



INNOVATIVE: Journal Of Social Science Research

Volume 4 Nomor 3 Tahun 2024 Page 13232-13241

E-ISSN 2807-4238 and P-ISSN 2807-4246

Website: <https://j-innovative.org/index.php/Innovative>

## Penerapan Metode Fuzzy K-Means Clustering untuk Pengelompokan Konten Halaman Web secara Otomatis

Wahyu Budiono<sup>1✉</sup>, Bangkit Indarmawan Nugroho<sup>2</sup>, Nugroho Adhi Santoso<sup>3</sup>, Gunawan  
Gunawan<sup>4</sup>

STMIK YMI Tegal

Email: [budionowahyu505@gmail.com](mailto:budionowahyu505@gmail.com)<sup>✉</sup>

### Abstrak

Penerapan Metode Fuzzy K-Means Clustering untuk Pengelompokan Konten Halaman Web Secara Otomatis adalah penelitian yang bertujuan untuk mengotomatisasi proses pengelompokan konten halaman web menggunakan pendekatan clustering fuzzy. Dalam konteks ini, algoritma Fuzzy K-Means digunakan untuk mengelompokkan konten halaman web menjadi beberapa kategori berdasarkan kesamaan karakteristik tertentu. Metode ini memanfaatkan kelebihan pendekatan clustering fuzzy dalam menangani ketidakpastian dalam data dan kemampuan K-Means dalam mengelompokkan data menjadi beberapa cluster. Penelitian ini mencakup tahapan pra-pemrosesan data, ekstraksi fitur, dan implementasi algoritma Fuzzy K-Means Clustering. Eksperimen dilakukan menggunakan dataset yang berisi konten halaman web dari berbagai domain. Hasil evaluasi menunjukkan bahwa metode ini dapat menghasilkan pengelompokan konten halaman web yang sesuai dengan karakteristiknya secara otomatis, dengan tingkat akurasi dan interpretabilitas yang baik. Implementasi metode ini dapat memberikan kontribusi signifikan dalam pengelolaan dan penyaringan konten web secara efisien.

*Kata Kunci: Algoritma Fuzzy K-Means, Clustering Fuzzy, Interpretabilitas.*

## Abstract

Application of the Fuzzy K-Means Clustering Method for the Automatic Grouping of Web Page Content is a research aimed at automating the process of grouping web page content using a fuzzy clustering approach. In this context, the K- Means algorithm is used to group the contents of web pages into several categories based on the similarity of certain characteristics. The method takes advantage of the fuzzy clustering approach in dealing with uncertainty in data and the ability of K-means to group data into several clusters. The research includes the pre-processing phases of data, feature extraction, and the implementation of the K - Me clustering algorithm. Experiments were conducted using datasets containing web page contents from various domains. The evaluation results showed that this method can result in a grouping of web page content that matches its characteristics automatically, with a good degree of accuracy and interpretability. Implementation of this method can make a significant contribution to managing and filtering web content efficiently.

*Keywords: Fuzzy K-Means Algorithms, Fuzzy Clustering, Interpretability.*

## PENDAHULUAN

Dalam era digital yang semakin berkembang, jumlah halaman web terus meningkat secara eksponensial, menyebabkan kesulitan dalam mengelompokkan dan mengelola konten tersebut secara efisien (Hidayati et al., 2021). Pengelompokan konten halaman web menjadi kategori yang relevan adalah langkah penting dalam meningkatkan pengalaman pengguna dan efisiensi pencarian informasi. Metode clustering telah terbukti efektif dalam mengatasi tantangan ini dengan mengelompokkan data ke dalam beberapa kelompok berdasarkan kesamaan karakteristik tertentu (Arizal & Dengen, 2017). Namun, penggunaan metode clustering konvensional seperti K-Means dapat menghasilkan pengelompokan yang kurang akurat karena ketidakpastian dalam data web (Ikhwan, 2019). Hal ini tidak hanya mengurangi efisiensi pencarian informasi tetapi juga menurunkan kepuasan pengguna terhadap situs web. Dalam konteks ini, pendekatan clustering fuzzy menawarkan solusi yang lebih adaptif dengan memperhitungkan ketidakpastian dalam data. Metode Fuzzy K-Means Clustering menggabungkan keunggulan clustering fuzzy dengan kejelasan algoritma K-Means, memungkinkan pengelompokan yang lebih tepat dan representatif (Indarmawan Nugroho et al., 2023). Dengan menerapkan teknik ini pada konten halaman web, kami bertujuan untuk mengotomatisasi proses pengelompokan konten web secara efisien dan akurat (Muhazzir et al., 2019).

