

STUDI KOMPARATIF ALGORITMA K-MEANS CLUSTERING DAN FUZZY C-MEANS PADA MULTIVARIATE FUZZY TIME SERIES

Tri Setya Darmawan¹

Institut Teknologi Bisnis dan Kesehatan Bhakti Putra Bangsa Indonesia

Ummu Wachidatul Latifah²

Institut Teknologi Bisnis dan Kesehatan Bhakti Putra Bangsa Indonesia

ummuwl2@gmail.com

Yuning Tyas A'aliyah Sari³

Institut Teknologi Bisnis dan Kesehatan Bhakti Putra Bangsa Indonesia

ABSTRAK

Prediksi merupakan suatu hal yang penting untuk mengetahui suatu langkah yang dapat dilakukan di masa yang akan datang. Prediksi deret waktu yang sering digunakan, yaitu *Fuzzy Time Series*. Model *Fuzzy Time Series* dapat digunakan untuk meramalkan satu faktor, sedangkan prediksi yang dipengaruhi oleh beberapa faktor dapat menggunakan *Multivariate Fuzzy Time Series*. Prediksi dengan banyak faktor dapat mempengaruhi beban komputasi dan hasil akurasi prediksi yang diperoleh. Tujuan penelitian ini, yaitu membandingkan algoritma K-Means *Clustering* dan Fuzzy C-Means. Hasil perbandingan algoritma *clustering* dimodelkan dengan *Multivariate Fuzzy Time Series*. Hasil penelitian menunjukkan bahwa berdasarkan evaluasi hasil *cluster* menunjukkan bahwa algoritma Fuzzy C-Means lebih baik dari algoritma K-Means. Berdasarkan nilai akurasi model *Multivariate Fuzzy Time Series* tergolong sangat akurat. Nilai MAPE dan RMSE yang diperoleh sebesar 0.4797% dan 0.84666.

Kata Kunci: K-Means, Fuzzy C-Means, Multivariate, Time Series, Clustering

ABSTRACT

Prediction is an important thing to know a step that can be taken in the future. Time series prediction is often used, namely Fuzzy Time Series. The Fuzzy Time Series model can be used to forecast one factor, while predictions influenced by several factors can use Multivariate Fuzzy Time Series. Predictions with many factors can affect the computational load and the accuracy of the predictions obtained. The purpose of this research is to compare K-Means Clustering and Fuzzy C-Means algorithms. The results of the clustering algorithm comparison are modeled with Multivariate Fuzzy Time Series. The results showed that based on the evaluation of cluster results, the Fuzzy C-Means algorithm is better than the K-Means algorithm. Based on the accuracy value, the

Multivariate Fuzzy Time Series model is classified as very accurate. The MAPE and RMSE values obtained are 0.4797% and 0.84666.

Keywords: *K-Means, Fuzzy C-Means, Multivariate, Time Series, Clustering*

1. PENDAHULUAN

Ilmu prediksi adalah teknik untuk memprediksi hasil tertentu untuk periode waktu mendatang dengan mempertimbangkan data masa lalu serta data saat ini. (Rahmah Binaiya et al., 2019). Prediksi yang dilakukan dengan mengumpulkan data masa lalu disebut deret waktu (*time series*). Data time series yang memiliki beberapa faktor atau variable disebut data *time series* multivariat. Prediksi data time series multivariat dapat dilakukan dengan metode *Multivariate Fuzzy Time Series* (MFTS).

Model MFTS merupakan pengembangan dari model FTS yang dapat digunakan untuk melakukan prediksi data time series yang terdiri dari beberapa variable (Fakhru Rosyad et al., 2020). Pada penelitian yang dilakukan oleh (SATRIA et al., 2015) mengatakan bahwa perbedaan FTS dan MFTS hanya terletak pada faktor yang digunakan dalam prediksi. FTS hanya menggunakan satu faktor, sedangkan MFTS mempertimbangkan lebih dari satu faktor, yaitu faktor utama dan faktor pendukung.

Beberapa hasil penelitian yang berkaitan dengan MFTS, yaitu penelitian (Li & Wei, 2020) menyatakan bahwa metode *Fuzzy clustering based on feature weights for multivariat time series* dapat meningkatkan akurasi pengelompokan untuk kumpulan data deret waktu multivariat. MFTS dianggap sebagai model yang efektif untuk menganalisis tipe data deret waktu (Tran et al., 2018). Penelitian lain menyatakan metode *multivariate high order fuzzy time series*, didapatkan bahwa faktor pendukung dan banyaknya orde dapat meningkatkan keakuratan prediksi (Kartini et al., 2019).

Penelitian (Tinh & Dieu, 2019) menyatakan bahwa model FTS partisi optimal lebih baik daripada model FTS. Namun, hal tersebut belum tentu berlaku apabila diterapkan pada data yang memiliki lebih dari satu variable. Oleh karena itu, pada penelitian ini akan dikembangkan model MFTS partisi optimal. Salah satu algoritma yang dapat menentukan interval kelas, yaitu algoritma *clustering*. Algoritma ini bekerja dengan melakukan pengoptimalan untuk menemukan *cluster* terbaik sehingga rentang interval yang diperoleh memiliki jarak optimal. Algoritma *clustering* yang akan digunakan pada penelitian ini, yaitu K-Means dan Fuzzy C-Means.

