

**PERBANDINGAN ANTARA METODE *K-MEANS CLUSTERING*  
DENGAN *GATH-GEVA CLUSTERING*  
(STUDI KASUS PADA VOLUME EKSPOR NON MIGAS PAKAIAN JADI)**

Oleh :  
Siti Lailiyah<sup>1)</sup>, Moh. Hafiyusholeh<sup>2)</sup>

<sup>1)</sup>Pendidikan Matematika UIN Sunan Ampel Surabaya, siti03\_math\_its@yahoo.com

<sup>2)</sup>Matematika UIN Sunan Ampel Surabaya, hafiyusholeh@uinsby.ac.id

**ABSTRAK**

*Perdagangan luar negeri Indonesia sedang ditata kembali format dan kinerjanya, agar pemerintah tidak membuat kesalahan dalam mengambil keputusan untuk meningkatkan ekspor non migas, maka pemerintah harus mampu memprediksi volume ekspor non migas. Prediksi pada dasarnya merupakan suatu perkiraan tentang terjadinya suatu kejadian di waktu yang akan datang. Salah satu cara yang dapat digunakan untuk memprediksi nilai ekspor tersebut adalah dengan k-means clustering dan gath-geva clustering. Kemudian dibentuk Fuzzy Inference System (FIS) untuk memperoleh hasil prediksi sehingga didapatkan error dan validasi hasil prediksi. Berdasarkan hasil analisa RMSE, cek maksimum dan cek minimum maka dapat disimpulkan bahwa metode Gath-Geva (GG) Clustering lebih teliti dibandingkan dengan metode K-means clustering.*

**Kata kunci:** *Sistem Fuzzy, K-means clustering, Gath-Geva clustering.*

**1. Pendahuluan**

Perdagangan luar negeri Indonesia sekarang ini sedang ditata kembali format dan kinerjanya ke kondisi awal dimasa-masa sebelum krisis ekonomi melanda bangsa Indonesia. Meskipun belum pulih secara keseluruhan, kegiatan perdagangan luar negeri Indonesia mulai menunjukkan kekuatannya kembali.

Sudah beberapa tahun terakhir ini hasil industri pakaian jadi Indonesia banyak diminati oleh negara-negara asing khususnya Amerika dan Eropa karena dianggap cukup berkualitas dengan harga yang bersaing. Bahkan di beberapa negara, hasil industri pakaian jadi Indonesia dikenakan kuota agar industri pakaian jadi negara tersebut tidak kalah bersaing dengan produk Indonesia.

Agar pemerintah tidak membuat kesalahan dalam mengambil keputusan untuk meningkatkan ekspor non migas khususnya untuk sektor industri, maka

pemerintah harus mampu melihat ke depan atau memprediksi volume ekspor non migas.

Cara yang dapat digunakan untuk memprediksi nilai ekspor diantaranya dengan metode *K-means clustering*, dan metode *Gath-Geva clustering*.

Beberapa penelitian yang terkait adalah Yunianti [10], dalam “Prediksi Volume Ekspor Non Migas Dengan Metode Subtractive Clustering” menyimpulkan bahwa pengolahan data dengan metode *subtractive clustering* menghasilkan jangkauan data dapat berpengaruh terhadap pembentukan cluster sehingga mempengaruhi hasil prediksi meskipun diberikan radius yang sama untuk data-data yang ada.

Penelitian yang lain, Gunawan [7], dalam “Perbandingan Antara Metode Fuzzy Subtractive Clustering Dengan Metode Gustafson Kessel Clustering (studi kasus data time series volume pinjaman kredit

gadi)” menyimpulkan bahwa hasil RMSE, error maksimum dan error minimum metode *gustafson kessel clustering* lebih teliti dibandingkan dengan metode *subtractive clustering*.

Sebagai kelanjutan dari penelitian sebelumnya, maka pengkajian dari penelitian Penelitian ini dilakukan dengan menganalisa secara khusus terhadap data volume ekspor non migas pada pakaian jadi menggunakan metode *K-means clustering* dan metode *Gath-Geva clustering*. Perbandingan hasil analisis terhadap kedua algoritma metode tersebut akan menentukan metode yang lebih teliti untuk memprediksi volume ekspor non migas dengan nilai-nilai parameter dan besaran yang terkait

Permasalahan utama yang diangkat dalam Penelitian ini adalah: Bagaimana perbandingan antara metode K-means clustering dan Gath-Geva Clustering untuk mendapatkan cluster optimal.