Penelitian ini dilakukan untuk mengatasi masalah pengelompokan konten halaman web secara otomatis, dengan tujuan untuk meningkatkan aksesibilitas dan relevansi informasi bagi pengguna. Kami berargumen bahwa dengan meningkatkan metode

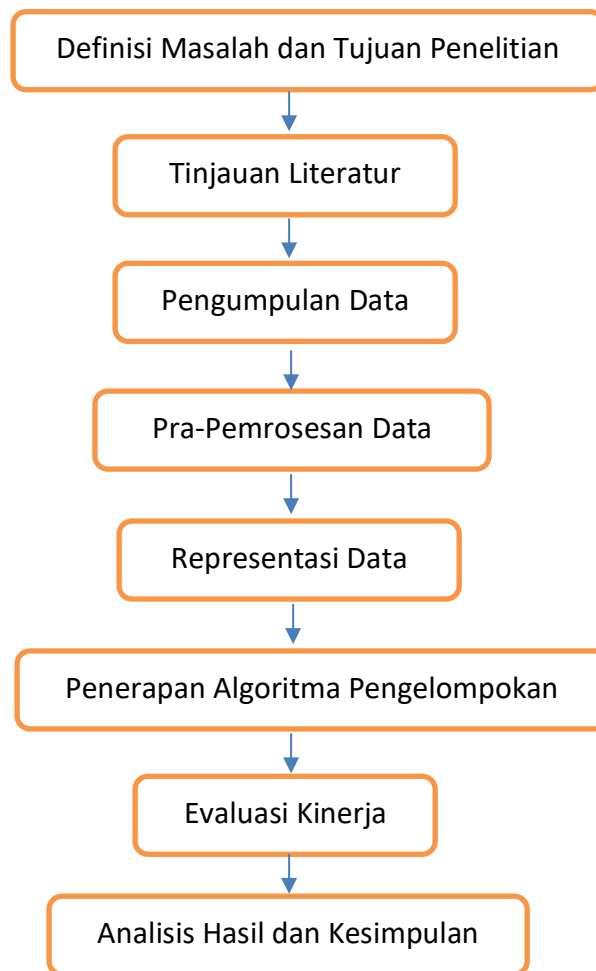
pengelompokan, dapat menciptakan pengalaman pengguna yang lebih intuitif dan memuaskan (Mahendra & Azizah, 2016). Masalah utama yang akan dipecahkan dalam penelitian ini adalah bagaimana meningkatkan efektivitas pengelompokan konten web secara otomatis menggunakan metode fuzzy k-means clustering. Ini penting karena akan memudahkan pengguna dalam menavigasi dan menemukan konten yang relevan dengan kebutuhan mereka, sehingga meningkatkan kepuasan pengguna dan efisiensi pencarian informasi (Sugumonrong et al., 2019). Untuk mengatasi masalah ini, penelitian ini akan menerapkan metode fuzzy k-means clustering, yang dikenal dengan kemampuannya untuk menangani ketidakpastian dan ketidakjelasan dalam pengelompokan data (Purwandito & Suyitno, 2019). Penelitian ini dilakukan dengan alasan bahwa metode ini dapat memberikan solusi yang lebih fleksibel dan akurat dalam pengelompokan konten web dibandingkan dengan metode pengelompokan tradisional. Penelitian ini diharapkan dapat berkontribusi dalam mengisi kesenjangan yang ada dalam literatur mengenai pengelompokan konten halaman web, khususnya dalam penerapan fuzzy k-means clustering (Marpaung & Marbun, 2021). Dengan mengeksplorasi dan mengevaluasi efektivitas metode ini, penelitian ini berambisi untuk menghasilkan wawasan baru dan praktek terbaik dalam pengelolaan konten web. Dalam hal metode, penelitian ini akan menggunakan pendekatan eksperimental untuk mengevaluasi penerapan fuzzy k-means clustering dalam pengelompokan konten web, dengan membandingkannya dengan metode pengelompokan lain dalam aspek efektivitas dan efisiensi (Suwanto et al., 2019). Sebagai state of the art, penelitian ini akan meninjau literatur yang ada mengenai teknik pengelompokan data, khususnya fuzzy k-means clustering, dan aplikasinya dalam berbagai konteks, termasuk pengelolaan konten web. Inovasi yang diusulkan meliputi pengembangan dan implementasi model fuzzy k-means clustering yang disesuaikan khusus untuk pengelompokan konten halaman web, dengan mempertimbangkan karakteristik unik dari data web (Ilmiah & Grafis, 2021).