Algoritma K-Means merupakan algoritma berbasis partisi yang mencoba mempartisi data menjadi dua *cluster* atau lebih dengan menggunakan nilai rata-rata sebagai pusat *cluster* (Oktarina et al., 2020). Kelebihan algoritma K-Means adalah algoritma yang sederhana dan efisien, sedangkan kekurangan algoritma K-Means adalah harus menentukan parameter dan sensitive terhadap data asing (Kolay, 2017).

FCM adalah salah satu metode *clustering* yang merupakan bagian dari *Hard K-Means*. FCM menggunakan model pengelompokan fuzzy sehingga pengelompokan data dalam satu terbentuk berdasarkan derajat atau tingkat keanggotaan yang berbeda dengan rentang 0 hingga 1 (Rahakbauw et al., 2017). Kelebihan FCM adalah penempatan *cluster* yang lebih tepat dibandingkan dengan algoritma yang lain (Agustini, 2017). Pengelompokan menggunakan FCM memiliki beberapa outlier karena penentuan pusat *cluster* awal diambil secara acak.

Model MFTS akan diimplementasikan pada suatu data. Data yang digunakan dalam penelitian ini, yaitu data Indeks Pembangunan (IPM) Jawa Tengah. IPM sangat penting dalam menilai kemampuan suatu negara untuk meningkatkan populasi dan kualitas hidupnya di wilayah atau negara tertentu, seperti Indonesia. (Kirana et al., 2019). Tinggi rendahnya IPM disebabkan oleh beberapa komponen. Menurut Badan Pusat Statistika (BPS) dan *United Nations Development Program* (UNDP) menyatakan bahwa IPM memiliki tiga komponen utama, yaitu lamanya hidup, diukur berdasarkan harapan hidup pada saat lahir, lama hidup yang diukur dengan harapan lahir, tingkat pendidikan yang diukur dengan tingkat melek huruf orang dewasa, dan standar hidup yang dinilai dengan pengeluaran per kapita (Utami, 2020).

Provinsi Jawa Tengah merupakan salah satu provinsi dengan jumlah penduduk terbanyak ke-3, yaitu dengan jumlah 36,52 juta jiwa (BPS Jateng, 2021). Jumlah penduduk yang besar dapat menjadi pendorong pembangunan jika didukung oleh kualitas dan kemampuan penduduk yang tinggi. Namun Provinsi Jawa Tengah menempati urutan ke-13 dari 34 provinsi dengan nilai 71,92. Meskipun berada pada kategori tinggi (70-80), IPM Provinsi Jawa Tengah masih berada pada nilai ambang batas, yaitu mendekati kategori sedang (70) (Rizqi & Khotimah, 2021). Hal tersebut perlu mendapatkan kepedulian dari pemerintah terhadap pembangunan manusia karena kurang mampu meminimalisir kesenjangan ekonomi, sosial dan kemiskinan.

Salah satu upaya yang dapat dilakukan, yaitu perlu dilakukan prediksi tingkat Laporan IPM dari tahun-tahun sebelumnya, agar pemerintah Indonesia memiliki acuan dan strategi yang jelas untuk diikuti agar pemerintah dapat meningkatkan IPM di Jawa Tengah.

2. TINJAUAN PUSTAKA

2.1 Prediksi

Prediksi adalah suatu proses meramalkan apa yang akan terjadi di masa depan dengan menggunakan suatu metode tertentu (Safitri et al., 2018). Ada dua metode prediksi, yaitu: (Aswi, 2006)

- 1) Metode Kualitatif merupakan prediksi berdasarkan pendapat suatu pihak dan data tidak dapat direpresentasikan secara tegas menjadi suatu angka/nilai.
- 2) Metode Kuantitatif merupakan metode prediksi berdasarkan yang dipresentasikan dalam bentuk angka. Menurut Makridakis, dkk (1999), metode kuantitatif terbagi menjadi 2, yaitu metode *time series* dan metode kasual. Metode *time series* merupakan metode prediksi berdasarkan analisis pola hubungan antara variabel yang diperkirakan dengan variasi runtun waktu (*time series*). Beberapa metode *time series* yang berkembang, yaitu *Fuzzy Time Series*, ARIMA (*Autoregressive Integrate Moving Average*), *Moving Average*, *Exponential Smoothing* dan *Time Series Regresion* (Fauziah et al., 2016). Metode kasual yaitu metode prediksi berdasarkan analisa pola hubungan antara variabel. Metode ini mengasumsikan bahwa faktor yang diprediksi menunjukkan suatu hubungan sebab akibat dengan satu atau lebih variabel bebas (Makridakis, S, Wheelright, S, C, dan McGee, 1999).

