

Implementasi Algoritma Fuzzy C-Means Untuk Pengelompokan Kabupaten Atau Kota di Jawa Timur Berdasarkan Tingkat Kesejahteraan
Implementation of the Fuzzy C-Means Algorithm for Grouping Regencies or Cities in East Java Based on Welfare Levels

Umar Dani¹, Hardian Oktavianto², Habibatul Azizah Al Faruq³

¹Mahasiswa Fakultas Teknik, Universitas Muhammadiyah Jember

e-mail: daniumar299@gmail.com¹

^{2,3}Dosen Fakultas Teknik, Universitas Muhammadiyah Jember

Abstrak

Keberhasilan pembangunan sebuah negara dapat diukur melalui tingkat kesejahteraannya. Peningkatan kesejahteraan adalah bukti keberhasilan dari rencana pembangunan setiap negara, termasuk Indonesia. Pembenahan dilakukan di setiap provinsi Indonesia. Namun, perbedaan kandungan sumber daya, sumber daya demografi, dan pembangunan di setiap wilayah menyebabkan perbedaan tingkat kesejahteraan masyarakat yang bersangkutan, salah satunya dipengaruhi oleh disparitas pembangunan di Provinsi Jawa Timur. Dalam hal ini, untuk mengatasi ketimpangan kesejahteraan, diperlukan upaya untuk mengetahui tingkat kesejahteraan suatu wilayah, kabupaten, atau kota. Pengelompokan dapat digunakan untuk menentukan wilayah mana yang paling penting. Salah satu metode untuk mengetahui cluster mana yang paling baik adalah dengan menggunakan algoritma fuzzy c-means. Data yang digunakan terdiri dari 38 kabupaten/kota di Jawa Timur dari tahun 2018 hingga 2019. Berdasarkan pengujian dari tiga hingga sepuluh cluster, tiga cluster dianggap sebagai cluster terbaik dengan nilai partition coefficient index 0,531075. Cluster 1 memiliki 8 anggota kabupaten/kota, cluster 2 memiliki 13 anggota, dan cluster 3 memiliki 17 anggota.

Kata kunci: kesejahteraan, *fuzzy c-means*, *partitioning coefficient index*, jawa timur, *clustering*.

Abstract

The success of a country's development can be measured by its level of welfare. Increased welfare is proof of the success of the development plans of every country, including Indonesia. Improvements were made in every province of Indonesia. However, differences in resource content, demographic resources, and development in each region lead to differences in the level of welfare of the people concerned, one of which is influenced by development disparities in East Java Province. In this case, to overcome inequality in welfare, efforts are needed to determine the level of welfare of a region, district or city. Grouping can be used to determine which areas are most important. One method to find out which cluster is the best is to use the fuzzy c-means algorithm. The data used consists of 38 regencies/cities in East Java from 2018 to 2019. Based on tests of three to ten clusters, three clusters are considered the best clusters with a partition coefficient index value of 0.531075. Cluster 1 has 8 district/city members, cluster 2 has 13 members, and cluster 3 has 17 members.

Keywords: *welfare, fuzzy c-means, jawa timur, clustering, partition coefficient index.*

1. PENDAHULUAN

Tingkat kesejahteraan sebuah negara merupakan sebuah indikator keberhasilan pembangunan sebuah negara. Peningkatan kesejahteraan merupakan cerminan keberhasilan dari pembangunan yang diharapkan oleh setiap negara, salah satunya di Indonesia. (Sagita & Dina 2021).

Berdasarkan penjelasan diatas, sebagai sebuah bangsa yang memiliki cita-cita untuk menjadi bangsa yang makmur secara merata, Indonesia terus melakukan pembenahan dan inovasi. Pembenahan tersebut dilakukan di seluruh provinsi di Indonesia (Sagita & Dina 2021). Namun, karena adanya perbedaan kandungan sumber daya, perbedaan sumberdaya demografi serta pembangunan setiap wilayah menyebabkan terjadi ketimpangan tingkat kesejahteraan masyarakat yang bersangkutan dipengaruhi oleh adanya disparitas pembangunan (Simbolon, 2017) salah satunya di Provinsi Jawa Timur. Penelitian (Abidin,2017) menyebutkan bahwa indikator kesejahteraan di Provinsi Jawa Timur lebih banyak tersedia di wilayah perkotaan sehingga akses untuk memperoleh sarana dan prasarana lebih mudah dibandingkan oleh wilayah lain seperti kabupaten. Oleh karena itu, hal tersebut menyebabkan terjadinya ketimpangan kesejahteraan wilayah antara satu dengan yang lain.