Tujuan dari penelitian ini adalah untuk menginvestigasi kemampuan Metode Fuzzy K-Means Clustering dalam mengelompokkan konten halaman web secara otomatis. Kami akan menguraikan langkah-langkah pra-pemrosesan data, ekstraksi fitur, dan implementasi algoritma clustering fuzzy (Lutfi Fuadi et al., 2022). Selain itu, kami juga akan melakukan eksperimen menggunakan dataset yang mencakup berbagai jenis konten halaman web untuk mengevaluasi kinerja metode yang diusulkan. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam pengelolaan konten web yang efisien dan penyaringan informasi yang lebih baik (Lumbantobing, 2021). Penelitian ini akan

menghasilkan metode yang lebih efektif dan efisien untuk pengelompokan konten halaman web, memberikan kontribusi yang signifikan terhadap peningkatan navigasi dan pengalaman pengguna di situs web. Selain itu, hasil penelitian ini diharapkan dapat memberikan panduan bagi pengembang web dan desainer dalam mengimplementasikan strategi pengelompokan konten yang lebih baik (Redy Hermawan & Alam, 2020).

## METODE PENELITIAN

Langkah-langkah penelitian ditunjukkan pada Gambar 1 sebagai berikut:



Gambar 1. Langkah-langkah Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain quasi eksperimental. Pendekatan ini dipilih untuk mengukur efektivitas metode fuzzy k-means clustering dalam pengelompokan konten halaman web secara otomatis (Azadnia et al., 2023). Desain quasi eksperimental memungkinkan peneliti untuk mengamati perbedaan yang terjadi setelah penerapan metode fuzzy k-means clustering tanpa memanipulasi

variabel secara langsung (Ichwan & Hadi, 2023). Penjelasan dari langkah-langkah penelitian adalah sebagai berikut:

1. Definisi Masalah dan Tujuan Penelitian

Langkah awal dalam penelitian pengelompokan konten web adalah mendefinisikan masalah dan tujuan penelitian secara jelas. Hal ini penting untuk memastikan bahwa penelitian memiliki arah yang jelas dan fokus pada pertanyaan penelitian yang spesifik.

2. Tinjauan Literatur

Setelah mendefinisikan masalah dan tujuan penelitian, langkah selanjutnya adalah melakukan tinjauan literatur untuk mempelajari penelitian sebelumnya tentang pengelompokan konten web. Hal ini penting untuk memahami state-of-the-art dalam bidang ini dan mengidentifikasi teknik dan algoritma pengelompokan konten web yang relevan dengan penelitian Anda.

3. Pengumpulan Data

Langkah selanjutnya adalah mengumpulkan data web yang akan digunakan dalam penelitian. Data web dapat dikumpulkan dari berbagai sumber, seperti:

- a. Mesin pencari web
- b. Situs web
- c. Repositori data web

Penting untuk memastikan bahwa data web yang dikumpulkan relevan dengan tujuan penelitian dan berkualitas tinggi.

4. Pra-pemrosesan Data

Sebelum data web dapat digunakan untuk pengelompokan, perlu dilakukan pra-pemrosesan data. Pra-pemrosesan data dapat meliputi:

- a. Pembersihan teks: Menghapus noise dan karakter yang tidak relevan dari teks.
- b. Stemming: Mengubah kata menjadi bentuk dasarnya.
- c. Pembuangan kata berhenti: Menghapus kata-kata umum yang tidak memiliki makna signifikan.

5. Representasi Data

Data web perlu direpresentasikan dalam format yang dapat digunakan oleh algoritma pengelompokan. Representasi data yang umum digunakan dalam pengelompokan konten web adalah:

- a. Bag-of-Words (BoW): Merepresentasikan dokumen sebagai vektor yang berisi frekuensi kata-kata dalam dokumen.

