

# **METODE *HIGH ORDER FUZZY TIME SERIES MULTI FACTORS* DENGAN ALGORITMA *FUZZY C-MEANS***

**(*HIGH ORDER FUZZY TIME SERIES MULTI FACTORS METHOD WITH C-MEANS FUZZY ALGORITHM*)**

**Yuni Wulandari<sup>1\*)</sup>, Dewi Retno Sari Saputro<sup>2)</sup>**

<sup>1,2)</sup>Program Studi Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas  
Sebelas Maret, Jl. Ir Sutami No. 36A, Surakarta 57126, Indonesia  
e-mail: yuniwld14@student.uns.ac.id\* dewiretnoss@staff.uns.ac.id

\*)penulis korespondensi

**Abstract.** Clustering is the process of grouping data into several clusters so that the data in a cluster has a high degree of similarity between data with one another but is very different from the data in other clusters. Fuzzy clustering is a technique to determine the optimal cluster in a vector space based on the Euclidian normal form for the distance between vectors. Fuzzy clustering is very useful for fuzzy modeling, especially in identifying fuzzy rules. There are various kinds of fuzzy clustering techniques, one of which is Fuzzy Cluster-Means (FCM). Fuzzy C-Means clustering is a data clustering technique in which the existence of each data point in a cluster is determined by the degree of membership. The purpose of this study is to examine the High Order Fuzzy Time Series Multi Factors method with Fuzzy C-Means in order to get  $k$  locations of the data cluster center points as many as  $k$  which are then used to form subintervals. The results show that Fuzzy C-Means replaces the process in the High Order Fuzzy Time Series Multi Factors method, which is when the subinterval is formed.

**Keywords:** *Clustering, Fuzzy C-Means, Metode High Order Fuzzy Time Series Multi Factors.*

## **1. PENDAHULUAN**

Informasi merupakan suatu elemen penting dalam kebutuhan kehidupan sehari-hari. Untuk mendapatkan sebuah informasi yang penting dan akurat sering kali tidak mudah kita dapatkan. Informasi yang tersedia dalam jumlah yang besar terkadang masih harus kita gali terlebih dahulu supaya informasi tersebut dapat menyajikan informasi yang tepat dan sesuai dengan kebutuhan. Untuk dapat menggali informasi yang berpotensi dari gudang data tidak bisa hanya mengandalkan data operasional saja, namun diperlukan suatu analisis data yang tepat sehingga menghasilkan informasi yang lebih berharga sehingga dapat menunjang kegiatan operasional. Untuk membantu mempermudah pengambil keputusan dalam menganalisis dan mengekstraksi data maka lahirlah cabang ilmu baru yang disebut Data Mining (Asriningtias dan Mardhiyah [1]). Data mining adalah penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar (Santoso dkk. [11]).

*Clustering* adalah metode yang digunakan dalam data mining dengan cara kerjanya mencari data dan mengelompokkan data yang mempunyai kemiripan karakteristik antara

data satu dengan data lainnya yang telah diperoleh (Sihombing [13]). Pengembangan metode clustering menghasilkan pendekatan berbasis *fuzzy*. Terdapat berbagai macam teknik *fuzzy clustering* salah satunya yaitu *Fuzzy Cluster-Means* (FCM). *Fuzzy C-Means* menerapkan pengelompokan *fuzzy*, dimana setiap data dapat menjadi anggota dari beberapa *cluster* dengan derajat keanggotaan yang berbeda-beda pada setiap *cluster*.

Metode *High Order Fuzzy Time Series Multi Factors* dibuat untuk mengatasi permasalahan pada peramalan *fuzzy time series* yang hanya bisa meramalkan suatu keadaan *time series* hanya berdasarkan satu faktor yang dipertimbangkan. Dengan Metode *high order fuzzy time series multi factors* suatu keadaan yang mempunyai banyak faktor yang menjadi penyebab kejadian dapat diramalkan dengan lebih akurat dikarenakan ada faktor-faktor pendukung yang memperkuat terjadinya suatu kejadian. Pada metode ini pembentukan subinterval yang berfungsi dalam pendefinisian *fuzzy set* dilakukan dengan menggunakan teknik *Fuzzy C-means clustering*. *Fuzzy C-Means* merupakan algoritma iteratif, yang menerapkan iterasi pada proses *clustering* data. *Fuzzy C-Means* digunakan untuk *clustering* data guna mendapatkan  $k$  buah letak titik pusat *cluster* data yang kemudian dijadikan untuk pembentukan subinterval (Wijaya dkk. [15]). Dengan demikian penelitian ini dikaji ulang mengenai metode *high order fuzzy time series multi factors* dengan *Fuzzy C-Means*.