Salah satu upaya yang dapat dilakukan untuk mengatasi ketimpangan adalah dengan melakukan metode pengelompokan atau *clustering* guna mengetahui bagian wilayah mana yang memiliki tingkat kesejahteraan rendah sehingga dapat dilakukan kebijakan lebih lanjut serta menentukan daerah prioritas pembangunan agar tingkat kesejahteraan masyarakat dapat ditingkatkan dan terjadi secara merata. Terdapat beberapa algoritma pengelompokan atau *clustering* diantara algoritma *Fuzzy-Cmeans* dan algoritma *K-Means*.

Penelitian sebelumnya dilakukan oleh (Murtadlo, 2021) dengan studi kasus “Pengelompokan Tingkat Buta Huruf di

Provinsi Indonesia Tahun 2015-2019”. Pada penelitian ini dilakukan perbandingan antara algoritma *Fuzzy-Cmeans* dengan algoritma *K-Means* dengan skenario pengujian 3 *cluster* sampai 10 *cluster*. Penelitian ini menunjukkan nilai performa validitas algoritma *Fuzzy-Cmeans* (0,768) lebih maksimal jika di bandingkan algoritma *K-Means* (0,4966). Penelitian lainnya dilakukan (Vera & Darwis, 2021) dengan studi kasus “Analisis *Clustering* Fasilitas Kesehatan menggunakan Metode *Fuzzy C-Means*”. Pada penelitian ini *cluster* pertama menghasilkan 479 *cluster* dan *cluster* dua menghasilkan 580 *cluster* serta menunjukkan nilai tingkat akurasi dari nilai keanggotaan (0,99998) yang artinya tingkat akurasi dari nilai keanggotaan klaster cukup baik.

Berdasarkan beberapa paparan diatas, maka akan dilakukan penelitian *clustering* tingkat kesejahteraan di Provinsi Jawa Timur, sehingga mudah dalam meningkatkan tingkat kesejahteraannya, dengan data yang di ambil dari badan pusat statistik (BPS) pada tahun 2018-2019 pada kabupaten atau kota di Jawa Timur. Penelitian ini diharapkan dapat menjadi salah satu alternatif penelitian untuk mengelompokkan data tingkat kesejahteraan masyarakat untuk memastikan kesejahteraan di seluruh wilayah jawa timur dengan mengidentifikasi wilayah mana yang perlu diprioritaskan dalam hal pembangunan sekolah, sarana kesehatan, dan peningkatan lainnya.

2. STUDI PUSTAKA

A. *Data Mining*

Data mining adalah bagian dari beberapa bidang ilmu sebagai sebuah metode untuk menggali informasi dan mendapatkan pola serta pengetahuan yang tidak diketahui keberadaannya. *Data mining* adalah suatu multi disiplin yang menggambarkan karya dari berbagai bidang termasuk teknologi pada *database*, *machine learning*, *artificial intelligence*, *neural network*, *statistic*, *pattern*

recognition, sistem dengan basis pengetahuan, pengetahuan akuisisi, pencarian informasi, komputasi kinerja tinggi dan visualisasi data (Hand dan Kamber, 2006)

Menurut Hand dkk, pada Larose (2014), data mining adalah bidang interdisipliner yang menyatukan teknik pembelajaran mesin, pengenalan statistik, pola, database dan visualisasi untuk mengatasi masalah ekstraksi dari basis data yang besar, sedangkan menurut Prescott dkk (2005), data *mining* ialah pengetahuan hasil dari penemuan atau penelitian memakai cara-cara yang tergabung dari tradisional, statistik, grafik komputer dan *artificial intelligence*.

B. Clustering

Dalam *data mining* terdapat teknik yang disebut *clustering* atau pengelompokan. *Clustering* merupakan pengelompokan berbagai data atau objek ke dalam satu grup yang sama sehingga menjadikan satu *cluster* yang memiliki kesamaan antara data satu dengan yang lainnya dan berbeda data atau objek dalam *cluster* yang lainnya. *Cluster analysis* adalah suatu kumpulan teknik yang cukup luas cakupannya, yang mana dirancang sedemikian rupa untuk menemukan kelompok dari suatu item serupa atau sejenis dalam sekumpulan data (Holland, 2019).