2.2 Data Time Series

Data time series merupakan data yang dikumpulkan dari waktu ke waktu, data dapat berupa periode waktu harian, mingguan, bulanan, tahunan, atau periode waktu tentu lainnya dalam rentang waktu yang sama (Makridakis, S, Wheelright, S, C, dan McGee, 1999; Sumartini et al., 2017). Analisis time series adalah salah satu metode statistika yang digunakan untuk meramalkan keadaan yang akan terjadi di masa yang akan datang dalam rangka pengambilan keputusan (Aswi, 2006).

2.3 Himpunan Crisp

Himpunan crisp atau himpunan tegas merupakan himpunan yang membagi sekelompok individu menjadi dua kelompok, yaitu anggota dan bukan anggota (Haqi, 2011). Himpunan ini disimbolkan dengan huruf besar (A, B, P , dan lain-lain) dan anggota (elemen) himpunan disimbolkan dengan huruf kecil (a, b, x, y , dan lain-lain). Himpunan crisp hanya memiliki 2 nilai keanggotaan (μ), yaitu satu (1), menunjukkan bahwa suatu item adalah anggota suatu himpunan, atau; (0), yang berarti bahwa suatu item tidak menjadi anggota dalam suatu himpunan (S. Kusumadewi & Purnomo, 2010).

Definisi 2.1 (Davvaz et al., 2021)

Misalkan U merupakan himpunan semesta dan $A \subseteq U$. Maka

$$\mu_A(u) = \begin{cases} 1, & \text{jika } u \in A \\ 0, & \text{jika } u \notin A \end{cases} \quad (1)$$

2.4 Logika Fuzzy

Menurut (S. dan P. H. Kusumadewi, 2010; Wardani et al., 2017) fuzzy terdiri dua atribut, yaitu linguistik dan numerik. Atribut linguistik digunakan untuk penamaan suatu grup yang mewakili suatu keadaan atau kondisi tertentu dengan menggunakan bahasa alami, seperti muda, tua, sedangkan atribut numerik suatu nilai yang menunjukkan ukuran dari suatu variabel

Pada tahun 1965 L.A Zadeh mengenalkan himpunan fuzzy sebagai perluasan himpunan klasik. Himpunan fuzzy merupakan dasar logika fuzzy. Pada teori himpunan fuzzy, peranan derajat keanggotaan atau nilai keanggotaan digunakan untuk menentukan posisi elemen dalam suatu himpunan. Himpunan tegas hanya memiliki dua kemungkinan, yaitu 0 dan 1, sedangkan pada himpunan fuzzy, nilai keanggotaan berkisar dari 0 hingga 1.

Apabila x memiliki nilai keanggotaan fuzzy $\mu_A[x] = 0$, berarti x tidak menjadi anggota himpunan A , demikian pula apabila x memiliki nilai keanggotaan fuzzy $\mu_A[x] = 1$, berarti menjadi anggota penuh pada himpunan A (S. dan P. H. Kusumadewi, 2010; Wardani et al., 2017).

Definisi 2.2 (J.ROSS, 2010)

Himpunan fuzzy \tilde{A} pada semesta pembicaraan U dapat didefinisikan sebagai sebuah himpunan pasangan berurut:

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) | x \in U\} \quad (2)$$

dengan $\mu_{\tilde{A}}(x)$ adalah fungsi keanggotaan atau karakteristik himpunan A . Simbol $\mu_{\tilde{A}}(x)$ menunjukkan derajat keanggotaan $\mu_{\tilde{A}}(x)$ di \tilde{A} yang merupakan derajat keanggotaan x pada himpunan \tilde{A} . Fungsi keanggotaan $\mu_{\tilde{A}}(x)$ memetakan U ke ruang keanggotaan $\mu = [0,1]$.

2.5 Multivariate Fuzzy Time Series

MFTS merupakan pengembangan dari metode FTS dengan menambahkan beberapa variabel yang dapat mempengaruhi prediksi (Fakhru Rosyad et al., 2020). Perbedaan utama antara FTS dan MFTS terletak pada faktor yang digunakan dalam pembentukan model prediksi. Pada FTS faktor yang dipertimbangkan hanya satu, sedangkan MFTS mempertimbangkan lebih dari satu faktor, satu sebagai faktor utama dan lainnya merupakan faktor pendukung. Berikut ini langkah-langkah MFTS (SATRIA et al., 2015).