## 2. Metodologi

Metode penelitian yang digunakan dalam menyusun penelitian ini yakni dengan mempelajari kajian teori dari artikel, buku, dan jurnal terkait metode *High Order Fuzzy Time Series Multi Factors* dan *Fuzzy C-Means*. Langkah-langkah yang dilakukan dalam penelitian ini dengan mengkaji ulang metode *High Order Fuzzy Time Series Multi Factors* dan *Fuzzy C-Means*

## 3. Hasil dan Pembahasan

### 3.1 Logika *Fuzzy*

Teori *fuzzy* pertama kali diperkenalkan oleh Dr. Lotfi Zadeh pada tahun 1965 dari University of California untuk mengembangkan konsep kualitatif yang tidak memiliki batas yang pasti, misalnya tidak ada nilai yang jelas atau pasti yang merupakan batas antara normal dan rendah, normal atau tinggi (Saepullah [9]). Menurut Sitohang dan Denson [14] logika *fuzzy* didefinisikan sebagai suatu jenis logika yang bernilai ganda dan berhubungan dengan ketidakpastian dan kebenaran parsial. Logika *fuzzy* nilai keanggotaan terletak pada rentang 0 sampai 1. Artinya, bisa saja suatu keadaan mempunyai dua nilai “Ya dan Tidak”, “Benar dan Salah” secara bersamaan, namun besar

nilainya tergantung pada bobot keanggotaan yang dimilikinya (Dorteus [4])

### 3.2 Himpunan *Fuzzy*

Himpunan *fuzzy* adalah pengelompokan sesuatu berdasarkan variabel linguistik yang dinyatakan dengan fungsi keanggotaan dalam semesta  $U$  (Saleh dkk. [10]). Himpunan *fuzzy* memiliki dua atribut. Atribut linguistik yaitu penamaan suatu grup yang mewakili suatu keadaan atau kondisi tertentu dengan menggunakan bahasa alami. Kemudian atribut numerik yaitu suatu nilai yang menunjukkan ukuran dari suatu variabel.

### 3.3 *Fuzzy Time Series*

*Fuzzy time series* adalah metode yang diperkenalkan oleh Song dan Chissom berdasarkan teori *fuzzy set* dan konsep variabel linguistik. *Fuzzy time series* digunakan untuk menyelesaikan masalah peramalan yang mana data historis adalah nilai-nilai linguistik (Rohmawati dkk. [8]). Metode ini mudah diterapkan dan sangat fleksibel, menyediakan cara untuk menangani data numerik dan non-numerik (De Lima Silva *et al.* [3]).

### 3.4 *Fuzzy C-means*

*Fuzzy C-means* adalah suatu teknik *clustering* data yang mana keberadaan tiap-tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan yang mudah diimplementasikan, efisien dan lugas (Izakian dan Abraham [6]). Tujuan algoritme *Fuzzy C-means* (*FCM*) adalah untuk menemukan pusat *cluster* (*centroid*) dengan meminimumkan fungsi objektif (Bezdek [2]). Berikut adalah algoritme *clustering Fuzzy C-means* (Sanusi dkk. [12]):