Clustering membutuhkan beberapa syarat (Han dan Kamber, 2012): skalabilitas, kemampuan untuk menganalisis berbagai bentuk data; kemampuan untuk menemukan cluster dengan bentuk yang tidak terduga, kemampuan untuk menangani suara, kemampuan untuk melakukan clustering untuk data dimensi tinggi, dan interpretasi dan kegunaan.

C. Normalisasi Data

Normalisasi data adalah proses menskalakan nilai atribut data sehingga dapat diposisikan dalam rentang tertentu (Nasution, Khotimah, & Chamidah, 2019). Normalisasi min-max, normalisasi skor z, scaling decimal, dan sigmoidal adalah beberapa metode yang dapat digunakan untuk normalisasi data (Alihar, 2018).

Metode normalisasi min-man dapat dilakukan dengan rumus sebagai berikut:

$$x_j = \frac{x_i - \text{value of } x}{\text{value of } x - \text{value of } x} \quad (1)$$

Keterangan

x_j = data setelah normalisasi

x_i = data ke-*i*

D. Fuzzy C-Means

Menurut (Aminah, dkk. 2019), *Fuzzy C-Means* merupakan teknik yang ada pada metode *clustering*, cara kerja dari teknik ini yaitu dengan mengelompokkan data yang ada dalam satu *cluster* yang telah ditentukan menurut nilai derajat keanggotaan. Teori dasar *Fuzzy C-Means* adalah menentukan pusat *cluster* yang akan menandai lokasi rata-rata untuk setiap *cluster*. Pada kondisi awal, kemungkinan besar pusat *cluster* masih belum akurat. Selain itu, setiap data memiliki derajat keanggotaan untuk setiap *cluster* dan nilai keanggotaan setiap data secara iteratif, maka pusat *cluster* akan berpindah mendekati lokasi yang tepat dalam dataset. Iterasi ini di dasarkan pada meminimalkan fungsi objek yang mewakili jarak dari titik data mana saja yang di berikan ke pusat *cluster* dan diperhitungkan dengan derajat keanggotaan titik data tersebut (Gelley, 2000).

E. Uji validitas

Setelah melakukan *clustering*, maka selanjutnya yang perlu dilakukan adalah menganalisis *cluster*. Analisis *cluster* adalah salah satu langkah dari uji validasi yang bertujuan untuk mengambil keputusan dari hasil *cluster* secara objektif dan kuantitatif.

F. Partition Coefficient Index

Menurut (Prasetyo, 2014) nilai *Partition Coefficient Index* (PCI) hanya mengoreksi nilai derajat keanggotaan tanpa dipengaruhi data yang biasanya mengandung informasi geometrik. Nilai dalam rentang [0,1] nilai yang semakin tinggi mendekati 1 mempunyai arti bahwa kualitas *cluster* yang

dapat semakin baik, *Partition Coefficient Index* didefinisikan sebagai berikut :

$$\begin{aligned}
 & \text{Partition Coefficient Index (C)} \\
 & = \frac{1}{N} \sum_{i=1}^C \sum_j^N (\mu_{ij})^2
 \end{aligned}$$

Keterangan:

- C = Jumlah *cluster*
- N = Jumlah data
- μ_{ij} = Derajat keanggotaan data ke-*j* pada *cluster* ke-*i*.

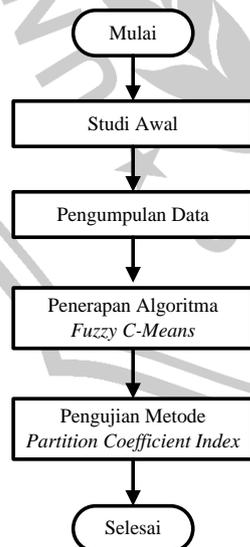
3. METODOLOGI PENELITIAN

A. Metode Penelitian

Penelitian ini memakai metode penelitian kuantitatif. Penelitian kuantitatif merupakan pengumpulan data dan penguraian hasilnya menggunakan angka. Karena data yang dipakai pada penelitian ini merupakan data numerik, sehingga penelitian ini termasuk penelitian kuantitatif (Arikunto, 1997).