1. Menentukan faktor utama dan faktor pendukung.
2. Definisikan himpunan semesta G dari faktor utama dengan persamaan berikut:

$$G = [D_{min}, D_{max}] \quad (3)$$

dimana D_{min} dan D_{max} merupakan nilai minimum dan maksimum data historis faktor utama, sedangkan D_1 dan D_2 merupakan bilangan real positif sembarang untuk membagi himpunan semesta menjadi beberapa interval g_1, g_2, \dots, g_i dengan panjang yang sama. Kemudian, dengan cara yang sama definisikan himpunan semesta faktor pendukung dengan $V_i, i = 1, 2, 3, \dots, m - 1$ dengan

$$V_i = [(E_i)_{min} - E_{i1}, (E_i)_{max} - E_{i2}],$$

dimana

$$(E_i)_{min} = (E_1)_{min}, (E_2)_{min}, \dots, (E_m)_{min}$$

dan $(E_i)_{max}, (E_1)_{max}, (E_2)_{max}, \dots, (E_m)_{max}$ adalah nilai minimum dan maksimum pada setiap data faktor pendukung. E_{i1} dan E_{i2} adalah vektor bilangan positif untuk membagi setiap bilangan universal dalam faktor pendukung $V_i, i = 1, 2, 3, \dots, m - 1$ ke dalam interval yang sama, selanjutnya disebut sebagai $v_{1,l}, v_{2,l}, \dots, v_{m-1,l}$, $l = 1, 2, \dots, p$ dengan $v_{1,1}, v_{1,2}, \dots, v_{1,p}$ mewakili n interval dengan panjang yang sama dari himpunan semesta v_1 untuk faktor pendukung deret waktu fuzzy time series.

3. Definisikan nilai linguistik A_i , yang didefinisikan oleh himpunan *fuzzy* dari faktor utama dan nilai linguistik dari faktor pendukung yang direpresentasikan dengan B_{ij} , dimana $i = 1, 2, \dots, m - 1, j = 1, 2, \dots, n$. Didefinisikan oleh himpunan *fuzzy* dari faktor-faktor utama dengan aturan di bawah ini:

(4)

$$\begin{aligned}
 A_1 &= \frac{1}{g_1} + \frac{0.5}{g_2} + \frac{0}{g_3} + \frac{0}{g_4} + \cdots + \frac{0}{g_{l-2}} + \frac{0}{g_{l-1}} + \frac{0}{g_l} \\
 A_2 &= \frac{0.5}{g_1} + \frac{1}{g_2} + \frac{0.5}{g_3} + \frac{0}{g_4} + \cdots + \frac{0}{g_{l-2}} + \frac{0}{g_{l-1}} + \frac{0}{g_l} \\
 A_3 &= \frac{0}{g_1} + \frac{0}{g_2} + \frac{0}{g_3} + \frac{0}{g_4} + \cdots + \frac{0}{g_{l-2}} + \frac{0.5}{g_{l-1}} + \frac{1}{g_l} \\
 &\vdots \\
 A_m &= \frac{0}{g_1} + \frac{0}{g_2} + \frac{0}{g_3} + \frac{0}{g_4} + \cdots + \frac{0}{g_{l-2}} + \frac{0.5}{g_{l-1}} + \frac{1}{g_l}
 \end{aligned}$$

4. Fuzzifikasiakan data historis dari faktor utama dan faktor pendukung.

a) Temukan interval $g_l, l = 1, 2, \dots, p$, yang mana terdapat nilai dari faktor utama pada data historis, dengan beberapa kondisi berikut:

Kondisi 1. Jika nilai dari faktor utama terdapat pada interval u_1 , maka nilai dari faktor utama tersebut difuzzifikasiakan ke dalam $\frac{1}{A_1} + \frac{0.5}{A_2} + \frac{0.0}{A_3}$, yang dinotasikan dengan X_i

Kondisi 2. Jika nilai dari faktor utama terdapat pada $u_l, l = 2, 3, \dots, p-1$ maka nilai faktor utama difuzzifikasi menjadi $\frac{0.5}{A_{l-1}} + \frac{1}{A_l} + \frac{0.5}{A_{l+1}}$

Kondisi 3. Jika nilai dari faktor utama terdapat pada interval u_p maka nilai faktor utama difuzzifikasi menjadi $\frac{0}{A_{i-2}} + \frac{0}{n-1} + \frac{1}{A_n}$, dinotasikan oleh X_n .

b) Untuk faktor pendukung, temukan interval $v_{i,l}$, yang mana terdapat nilai dari faktor utama pada data historis, dengan beberapa kondisi berikut:

Kondisi 1. Jika nilai dari faktor pendukung terdapat pada interval $v_{i,l}$, maka nilai dari faktor utama tersebut difuzzifikasiakan ke dalam $\frac{1}{B_{i,1}} + \frac{0.5}{B_{i,2}} + \frac{0.0}{B_{i,3}}$, yang dinotasikan dengan $Y_{i,1} = Y_{1,1}, Y_{2,1}, \dots, Y_{m-1,1}$

Kondisi 2. Jika nilai dari faktor pendukung ke- i terdapat pada interval $v_{i,l}, l = 2, 3, \dots, p-1$ maka nilai dari faktor pendukung tersebut difuzzifikasiakan ke dalam $\frac{0.5}{B_{i,j-1}} + \frac{1}{B_{i,j}} + \frac{0.5}{B_{i,j+1}}$ dimana $j = i = 2, 3, \dots, n-1$ yang dinotasikan dengan $Y_{i,j}, j = 2, 3, \dots, n-1$.