1. Memasukan data  $X$  yang akan di-*cluster*, berupa matriks berukuran  $m \times n$  ( $m$  : jumlah sampel data,  $n$  : atribut setiap data).  $X_{ij}$  : data sampel ke- $i$  ( $i=1, 2, \dots, m$ ), atribut ke- $j$  ( $j = 1, 2, 3, \dots, n$ ).
2. Menentukan:
  - a. Jumlah *cluster* ( $c \geq 2$ )
  - b. Pembobot ( $\infty > w > 1$ )
  - c. Maksimum iterasi (*maxlter*)
  - d. *Error* terkecil yang diharapkan ( $\xi =$  nilai positif yang sangat kecil)
  - e. Fungsi objektif awal ( $P_0 = 0$ )
  - f. Iterasi awal ( $t = 1$ )
3. Membangkitkan bilangan random  $\mu_{ik}; i = 1, 2, \dots, n$  dan  $k = 1, 2, 3, \dots, c$  sebagai elemen-elemen matriks partisi awal  $U$ .
4. Menghitung pusat *cluster* ke- $k$ ;  $V_{kj}$  dengan  $k = 1, 2, 3, \dots, c$  dan  $j = 1, 2, 3, \dots, m$ .

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w \times X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w}$$

dengan:

$V_{kj}$  : Pusat *cluster* ke- $k$ , atribut ke- $j$

$\mu_{ik}$  : Derajat keanggotaan pada data ke- $i$ , *cluster* ke- $k$

$X_{ij}$  : Data sampel pada data ke- $i$ , atribut ke- $j$

$w$  : Pembobot

5. Menghitung fungsi objektif pada iterasi ke- $t$ ,  $P_t$  dapat dituliskan sebagai

$$P_t = \sum_{k=1}^n \sum_{i=1}^c \left( \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right)$$

6. Menghitung perbaikan matriks partisi

$$\mu_{ik} = \frac{\left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{-1}}{\sum_{k=1}^c \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{-1}}$$

dengan:

$X_{ij}$  : Data sampel ke- $i$ , atribut ke- $j$

$V_{kj}$  : Pusat *cluster* ke- $k$ , atribut ke- $j$

$w$  : Pembobot

7. Mengecek kondisi berhenti:

a. Jika  $(|P_t - P_{t-1}| < \xi)$  atau  $(t > \max t_{iter})$ , akan berhenti

b. Jika tidak:  $t = t + 1$ , ulangi langkah ke-4

### 3.5 Metode *High Order Fuzzy Time Series Multi Factors*

Langkah-langkah metode *High Order Fuzzy Time Series Multi Factors* menurut Lin dan Yang [7] dituliskan sebagai berikut:

1. Menentukan *Universe of Discourse* ( $U$ )

*Universe of Discourse* ditentukan sebagai :

$$U = [D_{min} - S, D_{max} + S]$$

dengan:

$D_{min}$  : Nilai terkecil dari data sampel

$D_{max}$  : Nilai terbesar dari data sampel

$S$  : Standar deviasi data

2. Menentukan jumlah *cluster* data, *clustering* bertujuan untuk membagi himpunan semesta  $U$  ke dalam beberapa subinterval  $u_i$ .

$$k = \left\lceil |D_{min} - D_{max}| / \frac{\sum_{t=2}^n |x(t) - x(t-1)|}{n-1} \right\rceil$$

dengan:

$k$  : Jumlah *cluster*

$D_{min}$  : Nilai terkecil dari data sampel

$D_{max}$  : Nilai terbesar dari data sampel

$n$  : Banyaknya data sampel

$x(t)$  : Data pada waktu  $t$

Apabila hasil perhitungan  $k$  bernilai pecahan, maka  $k$  dibulatkan.

### 3. Pembentukan subinterval

Proses pengelompokan data menggunakan *Fuzzy C-Means* untuk mendapatkan pusat *cluster* sebanyak  $k$ . *Universe of Discourse* dibagi dalam  $k$  interval:  $(D_{min}, d_1)$ ,  $(d_1, d_2)$ ,  $(d_2, d_3)$ , ...,  $(d_{k-1}, D_{max})$ , dimana  $d_i$  ( $i = 1, 2, \dots, k - 1$ ) adalah titik tengah antara dua pusat *cluster*.  $K$  subinterval dinyatakan sebagai  $u_1, u_2, \dots, u_k$ . Subinterval dengan panjang berbeda dapat merefleksikan struktur data kompleks dengan lebih baik dibanding pembagian subinterval sama panjang.