B. Diagram Alur Penelitian

Adapun tahapan penelitian yang dilakukan dalam penelitian ini yaitu studi awal, pengumpulan data, penerapan algoritma *Fuzzy C-Means* dan pengujian menggunakan metode *partition coefficient index* berikut adalah diagram alur penelitian:



C. Studi Awal

Pada tahapan awal pada penelitian ini yakni mempelajari dan menemukan topik permasalahan yang akan di teliti yaitu tingkat kesejahteraan masyarakat di Provinsi Jawa Timur, kemudian menentukan apa saja yang termasuk dalam ruang lingkup masalah, latar belakang serta mengkaji literatur-literatur yang berhubungan dengan topik masalah, lalu kemudian menemukan solusi dari masalah tersebut agar tercapai tujuan tersebut tentu peneliti harus membaca dan memahami literatur-literatur yang berkaitan dengan topik penelitian, kemudian literatur tersebut dipakai dalam penelitian yang diharapkan menjadi salah satu solusi permasalahan kesejahteraan masyarakat di Provinsi Jawa Timur.

D. Pengumpulan Data

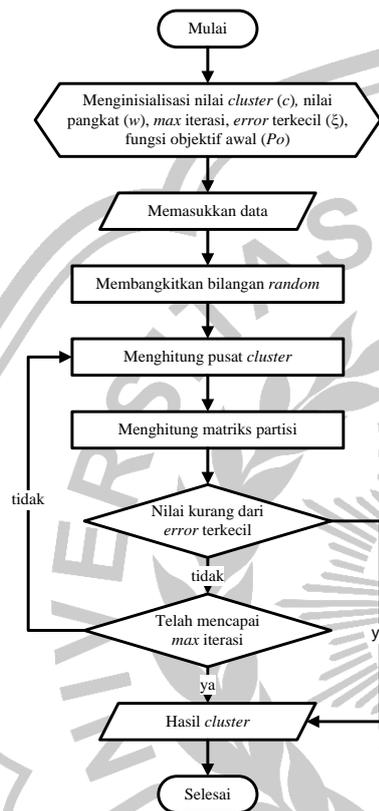
Data yang di pakai pada penelitian ini merupakan data yang di akses kemudian di ambil secara *opensource* dari *website* resmi Badan Pusat Statistik (BPS) Provinsi Jawa Timur dengan judul “Kesejahteraan Masyarakat di Provinsi Jawa Timur”. Dataset yang dipakai pada penelitian ini ialah kesejahteraan masyarakat di Provinsi Jawa Timur, data yang diambil ialah sebanyak 38 data. Data ini akan di *cluster* menggunakan algoritma *Fuzzy C-Means* yang nantinya akan divalidasi algoritma.

Tabel 3.1 Data Kabupaten/Kota

Kab/Kot	TK	FK	JSK	...	IPM
Pacitan	1396	42	709	...	68,16
Ponorogo	989	64	1009	...	70,56
Trenggalek	1793	34	730	...	69,46
Tulungagung	1403	75	981	...	72,62
Blitar	1997	65	1139	...	70,57
Kediri	1674	103	1234	...	71,85
Malang	1144	171	2291	...	70,35
Lumajang	1870	51	1111	...	65,33
Jember	674	119	2321	...	66,69
...
Kota Batu	676	34	148	...	75,88

E. Implementasi Algoritma Fuzzy C-Means

Adapun alur diagram dalam pengimplementasian algoritma Fuzzy C-Means yang diterapkan pada data tingkat kesejahteraan masyarakat di Provinsi Jawa Timur dapat digambarkan sebagai berikut:



Gambar 3.2 Flowchart Fuzzy C-Means

Tahapan dalam proses clustering menggunakan Fuzzy C-Means, sebagai berikut :

Pada proses pengelompokan algoritma fuzzy c-means tahap pertama yaitu dengan memasukkan data yang akan di cluster. Sejumlah 38 data Kabupaten atau Kota dengan 7 parameter, yaitu fasilitas kesehatan, tenaga kesehatan, jumlah sekolah, guru, sumber penerangan listrik, tingkat partisipasi angkatan kerja dan indeks pembangunan manusia. Selanjutnya menginisialisasi parameter awal yang dibutuhkan. Lalu membangkitkan bilangan acak sebagai nilai awal derajat keanggotaan. Kemudian lakukan perhitungan untuk mencari pusat cluster. Selanjutnya lakukan perhitungan

fungsi objektif. Setelah itu lakukan perhitungan perubahan partisi. Tahapan selanjutnya yaitu memeriksa syarat berhenti, Jika telah sampai error terkecil dan batas maksimal iterasi terpenuhi maka tahapan iterasi berakhir.

