L., & Wahyuni, E. G. (2024). Performance of Fuzzy C-Means (FCM) and Fuzzy Subtractive Clustering (FSC) on Medical Data Imputation. *ComTech: Computer, Mathematics and Engineering Applications*, 15(1), 29–40.
<https://doi.org/10.21512/comtech.v15i1.11002>
- Li, X., & Tan, H. (2020). K-Means Algorithm Based on Initial Cluster Center Optimization. In Z. Xu, R. M. Parizi, M. Hammoudeh,

- & O. Loyola-González (Eds.), *Cyber Security Intelligence and Analytics* (pp. 310–316). Springer International Publishing.
- M Aditya Yoga Pratama, Agus Rahmad Hidayah, & Tertia Avini. (2023). CLUSTERING K-MEANS UNTUK ANALISIS POLA PERSEBARAN BENCANA ALAM DI INDONESIA. *Jurnal Informatika Dan Teknologi Komputer (JITEK)*, 3(2), 108–114. <https://doi.org/10.55606/jitek.v3i2.1506>
- M. Lv. (2022). Application of an K-means Improved Clustering Analysis Algorithm in the Design of Resource Management Information System. *2022 World Automation Congress (WAC)*, 158–162. <https://doi.org/10.23919/WAC55640.2022.9934387>
- M. Z. Ong, N. J. Y. Tan, S. M. Lim, & J. H. Ee. (2024). Unsupervised Analysis of Scientific Paper Abstracts: Exploring Clustering Algorithms. *2024 3rd International Conference on Digital Transformation and Applications (ICDXA)*, 231–236. <https://doi.org/10.1109/ICDXA61007.2024.10470550>
- Marín Díaz, G., Gómez Medina, R., & Aijón Jiménez, J. A. (2024). Integrating Fuzzy C-Means Clustering and Explainable AI for Robust Galaxy Classification. *Mathematics*, 12(18), 2797. <https://doi.org/10.3390/math12182797>
- Mohammad, N. (2023). *A Computational Theory and Semi-Supervised Algorithm for Clustering* (arXiv:2306.06974). arXiv. <https://doi.org/10.48550/arXiv.2306.06974>
- Mussabayev, R., & Mussabayev, R. (2024). *Comparative Analysis of Optimization Strategies for K-means Clustering in Big Data Contexts: A Review* (arXiv:2310.09819). arXiv. <https://doi.org/10.48550/arXiv.2310.09819>
- Mweshi, G. K., & Muhyila, M. (2024). Determining a Statistical Analysis for the Quantitative Study in Research. *Advances in Social Sciences Research Journal*, 11(7), 187–231. <https://doi.org/10.14738/assrj.117.17018>
- Mwita, K. (2022). Factors to consider when choosing data collection methods. *International Journal of Research in Business and Social Science (2147- 4478)*, 11(5), 532–538. <https://doi.org/10.20525/ijrbs.v11i5.1842>
- Patel, S. S., Kumar, N., Aswathy, J., Vaddadi, S. K., Akbar, S. A., & Panchariya, P. C. (2022). K-Means Algorithm: An Unsupervised Clustering Approach Using Various Similarity/Dissimilarity Measures. In J. S. Raj, R. Palanisamy,