### 4. Pembentukan *fuzzy set*

*Fuzzy set*  $A_i$  ( $i = 1, 2, \dots, k$ ) dapat dibentuk

$$\begin{aligned} A_1 &= f_{11}/u_1 + f_{12}/u_2 + \dots + f_{1k}/u_k \\ A_2 &= f_{21}/u_1 + f_{22}/u_2 + \dots + f_{2k}/u_k \\ &\dots = \dots + \dots + \dots + \dots \end{aligned}$$

$$A_k = f_{k1}/u_1 + f_{k2}/u_2 + \dots + f_{kk}/u_k$$

Dengan  $f_{ij}$  menunjukkan derajat keanggotaan dari  $u_j$  dalam *fuzzy set*  $A_i$  ( $i = 1, 2, \dots, k; j = 1, 2, \dots, k$ ). Tanda “+” menunjukkan operator himpunan gabungan.

### 5. Fuzzifikasi

Disini nilai data yang masih berupa bilangan asli dari data sampel diubah menjadi nilai *fuzzy* dan nilai *fuzzy* tersebut dalam bentuk derajat keanggotaan. Dari beberapa nilai derajat keanggotaan yang ada, dipilih satu nilai dari *fuzzy set* yang memiliki derajat keanggotaan paling tinggi.

### 6. Pembentukan *Fuzzy Logic Relationship* (FLR)

Data sampel sebanyak  $n$  periode (*order*) dipilih sebagai data latih untuk membentuk *fuzzy time series* model. Misal  $t$  adalah periode yang hendak diramal, dengan *order* 3 maka tiga periode sebelum periode  $t$  adalah  $t-3, t-2, t-1$  dengan  $t = 4, 5, \dots, n$ . FLR dapat dihasilkan seperti pada persamaan

$$\begin{aligned} &(A_{(t-3,i3)}, B_{(t-3,j3)}, C_{(t-3,p3)}, D_{(t-3,q3)}), \\ &(A_{(t-2,i2)}, B_{(t-2,j2)}, C_{(t-2,p2)}, D_{(t-2,q2)}), \\ &(A_{(t-1,i1)}, B_{(t-1,j1)}, C_{(t-1,p1)}, D_{(t-1,q1)}) \\ &\quad \rightarrow A_{(t,i4)} \end{aligned}$$

### 7. Defuzzifikasi

Proses perubahan besaran *fuzzy* dalam bentuk himpunan *fuzzy* untuk mendapat kembali bentuk tegasnya. Defuzzifikasi dilakukan untuk mendapatkan hasil dari peramalan. Disini dilakukan pencocokan dari data latih dengan data uji yang diambil dari FLR. Data latih dinyatakan cocok dengan data uji bila selisih nilai absolut *antecedent factor* (faktor di ruas kiri FLR) dari kedua data tersebut lebih kecil dari nilai *threshold*. Kemudian untuk memperoleh nilai hasil peramalan digunakan metode *centroid*, metode tersebut ditulis sebagai:

$$r_T = \frac{\sum_{i=1}^{k_1} c_i \times f_i}{\sum_{i=1}^{k_i} f_i}$$

dengan:

$r_T$  : Hasil peramalan hari ke- $T$

- $c_i$  : Pusat *cluster*  $i$  dimana  $A(*, i)$  adalah *secedent factor* dari *FLR* hasil proses pencocokan  
 $f_i$  : Frekuensi munculnya  $A(*, i)$  pada saat proses pencocokan.

#### 4. Kesimpulan

Berdasarkan hasil dan pembahasan disimpulkan metode *High Order Fuzzy Time Series Multi Factors* adalah metode *Fuzzy Time Series* (FTS) yang melibatkan beberapa faktor yang mempengaruhi peramalan dengan berdasarkan beberapa periode waktu sebelum periode yang akan diramalkan dalam proses pembentukan *Fuzzy Logical Series* (FLS). Metode *High Order Fuzzy Time Series Multi Factors* dengan *Fuzzy C-Means* (FCM) dapat memperbaiki kelemahan algoritma *Multi-Factors High Order Fuzzy Time Series* pada saat pembentukan subinterval.