F. Metode Validasi Partition Coefficient Index

Pada proses pengelompokan menggunakan algoritma fuzzy c-means yang telah dilakukan sebelumnya, yakni pada rentang 3 sampai 10 cluster. Selanjutnya diterapkan metode validasi menggunakan partition coefficient index untuk menentukan cluster mana yang merupakan cluster terbaik atau optimum

4. HASIL DAN PEMBAHASAN

Selanjutnya data kabupaten atau kota akan dilakukan normalisasi data menggunakan normalisasi min-max tujuannya adalah mengubah nilai kolom numerik dalam himpunan data untuk menggunakan skala umum, tanpa mendistorsi perbedaan dalam rentang nilai atau kehilangan informasi

A. Algoritma Fuzzy C-Means

Data yang telah dilakukan normalisasi terlebih dahulu selanjutnya diproses memakai tools RStudio untuk diterapkan clustering menggunakan fuzzy c-means dengan skenario 3 sampai 10 cluster. Hasil keluaran dari perintah RStudio diantaranya berupa jumlah data yang ada pada masing-masing cluster, jumlah iterasi yang terdapat di masing-masing cluster, pusat cluster, fungsi objektif, nilai derajat keanggotaan. Selain itu, perintah RStudio juga dapat menampilkan cluster kedalam bentuk plot. Adapun contoh plot yang merupakan hasil dari fuzzy c-means pada RStudio yaitu berupa



Gambar 1. Plot 3 cluster

Berdasarkan pada **Gambar 1** diatas *cluster* 1 memiliki anggota yang terdiri dari 7 Kabupaten atau Kota, dan *cluster* 2 memiliki anggota yang terdiri dari 13 anggota Kabupaten atau Kota, dan *cluster* 3 memiliki anggota 17 Kabupaten atau Kota.

B. Penentuan Cluster Optimum

Menentukan *cluster* terbaik atau optimum dari perhitungan menggunakan algoritma *fuzzy c-means* sebelumnya, dilanjutkan dengan menerapkan metode validasi *partition coefficient index*. Adapun hasil keluaran yang didapat yaitu berupa nilai seperti pada **Tabel 2** berikut:

Tabel 4.1 Perhitungan PCi

Cluster	Partition Coefficient
3	0,53175
4	0,48128
5	0,43004
6	0,41248
7	0,40927
8	0,38762
9	0,38949
10	0,38468

Sumber: Perhitungan *Partition Coefficient Index* pada RStudio

Pada metode ini, nilai *cluster* yang menjadi penentu *cluster* terbaik atau optimum adalah nilai validitas *partition coefficient index* yang paling mendekati 1. Maka *cluster* yang sesuai kriteria tersebut dan menjadi *cluster* optimum adalah *cluster* 3.

5. Kesimpulan dan Saran

A. Kesimpulan

Berdasarkan hasil yang didapat dari perhitungan pada penelitian yang telah dilakukan, penerapan *Fuzzy C-Means* pada data Kabupaten/Kota untuk mengelompokkan Kabupaten/Kota di Jawa Timur menghasilkan *cluster* optimum pada 3 *cluster* dengan nilai *Partition Coefficient Index* senilai 0,53175.

Juga dapat diketahui bahwa nilai koefisien partisi 3 sampai 10 *cluster* memiliki nilai derajat keanggotaan yang berbeda-beda, dan nilai yang lebih tinggi serta mendekati nilai 1 adalah *cluster* yang optimum jika dibandingkan dengan *cluster* lainnya.