Kondisi 3. Jika nilai dari faktor pendukung ke- i terdapat pada interval $v_{i,p}$ maka nilai dari faktor pendukung tersebut difuzzifikasiakan ke dalam $\frac{0}{B_{i,n-2}} + \frac{0.5}{B_{i,n-1}} + \frac{1}{B_n}$ yang dinotasikan dengan $Y_{i,n}$

5. Menentukan *Fuzzy Logic Relationship* (FLR)

6. Menentukan *Fuzzy Logic Relationship Group* (FLRG)

7. Melakukan defuzzyifikasi data fuzzy

3. METODOLOGI PENELITIAN

3.1 Desain Penelitian

Penelitian ini merupakan penelitian yang dilakukan dengan menggunakan metode prediksi MFTS. Pada penelitian ini pendekatan yang digunakan adalah pendekatan kuantitatif. Pendekatan kuantitatif merupakan pendekatan yang digunakan pada suatu penelitian dimana data yang akan dianalisis berupa angka (numerik). Selain itu, penelitian ini juga akan lebih banyak menggunakan tabel ataupun grafik untuk menampilkan hasil analisis data.

3.2 Sumber Data dan Subjek Penelitian

Sumber data yang digunakan pada penelitian ini adalah data sekunder. Data sekunder adalah data yang digunakan pada penelitian tidak diambil melalui penelitian secara langsung. Karakteristik data yang dapat diterapkan dengan MFTS adalah data yang memiliki lebih dari satu variable atau data yang dipengaruhi oleh beberapa faktor. Data merupakan data numerik dan data *unsupervised* (non klasifikasi), seperti data IPM. Data IPM yang digunakan terdiri dari beberapa faktor, yaitu faktor pendukung dan faktor utama. Faktor pendukung terdiri dari Angka Harapan Hidup (AHH) (X_1), Rata-rata Lama Sekolah (RLS) (X_2) dan faktor utama terdiri dari nilai IPM (Y). Subjek dari penelitian ini merupakan penduduk Jawa Tengah.

3.3 Teknik Dan Instrumen Pengumpulan Data

Teknik pengumpulan data merupakan metode atau cara yang digunakan untuk mengambil data. Pada penelitian ini teknik yang digunakan yaitu dengan melakukan studi kepustakaan. Data diambil dari salah satu data base yang ada pada internet yaitu Badan Pusat Statistika (BPS) Jawa Tengah.

3.4 Teknik Analisis Data

Teknik analisis data dilakukan dengan bantuan software yaitu python. Langkah-langkah analisis data sebagai berikut:

1. Data yang digunakan di lakukan preprocessing terlebih dahulu.
2. Melakukan *clustering* data dengan algoritma K-Means dan FCM.
3. Mengevaluasi hasil *cluster* dari setiap algoritma dengan menggunakan rasio simpangan baku.
4. Melakukan prediksi dengan MFTS.
5. Mengevaluasi hasil prediksi dengan MAPE dan RMSE.

4. HASIL DAN PEMBAHASAN

4.1 Data Time Series

Karakteristik data yang digunakan dalam penelitian ini berupa data numerik dan data *unsupervised*, bukan termasuk data klasifikasi. Data diperoleh dari BPS Jawa Tengah dari tahun 2010-2021. Adapun variabel-variabel yang akan digunakan, yaitu

X_1 : Angka Harapan Hidup (AHH)

X_2 : Angka Rata-rata Lama Sekolah (RLS)

Y : Angka Indeks Pembangunan Manusia (IPM).

Data IPM Provinsi Jawa Tengah 1999-2021 dapat dilihat pada tabel berikut ini.

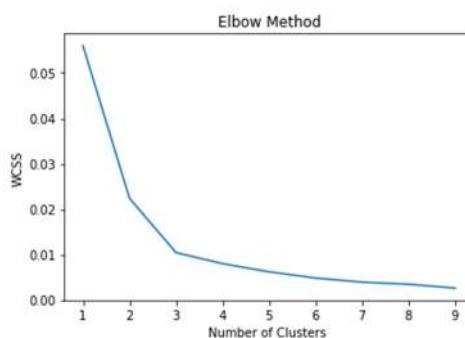
Tabel 1. Data IPM Provinsi Jawa Tengah 1999-2021

No	Tahun	AHH	RLS	IPM
1	1999	68.3	6.03	64.6
2	2000	62.67	6.14	61.43
3	2001	62.67	6.14	61.43
...
23	2021	74.47	7.75	72.16

4.2 Metode Partisi dengan Algoritma *Clustering*

4.2.1 Penentuan Jumlah *Cluster*

Jumlah *cluster* optimal dapat diperoleh menggunakan metode Elbow. Hasil dari metode Elbow yang dihitung berdasarkan persamaan (5) dapat dilihat pada Gambar 1 berikut ini.



Gambar 1. Jumlah *Cluster* Optimal Metode Elbow

Berdasarkan Gambar 1 terlihat bahwa grafik membentuk siku di jumlah *cluster* 3, hal tersebut menunjukkan bahwa jumlah yang optimal adalah 3 *cluster*.