#### Daftar Pustaka

- [1] Asriningtias, Y., & Mardhiyah, R. (2014). Aplikasi Data Mining Untuk Menampilkan Informasi. *Informatika*, 8(1), 837–848.
- [2] Bezdek, J. C. (1984). *FCM: THE FUZZY c-MEANS CLUSTERING ALGORITHM I*; *yk E Y ~ l*. 10(2), 191–203.
- [3] De Lima Silva, P. C., Sadaei, H. J., Ballini, R., & Guimaraes, F. G. (2020). Probabilistic Forecasting with Fuzzy Time Series. *IEEE Transactions on Fuzzy Systems*, 28(8), 1771–1784. <https://doi.org/10.1109/TFUZZ.2019.2922152>
- [4] Dorteus, L. R. (2015). Penerapan Logika Fuzzy Metode Sugeno Berdasarkan Data Persediaan Dan Jumlah Permintaan (Studi Kasus: Pabrik Roti Sarinda Ambon). *Jurnal Ilmu Matematika Dan Terapan*, 9(2), 121–134. <https://ojs3.unpatti.ac.id/index.php/barekeng/article/view/289/249>
- [5] Harapan, U. P., Ratnawati, D. E., & Dewi, C. (2018). Implementasi Metode Gabungan *Multi-Factors High Order Fuzzy Time Series* dan *Fuzzy C-Means* Untuk Peramalan Kebutuhan Energi Listrik di Indonesia Identifying Quality of Patchouli Leaves through Its Leave Image Using Learning Vector Quantization View project. 2(3), 905–914. <http://j-ptiik.ub.ac.id>
- [6] Izakian, H., & Abraham, A. (2011). Fuzzy C-means and fuzzy swarm for fuzzy clustering problem. *Expert Systems with Applications*, 38(3), 1835–1838. <https://doi.org/10.1016/j.eswa.2010.07.112>
- [7] Lin, Y. (2009). *Stock Markets Forecasting Based on Fuzzy Time Series Model*.
- [8] Rohmawati, F., Rohman, M. G., & Mujilahwati, S. (2017). Sistem Prediksi Jumlah

- Pengunjung Wisata Wego Kec.Sugio Kab.Lamongan Menggunakan Metode Fuzzy Time Series. *Jouticla*, 2(2). <https://doi.org/10.30736/jti.v2i2.66>
- [9] Saepullah, A. (2015). Comparative Analysis of Mamdani, Sugeno and Tsukamoto Method of Fuzzy Inference System for Air Conditioner Energy Saving. *Journal of Intelligent Systems*, 1(2), 143–147.
- [10] Saleh, M. N., Irwansyah, M. A., & Anra, H. (2017). Implementasi Peramalan Menggunakan Fuzzy Time Series pada Aplikasi Helpdesk Inventaris Perangkat Teknologi Informasi. ... *Dan Teknologi Informasi*, 5(2), 2–7. <https://jurnal.untan.ac.id/index.php/justin/article/view/18721>
- [11] Santoso, H., Hariyadi, I. P., & Prayitno. (2016). Data Mining Analisa Pola Pembelian Produk. *Teknik Informatika*, 1, 19–24. <http://ojs.amikom.ac.id/index.php/semnasteknomedia/article/download/1267/1200>
- [12] Sanusi, W., Zaky, A., & Afni, B. N. (2020). Analisis Fuzzy C-Means dan Penerapannya Dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Faktor-faktor Penyebab Gizi Buruk. *Journal of Mathematics, Computations, and Statistics*, 2(1), 47. <https://doi.org/10.35580/jmathcos.v2i1.12458>
- [13] Sihombing, E. G. (2017). Klasifikasi Data Mining pada Rumah Tangga Menurut Provinsi dan Status Kepemilikan Rumah Kontrak/Sewa Menggunakan K-Means Clustering Method. *Computer Engineering, System and Science Journal*, 2(2), 74–82.
- [14] Sitohang, S., & Denson Napitupulu, R. (2017). Fuzzy Logic Untuk Menentukan Penjualan Rumah dengan Metode Mamdani (Studi Kasus: Pt Gracia Herald). *Jurnal ISD*, 2(2), 91–101.
- [15] Wijaya, A. B., Dewi, C., & Rahayudi, B. (2018). Peramalan Curah Hujan Menggunakan Metode High Order Fuzzy Time Series Multi Factors. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(3), 930–939.