- b. TF-IDF (Term Frequency-Inverse Document Frequency): Merepresentasikan dokumen sebagai vektor yang berisi bobot kata-kata dalam dokumen, dengan mempertimbangkan frekuensi kata dalam dokumen dan frekuensi kata dalam korpus dokumen.
6. Penerapan Algoritma Pengelompokan

Setelah data web direpresentasikan, algoritma pengelompokan dapat diterapkan untuk mengelompokkan data ke dalam kategori yang berbeda. Algoritma pengelompokan yang umum digunakan dalam pengelompokan konten web adalah:

    - a. K-Means Clustering: Algoritma clustering yang populer yang mengelompokkan data ke dalam sejumlah cluster yang telah ditentukan sebelumnya.
    - b. Hierarchical Clustering: Algoritma clustering yang membangun hierarki cluster secara bertahap.
    - c. Density-Based Spatial Clustering of Applications with Noise (DBSCAN): Algoritma clustering yang mengidentifikasi cluster berdasarkan kepadatan data.
    - d. Self-Organizing Maps (SOMs): Algoritma clustering yang memetakan data ke dalam ruang dua dimensi.
  7. Evaluasi Kinerja

Setelah data web dikelompokkan, perlu dilakukan evaluasi kinerja algoritma pengelompokan untuk memastikan bahwa algoritma menghasilkan pengelompokan yang akurat dan bermakna. Metrik evaluasi yang umum digunakan dalam pengelompokan konten web adalah:

    - a. Purity: Mengukur proporsi data dalam setiap cluster yang berasal dari kelas yang sama.
    - b. Silhouette Coefficient: Mengukur seberapa baik data dalam setiap cluster dipisahkan dari data di cluster lain.
    - c. Entropy: Mengukur tingkat ketidakpastian dalam pengelompokan.
  8. Analisis Hasil dan Kesimpulan

Setelah evaluasi kinerja, perlu dilakukan analisis hasil untuk memahami pengelompokan yang dihasilkan dan interpretasikan maknanya. Kesimpulan penelitian kemudian ditarik berdasarkan hasil analisis.

## Klasterisasi

Tahap *clustering* ini menggunakan Algoritma K- Means. Pada tahap ini akan dicari *centroid* yang terdekat berdasarkan nilai *euclidean distance* yang dicari dari setiap

dokumen. Adapun contoh beserta langkah-langkah tahap *clustering* dengan menerapkan algoritma K-Means dengan sebagai berikut :

1. Menentukan jumlah *cluster* dan *centroid* awal. Banyaknya *cluster* harus lebih kecil dari banyaknya data ( $k < n$ ). Berdasarkan pada contoh kasus sebelumnya nilai  $k$  yang digunakan adalah 3 dan *centroid* awal yang terpilih secara random (D1, D4 dan D5). Maka *centroid* pertama pada *cluster* C1 diinisialisasikan pada D1, *centroid* kedua pada *cluster* C2 berada pada D4 dan *centroid* pada *cluster* C3 berada pada D5.
2. Langkah selanjutnya adalah mencari *centroid* terdekat dari setiap dokumen. Pada penelitian ini akan menerapkan pendekatan berdasarkan *euclidean distance* yang digunakan untuk mencari jarak antara masing- masing dokumen dengan *centroid*. Adapun rumus pengukuran *euclidean distance* yang didefinisikan pada rumus (1):

$$d_{(i,j)} = \sqrt{(|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{in} - x_{jn}|^2)} \quad (1)$$

Keterangan :

$d_{(i,j)}$  : jarak dokumen ke  $i$  ke dokumen ke  $j$

$x_{in}$  : kata ke  $n$  di dokumen ke- $i$

$x_{jn}$  : kata ke  $n$  di dokumen ke- $j$

3. Langkah selanjutnya menghitung ulang untuk menentukan *centroid* baru dari setiap *cluster*. Adapun rumus (2) untuk menentukan *centroid* baru :

$$M_k = \left(\frac{1}{n_k}\right) \sum_{i=1}^{n_k} X_{ik} \quad (2)$$

Keterangan :

$M_k$  :Nilai *centroid* dari suatu *cluster*

$n_k$  :Jumlah dokumen yang berada dalam satu *cluster*

$x_{ik}$  :Nilai vektor dari sampel dokumen ke- $i$  yang termasuk ke dalam *cluster* C $k$ .

4. Langkah terakhir adalah menghitung apakah *centroid* lama sama dengan *centroid* baru. Untuk itu akan digunakan kembali rumus *euclidean distance* antara *centroid* baru dengan *centroid* lama.