4.2.2 Algoritma K-Means

Metode K-Means merupakan metode *clustering* yang paling sederhana. Berikut merupakan hasil *clustering* data IPM berdasarkan algoritma K-Means menggunakan Python.

1. Menentukan jumlah *cluster*. Berdasarkan metode Elbow, diperoleh jumlah *cluster* optimal, yaitu 3 dengan iterasi maksimum 10.
2. Menentukan nilai *centroid* (titik pusat). Nilai *centroid* awal dipilih secara acak. Nilai *centroid* awal yang dipilih berdasarkan Tabel 4, yaitu data ke- 4 sebagai *centroid cluster* 0, data ke-15 sebagai *centroid cluster* 1 dan data ke-20 sebagai *centroid cluster* 2. Berikut ini merupakan nilai *centroid* awal.

Tabel 2. Nilai Centroid Awal

Cluster	X ₁	X ₂	Y
0	0,20253	0.19468	0.19878
1	0.21156	0.22253	0.22202
2	0.21806	0.22013	0.21323

3. Menghitung jarak antara titik *centroid* dengan titik tiap objek menggunakan *Euclidean Distance* dengan persamaan (7). Berikut hasil perhitungan jarak berdasarkan persamaan (7).
4. Realokasi data ke masing-masing *cluster* berdasarkan perbandingan jarak antara data dengan *centroid* masing-masing *cluster*. Untuk menentukan anggota *cluster*

adalah dengan memperhitungkan jarak minimum objek. Nilai yang diperoleh dalam keanggotaan data pada jarak matriks adalah 0 atau 1, dimana nilai 1 untuk data yang dialokasikan ke *cluster* dan nilai 0 untuk data yang dialokasikan ke *cluster* yang lain. Tabel hasil *cluster* pada iterasi 1 sebagai berikut.

Tabel 3. Hasil Iterasi-1 K-Means

No	Tahun	AHH	RLS	IPM	Cluster
1	1999	68,3	6.03	64.6	0
2	2000	62.67	6.14	61.43	0
3	2001	62.67	6.14	61.43	0
...
23	2021	74.47	7.75	72.16	2

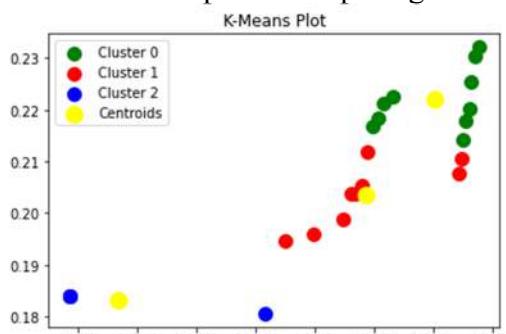
5. Menentukan nilai *centroid* pada iterasi berikutnya menggunakan persamaan (6). Perhitungan nilai *centroid* pada iterasi selanjutnya sebagai berikut.

Untuk *cluster* 0 variabel X_1 , diperoleh bahwa jumlah anggota *cluster* 0 sebanyak 9 data, yaitu data ke-1 sampai data ke-9. Berikut hasil yang diperoleh.

Tabel 3 Nilai Centroid Iterasi 2

Cluster	X_1	X_2	Y
0	0.19831	0.19211	0.19823
1	0.21064	0.21976	0.21950
2	0.21622	0.21753	0.21268

6. Apabila terjadi perubahan nilai *centroid* atau jumlah iterasi kurang dari jumlah iterasi maksimum maka lakukan langkah ke 3, jika tidak maka iterasi berhenti. Visualiasi plot hasil *cluster* dapat dilihat pada gambar berikut ini.



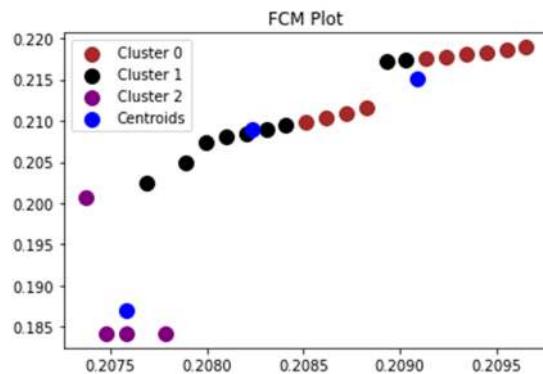
Gambar 2. Plot K-Means

Pada Gambar 2 terlihat bahwa *cluster* 0 merupakan *cluster* yang memiliki anggota dengan nilai tertinggi, *cluster* 1 berada di posisi kedua dan *cluster* 2 di posisi terakhir. Anggota *cluster* yang memiliki nilai tertinggi merupakan anggota yang memiliki nilai IPM terbaik dan *cluster* dengan nilai terendah merupakan *cluster* yang memiliki nilai IPM terrendah.

4.2.3 Algoritma Fuzzy C-Means

Berikut merupakan hasil *clustering* data IPM berdasarkan algoritma Fuzzy C-Means.