- I. Perikos, & Y. Shi (Eds.), *Intelligent Sustainable Systems* (pp. 805–813). Springer Singapore.
- R. Kumar, P. B. Pati, K. Deepa, & S. Yanan. (2023). Clustering the Various Categorical Data: An Exploration of Algorithms and Performance Analysis. *2023 4th International Conference for Emerging Technology (INCET)*, 1–6.
<https://doi.org/10.1109/INCET57972.2023.10170508>
- Rahmawati, T., Wilandari, Y., & Kartikasari, P. (2024). ANALISIS PERBANDINGAN SILHOUETTE COEFFICIENT DAN METODE ELBOW PADA PENGELOMPOKKAN PROVINSI DI INDONESIA BERDASARKAN INDIKATOR IPM DENGAN K-MEDOIDS. *Jurnal Gaussian*, 13(1), 13–24.
<https://doi.org/10.14710/j.gauss.13.1.13-24>
- Ramjan, Sarawut, S., Jirapon. (2023). Clustering. In *Principles and Theories of Data Mining With RapidMiner* (pp. 139–159). IGI Global. <https://doi.org/10.4018/978-1-6684-4730-7.ch007>
- Rosyid, H., Bin Lakulu, M. M., & Bt. Mailok, R. (2024). Hybrid ABC–K Means for Optimal Cluster Number Determination in Unlabeled Data. *Mobile and Forensics*, 6(2), 52–65.
<https://doi.org/10.12928/mf.v6i2.11529>
- Safouan, S., El Moutaouakil, K., & Patriciu, A.-M. (2024). Fractional Derivative to Symmetrically Extend the Memory of Fuzzy C-Means. *Symmetry*, 16(10), 1353.
<https://doi.org/10.3390/sym16101353>
- Semple, H. K., & Langbart, M. J. (2022). Margin of error: Accuracy of estimated excision margins. *Australasian Journal of Plastic Surgery*, 5(1), 13–16. <https://doi.org/10.34239/ajops.v5n1.257>
- Sudaryono, Rahardja, U., Aini, Q., Isma Graha, Y., & Lutfiani, N. (2019). Validity of Test Instruments. *Journal of Physics: Conference Series*, 1364(1), 012050.
<https://doi.org/10.1088/1742-6596/1364/1/012050>
- T. A. Malapane & N. K. Ndlovu. (2024). Assessing the Reliability of Likert Scale Statements in an E-Commerce Quantitative Study: A Cronbach Alpha Analysis Using SPSS Statistics. *2024 Systems and Information Engineering Design Symposium (SIEDS)*, 90–95.
<https://doi.org/10.1109/SIEDS61124.2024.10534753>
- Thatha, V. N., Sekharreddy, B. V., Kiran, G. U., Srilakshmi, V., & Sanapala, S. (2024). SCIENTIFIC ARTICLE CLUSTERING USING R TREES AND K-MEANS ALGORITHM WITH

- GWO OPTIMIZATION TECHNIQUE. In Dr. J. Raut, Dr. S. Kushwaha, Dr. J. Nithyashri, R. Kar, Dr. G Vijaya, Dr. S. Sowmyayani, Dr. I. Hemalatha, Dr. S. A. Bhyratae, Dr. R. R., Dr. T. H. Sheikh, Dr. N. Thakur, R. V. Gandhi, Dr. K. B, Dr. Jacinth, Dr. A. M. Patokar, V. Punjabi, Dr. A. Balaram, Dr. J. J. Jeya, Dr. Anshu, & S. Kumar (Eds.), *Futuristic Trends in Computing Technologies and Data Sciences Volume 3 Book 1* (First, pp. 171–177). Iterative International Publisher, Selfypage Developers Pvt Ltd.
<https://doi.org/10.58532/V3BGCT1P5CH3>
- The Graduate School of Natural and Applied Sciences, Dokuz Eylul University, 'Izmir, Turkiye, & Gurler, O. (2024). COMPREHENSIVE ANALYSIS OF FUZZY C - MEANS CLUSTERING AND ITS VARIANTS. *Journal of Modern Technology and Engineering*, 9(1), 69–93.
<https://doi.org/10.62476/jmte9269>
- ‘Uyun, S., Sulistiyowati, E. S., & Jati, T. A. (2024). Fuzzy C-Means Algorithm Modification Based on Distance Measurement for River Water Quality. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*. <https://doi.org/10.22219/kinetik.v9i3.1991>
- Velunachiyar, S., & Sivakumar, K. (2023). Some Clustering Methods, Algorithms and their Applications. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(6s), 401–410.
<https://doi.org/10.17762/ijritcc.v11i6s.6946>
- Vora, A., Mishra, M., & Muthukumar, S. (2023). An Applied Computational Linguistics Approach to Clustering Scientific Research Papers. *Advances in Science and Technology*, 124, 378–391. <https://doi.org/10.4028/p-s8m133>
- Wani, A. A. (2024). Comprehensive analysis of clustering algorithms: Exploring limitations and innovative solutions. *PeerJ Computer Science*, 10, e2286. <https://doi.org/10.7717/peerj-cs.2286>
- Yon, A. (2024). Collection Data: Sharing, Discovery, Inspiration, and Innovation. *Art Libraries Journal*, 49(3), 85–94.
<https://doi.org/10.1017/alj.2024.12>