## HASIL DAN PEMBAHASAN

Pengujian akurasi *clustering* informasi berita dilakukan untuk mengetahui tingkat akurasi pengelompokan informasi berita pendidikan yang dilakukan oleh algoritma K-Means pada sistem *news aggregator* dengan pengelompokan yang dilakukan secara manual. Pengujian dilakukan dengan menggunakan 132 dokumen yang diambil dari hasil Copyright @ Wahyu Budiono, Bangkit Indarmawan Nugroho, Nugroho Adhi Santoso, Gunawan

*crawling*, dimana dari hasil *crawling* tersebut ditemukan informasi berita yang dengan kata kunci "pendidikan" dan telah dilakukan proses *clustering* sebelumnya dengan nilai *k* ditentukan secara dinamis yang dapat dilihat pada Tabel Hasil pengelompokan informasi berita pendidikan yang ditunjukkan pada tabel 1 sebagai berikut:

Tabel 1. Hasil pengelompokan informasi berita pendidikan

	Informasi Pendidikan Dalam 1 Cluster	Informasi Pendidikan Bukan Dalam 1 Cluster
News Agregat	8	124
<i>or</i>		
Manual	6	126

Adapun tabel *confusion matrix* dimana untuk membandingkan hasil prediksi atau yang dilakukan dengan sistem dengan yang dilakukan secara manual yang dapat dilihat pada Tabel Confusion Matrix yang ditunjukkan pada tabel 2 sebagai berikut:

Tabel 2. Confusion Matrix

Predicted Class/ Actual Class	Manual Cluster	Manual Bukan Cluster
Sistem Cluster	6	2
Sistem Bukan Cluster	0	124

Setelah sistem melakukan *clustering*, lalu hitung nilai akurasi berdasarkan pada rumus Akurasi sebagai berikut:

$$\text{Akurasi} = \frac{6 + 124}{6 + 2 + 0 + 124} = 0.98$$

Berdasarkan pengujian akurasi, didapatkan hasil akurasi *clustering* berita dengan kata kunci "pendidikan" pada sistem *news aggregator* dengan menggunakan algoritma K-Means sebesar 98%. Kesimpulan yang diperoleh dari pengujian akurasi ini adalah bahwa algoritma K-Means dapat digunakan sebagai metode pengelompokan untuk topik berita tertentu dari hasil *crawling* situs berita karena tingkat akurasi yang besar, hanya saja kekurangan dari algoritma K-Means ini masih terdapat *outlier* atau dokumen yang seharusnya tidak dalam satu *cluster*.

## SIMPULAN

Metode Fuzzy K-Means Clustering memungkinkan dokumen memiliki derajat keanggotaan dalam lebih dari satu cluster, yang membantu dalam menangani

Copyright @ Wahyu Budiono, Bangkit Indarmawan Nugroho, Nugroho Adhi Santoso, Gunawan  
Gunawan

ketidakpastian dalam data teks. Ini membuat pengelompokan konten halaman web lebih fleksibel dan realistis. Hasil pengelompokan menunjukkan efektivitas metode ini dalam mengelompokkan konten berdasarkan kesamaan fitur. Penerapan metode Fuzzy K-Means Clustering untuk pengelompokan konten halaman web menunjukkan hasil yang efektif dan akurat dalam mengelompokkan artikel ke dalam kategori yang sesuai. Dengan fleksibilitas yang diberikan oleh metode fuzzy, pengelompokan menjadi lebih realistis dan sesuai dengan karakteristik alami dari data teks. Evaluasi kinerja yang baik menunjukkan bahwa metode ini layak untuk digunakan dalam aplikasi nyata di industri konten digital.