- 1) Menginput data yang akan di *cluster* X berupa matriks ukuran $n \times m = 23 \times 3$ (n merupakan jumlah sampel data dan m atribut setiap data. X_{ij} merupakan data sample ke- i , ($i = 1,2,3, \dots, n$), atribut ke- j , ($j = 1,2,3, \dots, m$).
- 2) Menentukan:
 - a. Jumlah *cluster* = $c = 3$
 - b. Pangkat = $w = 2$
 - c. Maksimum Iterasi = $MaxIter = 100$
 - d. Error terkecil yang diharapkan = $\xi = 0,001$
 - e. Fungsi objektif, $P_0 = 0$
 - f. Iterasi awal, $t = 1$
- 3) Menentukan bilangan random μ_{ik} , $i = 1,2,3, \dots, n$; $k = 1,2,3, \dots, c$; sebagai elemen-elemen matriks partisi awal U . Berdasarkan persamaan (9) diperoleh matriks partisi awal U .
- 4) Menghitung pusat *cluster* ke- k : V_{kj} dengan $k = 1,2,3, \dots, c$ dan $j = 1,2,3, \dots, m$. Berdasarkan persamaan (14), pada iterasi ke-1 diperoleh hasil sebagai berikut.
- 5) Menghitung fungsi objektif pada iterasi ke- t , P_t menggunakan persamaan.
- 6) Menghitung perubahan matriks partisi



Gambar 3. Plot FCM Clustering

Berdasarkan Gambar 5, terlihat bahwa *cluster* 0 disimbolkan dengan warna merah, *cluster* 1 dengan warna hitam, *cluster* 2 berwarna ungu dan *centroid* yang berwarna biru. Kedudukan *cluster* 0 lebih tinggi dibandingkan *cluster* 1, hal tersebut berarti nilai IPM masing-masing anggota *cluster* 0 lebih tinggi atau lebih baik dibandingkan dengan *cluster*

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan yang diperoleh dari baba sebelumnya, dapat disimpulkan bahwa.

1. Algoritma *clustering* yang memiliki hasil terbaik adalah algoritma Fuzzy C-Means. Fuzzy C-Means lebih baik dibandingkan dengan algoritma K-Means dan Fuzzy Subtractive *Clustering* dalam mengelompokan data IPM Provinsi Jawa Tengah. Hal tersebut dikarenakan Fuzzy C-Means memiliki nilai simpangan baku terkecil, yaitu 0.42628.
2. Hasil prediksi metode MFTS ditunjukkan dengan menggunakan *Root Mean Square Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE). Hasil prediksi model MFTS memiliki nilai MAPE 0.4797% dan nilai RMSE 0.8466.

5.2 Saran

Berdasarkan hasil analisis serta kesimpulan yang sudah diperoleh, maka penelitian selanjutnya dapat diterapkan pada data yang memiliki faktor pendukung lebih banyak. Untuk memperbaiki kelemahan yang ada pada penelitian ini dapat diterapkan algoritma clustering lain, seperti Fuzzy K-Medoids dan dapat dioptimasi dengan menerapkan algoritma optimasi pada metode prediksi.

DAFTAR PUSTAKA

Agustini, F. (2017). Implementasi Algoritma Fuzzy C-Means Studi Kasus. *Jurnal Ilmu Pengetahuan Dan Teknologi Komputer*, 3(1), 127–132.

Aswi, dan S. (2006). *Analisis Deret Waktu*. Andira Publisher.

BPS Jateng. (2021).

Bunkers, M.J., & dkk. (n.d.). Definition of Climate Region In The Nothern Plains Using an Objactive Cluster Modifation Technique. *Journal Climate*, Vol 9.

Chen, S. M. (1996). Forecasting enrollments based on fuzzy time series. *Fuzzy Sets and Systems*, 81(3), 311.

Davvaz, B., Mukhlash, I., & Soleha, S. (2021). Himpunan Fuzzy dan Rough Sets. *Limits: Journal of Mathematics and Its Applications*, 18(1), 79. <https://doi.org/10.12962/limits.v18i1.7705>

Faisal, M., Zamzami, E. M., & Sutarman. (2020). Comparative Analysis of Inter-Centroid K-Means Performance using Euclidean Distance, Canberra Distance and Manhattan Distance. *Journal of Physics: Conference Series*, 1566(1). <https://doi.org/10.1088/1742-6596/1566/1/012112>

Fakhru Rosyad, A., Farikhin, & Suseno, J. E. (2020). Information Systems of Forecasting Incidence Rates of Dengue Fever Disease Using Multivariate Fuzzy Time Series. *E3S Web of Conferences*, 202. <https://doi.org/10.1051/e3sconf/202020214005>

Fauziah, N., Wahyuningsih, S., & Nasution, Y. N. (2016). Peramalan Menggunakan Fuzzy Time Series Chen (Studi Kasus : Curah Hujan Kota Samarinda). *Statistika*, 4(2), 52–61.

Haqi, B. (2011). *Model penduga jumlah produksi pil kb trinordiol*-28: studi kasus pt.sunthi sepuri*. 4(1), 37–47.