Cluster optimal dari penelitian ini terdapat pada 3 *cluster*, yang memiliki keanggotaan masing-masing *cluster* yaitu *cluster* 1 memiliki 8 anggota, *cluster* 2 memiliki 13 anggota dan untuk *cluster* 3 memiliki 17 anggota. Karakteristik yang diperoleh dari *cluster* optimal tersebut yakni 3 *cluster* memperoleh *cluster* 2 sebagai *cluster* yang memiliki nilai rata-rata tertinggi diikuti oleh *cluster* 3 yang memiliki nilai rata-rata tertinggi kedua, diikuti oleh *cluster* 1 yang memiliki nilai rata-rata yang paling kecil.

B. Saran

Penentuan *cluster* terbaik dapat menggunakan metode pengganti yang lain seperti metode *Dunn Index*, *Elbow*, dan *Silhouette*

6. REFERENSI

- Badan Pusat Statistik. 2015. Indikator Kesejahteraan rakyat 2015. *Cerebrovascular Diseases*, 170-175. <https://doi.org/10.1159/000490421>
- Badan Pusat Statistik. 2019. Provinsi Jawa Timur Dalam Angka: Badan Pusat Statistik. URL : <https://www.jatim.bps.go.id/publication/2019/08/16/f68b9b7ca53a7998bc81453/provinsi-jawa-timur-dalam-angka-2019.html>
- Bezdek, J, C. 1981. *Pattern Recognition with Fuzzy Objective Function Algorithm*. New York: Plenum Press.
- Ade, Yugo, D. 2020. *Algoritma Fuzzy C-Means dengan Metode Elbow untuk Mengelompokkan Provinsi di Indonesia Berdasarkan Jumlah Kasus HIV. Skripsi*. Jember: Universitas Muhammadiyah.
- Han, J., & Kamber, M. 2006. *Data Mining: Concept and Techniques, Second*

- Edition*. Waltham: Morgan Kaufmann Publishers.
- Han, J., Kamber, M., & Pei, J. 2014. Data mining: Data mining concepts and techniques. In *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*. <https://doi.org/10.1109/ICMIRA.2013.45>
- Holland, S. M. 2006. *Cluster Analysis*. United States: University of Georgia
- Kusumadewi, S., & Purnomo, H. 2004. *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*. Yogyakarta: Graha Ilmu.
- Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition*. In *Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition* (Vol. 9780470908). <https://doi.org/10.1002/9781118874059>
- Larose, Daniel, T. 2005. *Discovering Knowledge In Data: An Introduce to Data Mining*. Canada: JohnWiley and Sons, Inc.
- Luthfi, E. T. 2010. “Fuzzy C-Means Untuk Clustering Data (Studi Kasus: Data Performance Mengajar Dosen)”. *Seminar Nasional Teknologi, Samarinda*.
- Merliana, N. P. E., Ernawati, & Santoso, A. J. 2015. Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means Clustering. Yogyakarta: *Prosiding Seminar Nasional Multi Disiplin Ilmu & Call for Papers Unisbank (Sendi_U)*, 978-979.
- Nengsih, W., & Fadly, A. 2017. *Klasterisasi Tingkat Buta Huruf di Indonesia berbasis Point-Based K-Means Analysis*. Riau: Program Studi Sistem Informasi, *Politeknik Caltex Riau*.
- Putri, A. L. R., & Dwidayati, N. 2021. Analisa Perbandingan K-Means Dan Fuzzy C-Means Dalam Pengelompokan Daerah Penyebaran Covid-19 Indonesia. *UNNES Journal Of Mathematics*, 10(2), 4–7.
- <http://journal.unnes.ac.id/sju/index.php/ujme>
- Rahakbauw, D. L., Ilwaru, V. Y. I., & Hahury, M. H. (2017). *Implementation of Fuzzy C-Means Clustering in Scholarship Determination*. *Jurnal Ilmu Matematika Dan Terapan*, 11, 1–12.
- Ramadhan, A., Efendi, Z., & Mustakim. (2017). Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling. *Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (SNTIKI) 9*, 219–226.
- Taqwim, Setiawan, & Bachtiar. (2019). Analisis Segmentasi Pelanggan Dengan RFM Model Pada Pt . Arthamas Citra Mandiri Menggunakan Metode Fuzzy C-Means Clustering. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(2), 1986–1993.
- Wakhidah, N. (2014). *Clustering Menggunakan K-Means Algorithm (K-Means Algorithm Clustering)*. *Fakultas Teknologi Informasi*, 21(1), 70–80.