#### DAFTAR PUSTAKA

- Arizal, M., & Dengen, N. (2017). APLIKASI LOGIKA FUZZY DALAM OPTIMISASI STOK BAHAN MENGGUNAKAN METODE TSUKAMOTO. *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, 2(1).
- Azadnia, R., Fouladi, S., & Jahanbakhshi, A. (2023). Intelligent detection and waste control of hawthorn fruit based on ripening level using machine vision system and deep learning techniques. *Results in Engineering*, 17. <https://doi.org/10.1016/j.rineng.2023.100891>.
- Hidayati, N. U., Saripurna, D., & Ginting, R. I. (2021). IMPLEMENTASI LOGIKA FUZZY DALAM MENENTUKAN PERSEDIAAN BARANG DENGAN METODE MAMDANI. *Jurnal CyberTech*, 4(2). <https://ojs.trigunadharma.ac.id/>.
- Ichwan, M., & Hadi, S. (2023). MIND (Multimedia Artificial Intelligent Networking Database Kinerja Model EfficientNetV2M dalam Klasifikasi Citra Tutupan dan Penggunaan Lahan. *Journal MIND Journal / ISSN*, 8(2), 203–216. <https://doi.org/10.26760/mindjournal.v8i2.203-216>.
- Ikhwan, A. (2019). Penerapan Fuzzy Mamdani Untuk Sistem Pendukung Keputusan Pemilihan Laptop. *JURNAL FASILKOM*, 9(2), 476–483. <https://doi.org/10.37859/jf.v9i2.1407>.
- Ilmiah, J., & Grafis, K. (2021). PENERAPAN METODE FUZZY MAMDANI DALAM MENENTUKAN HARGA JUAL PONSEL PINTAR BEKAS (STUDI KASUS PADA KAYYIS CELLULAR DEPOK). 14(2), 253–262. <http://journal.stekom.ac.id/index.php/pixel/page253>.
- Indarmawan Nugroho, B., Adam Hidayatullah, B., Alim Murtopo, A., YMI Tegal, S., Pendidikan No, J., & Lor, P. (2023). Sistem Pakar Diagnosa Penyakit Gigi Periodontal Menggunakan Metode Fuzzy Mamdani EXPERT SYSTEM FOR DIAGNOSIS OF

PERIODONTAL DENTAL DISEASE USING THE FUZZY MAMDANI METHOD.  
*TEKNOMATIKA*, 13(02).

- Lumbantobing, T. M. (2021). *Sistem Pendukung Keputusan Pemilihan Majelis Perbendaharaan Gereja dengan Metode Fuzzy Logic Modell Mamdani (Studi kasus di gereja HKBP Semarang Barat)*. 1(1), 26–39. <http://journal.politeknik-pratama.ac.id/index.php/JTIM>.
- Lutfi Fuadi, A., Kunci, K., Stok Beras, P., & Mamdani, F. (2022). *OKTAL: Jurnal Ilmu Komputer dan Science IMPLEMENTASI LOGIKA FUZZY MAMDANI UNTUK MENENTUKAN STOK BERAS DI TOKO AGUNG CAHAYA BERBASIS WEB*. 1(10).
- Mahendra, D., & Azizah, N. (2016). IMPLEMENTASI FUZZY INFERENCE SYSTEM TSUKAMOTO UNTUK PENENTUAN TOPIK TUGAS AHIR. *Jurnal SIMETRIS*, 1(1).
- Marpaung, D. A., & Marbun, M. (2021). Sistem Pendukung Keputusan Penentuan Tingkat Kecanduan Masyarakat Terhadap Rokok dengan Metode Fuzzy Mamdani. *Jurnal Nasional Komputasi Dan Teknologi Informasi*, 4(1).
- Muhazzir, A., Reza, F., Satria, B., Wahyuni, S., Lubis, Z., Annisa, S., & Nando Winata, H. (2019). SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN SMARTPHONE ANDROID MENGGUNAKAN METODE FUZZY TSUKAMOTO. In *Cetak) Buletin Utama Teknik* (Vol. 14, Issue 3). Online.
- Purwandito, R., & Suyitno, H. (2019). PENERAPAN SISTEM INFERENSI FUZZY METODE MAMDANI UNTUK PENENTUAN JUMLAH PRODUKSI EGGROLL. In *UJM* (Vol. 8, Issue 1). <http://journal.unnes.ac.id/sju/index.php/ujm>
- Redy Hermawan, M., & Alam, R. (2020). KLIK: Kajian Ilmiah Informatika dan Komputer Logika Fuzzy Mamdani Untuk Mendukung Keputusan Pembelian Laptop Asus M409BA Berdasarkan Spesifikasi Yang Tersedia. *Media Online*, 1(3), 99–103. <https://djournals.com/klik>.
- Sugumonrong, D. P., Handinata, A., & Tehja, A. (2019). Prediksi Harga Emas Menggunakan Metode Fuzzy Time Series Model Algoritma Chen. In *PSDKU Medan Jurusan Teknik Informatika INFORMATICS ENGINEERING RESEARCH AND TECHNOLOGY*.
- Suwanto, S., Bisri, M. H., Novitasari, D. C. R., & Asyhar, A. H. (2019). Classification of EEG Signals using Fast Fourier Transform (FFT) and Adaptive Neuro Fuzzy Inference System (ANFIS). *Jurnal Matematika "MANTIK,"* 5(1), 35–44. <https://doi.org/10.15642/mantik.2019.5.1.35-44>.