Irawan, J., Handayani, A. A. A. T., & Zohri, L. H. N. (2021). Operasionalisasi IBM SPSS 21 untuk Meningkatkan Kemampuan dan Keterampilan Olah Data Penelitian Mahasiswa. *Jurnal Pengabdian Magister Pendidikan IPA*, 4(2). <https://doi.org/10.29303/jpmi.v4i2.660>

J.ROSS, T. (2010). *Fuzzy Logic With Engineering Application*.

Junaidi, J. (2014). Deskripsi Data Melalui Box-Plot. *Statistika Deskriptif Dengan Microsoft Excel*, 6, 1–5. <https://doi.org/10.13140/RG.2.1.1057.8088>

Kartini, D., Rusdiani, H., & Farmadi, A. (2019). Analisis Pengaruh Banyak Orde pada Metode Multivariate High-Order Fuzzy Time Series untuk Prediksi Duga Muka Air Waduk. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 5(1), 9. <https://doi.org/10.26418/jp.v5i1.27265>

Khang, T. D., Vuong, N. D., Tran, M. K., & Fowler, M. (2020). Fuzzy C-means clustering algorithm with multiple fuzzification coefficients. *Algorithms*, 13(13), 1–11. <https://doi.org/10.3390/A13070158>

Kirana, I. O., Nasution, Z. M., & Wanto, A. (2019). Proyeksi Indeks Pembangunan Manusia Di Indonesia Menggunakan Metode Statistical Parabolic Dalam Menyongsong Revolusi Industri 4.0. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 16(2), 202. <https://doi.org/10.23887/jptk-undiksha.v16i2.18178>

Kolay, S. (2017). *K+ Means : An Enhancement Over K-Means Clustering Algorithm*. 1–7.

Kurniawan, I. D., Mhm, A., & Legowo, S. J. (2016). *STUDI PERBANDINGAN ESTIMASI KESALAHAN PADA LINEAR MODEL DAN INSTANTANEOUS MODEL DALAM MENGESTIMASI WAKTU PERJALANAN BERBASIS KECEPATAN SESAAT (LOKASI STUDI : RING ROAD UTARA SURAKARTA)*. 1264–1272.

Kusumadewi, S., & Purnomo, H. (2010). *Applikasi logika fuzzy untuk pendukung keputusan*. Graha Ilmu.

Kusumadewi, Sri. dan P. H. (2010). *Artificial Intelligence: Teknik dan Aplikasinya*. Graha Ilmu.

Li, H., & Wei, M. (2020). Fuzzy clustering based on feature weights for multivariate time series. *Knowledge-Based Systems*, 197, 105907. <https://doi.org/10.1016/j.knosys.2020.105907>

Makkulau, Linuwuh, S., Purhadi, & Mashuri, M. (2011). *Pendeteksian Outlier pada Pengamatan dalam Model Linear Multivariat dengan Metode Likelihood Displacement Statistic-Lagrange*. Christensen 1991, 62–68.

Makridakis, S, Wheelright, S, C, dan McGee, VE. (1999). *Metode dan Aplikasi Peramalan* Edisi 2 (2nd ed.). Binarupa Aksara.

Marutho, D., Hendra Handaka, S., Wijaya, E., & Muljono. (2018). The Determination of Cluster Number at k-Mean Using Elbow Method and Purity Evaluation on Headline News. *Proceedings - 2018 International Seminar on Application for Technology of Information and Communication: Creative Technology for Human Life, ISemantic 2018*, 533–538. <https://doi.org/10.1109/ISEMANTIC.2018.8549751>

Nainggolan, R., Perangin-Angin, R., Simarmata, E., & Tarigan, A. F. (2019). Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE)

optimized by using the Elbow Method. *Journal of Physics: Conference Series*, 1361(1). <https://doi.org/10.1088/1742-6596/1361/1/012015>

Nasution, D. A., Khotimah, H. H., & Chamidah, N. (2019). Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN. *Computer Engineering, Science and System Journal*, 4(1), 78. <https://doi.org/10.24114/cess.v4i1.11458>

Ni Luh Wiwik Sri Rahayu Ginantra, I. B. G. A. (2021). Penerapan Metode Single Exponential Smoothing dalam Peramalan Penjualan Benang. *Smart Comp: Jurnalnya Orang Pintar Komputer*, 10(3), 154–159. <https://doi.org/10.30591/smartcomp.v10i3.2887>

Ningrat, D. R., Maruddani, D. A. I., & Wuryandari, T. (2016). Analisis cluster dengan algoritma K-Means dan Fuzzy C-Means clustering untuk pengelompokan data obligasi korporasi. *None*, 5(4), 641–650.

Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika: Jurnal Pengembangan IT*, 4(1), 20–24. <https://doi.org/10.30591/jpit.v4i1.1253>

Nuryadi, Astuti, T. D., Utami, E. S., & Budiantara, M. (2017). *Buku ajar dasar-dasar statistik penelitian*.