Dalam pembahasan ini, kajian permasalahan akan dibatasi dengan menentukan nilai-nilai parameter dan besaran yang terkait, antara lain:

1. Jari-jari cluster.
2. Indeks validasi masing-masing algoritma yaitu Koefisien Partisi (PC), Entropi Klasifikasi (CE), Partisi Indeks (SC), Separasi Indeks (S) dan Jumlah Kuadrat Kesalahan (SSE).
3. RMSE (Root Means Square Error) dalam peramalan.

Tujuan dari Penelitian ini adalah (1) dapat memprediksi volume ekspor non migas untuk pakaian jadi dengan metode K-means Clustering dan Gath-Geva Clustering; (2) dapat mengetahui unjuk kerja dua algoritma pengklasteran diatas secara empiris.

## 2. Teori Dasar

### 1) Sistem Fuzzy

Sistem fuzzy terdiri dari himpunan fuzzy, fungsi keanggotaan, fuzzy inferensi sistem dan penalaran fuzzy.

Himpunan fuzzy  $A$  dalam semesta  $X$ , ditulis dengan  $\tilde{A}$  dan didefinisikan oleh pasangan:

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) / x \in X\}$$

dengan  $\mu_{\tilde{A}} : X \rightarrow [0, 1]$  adalah fungsi atau derajat keanggotaan dari himpunan fuzzy  $\tilde{A}$ .

Fungsi keanggotaan (*membership function*) adalah suatu kurva yang menunjukkan pemetaan titik-titik input data ke dalam nilai keanggotaannya yang memiliki interval antara 0 dan 1.

Pada dasarnya nilai keanggotaan fuzzy dapat digambarkan melalui beberapa representasi yaitu:

#### a. Representasi linear

Ada 2 keadaan himpunan fuzzy yang linear.

Pertama,

Fungsi keanggotaan :

$$\mu[x] = \begin{cases} 0 & ; x \leq a \\ (x - a)/(b - a) & ; a \leq x \leq b \\ 1 & ; x \geq b \end{cases}$$

Kedua,

Fungsi keanggotaan :

$$\mu[x] = \begin{cases} (b - x)/(b - a) & ; a \leq x \leq b \\ 0 & ; x \geq b \end{cases}$$

#### b. Representasi kurva segitiga

Fungsi ini diidentifikasi dengan 3 parameter.

Fungsi keanggotaan:

$$\mu[x] = \begin{cases} 0 & ; x \leq a \vee x \geq c \\ (x - a)/(b - a) & ; a \leq x \leq b \\ (b - x)/(c - b) & ; b \leq x \leq c \end{cases}$$

#### c. Representasi kurva trapesium

Ada 4 parameter yang dapat digunakan yaitu [ a b c d ].

Fungsi keanggotaan :

$$\mu[x] = \begin{cases} 0 & ; x \leq a \text{ atau } x \geq d \\ (x - a)/(b - a) & ; a \leq x \leq b \\ 1 & ; b \leq x \leq c \\ (d - x)/(d - c) & ; x \geq d \end{cases}$$

**2) Fuzzy Inference System (FIS)**

*Fuzzy Inference System* (FIS) adalah sistem yang didasarkan pada konsep teori himpunan fuzzy, aturan fuzzy, dan penalaran fuzzy. Secara umum, didalam logika fuzzy ada 5 langkah dalam melakukan penalaran, yaitu :

1. Memasukkan input fuzzy.
2. Mengaplikasikan operator fuzzy.
3. Mengaplikasikan metode implikasi
4. Mengkomposisi semua output
5. Defuzzifikasi

Implikasi adalah proses pembentukan himpunan fuzzy pada konsekuen yang didasarkan pada anteseden.

Apabila sistem terdiri dari beberapa aturan, maka inferensi diperoleh dari kumpulan dan korelasi antar aturan.

Input dari proses defuzzifikasi adalah suatu himpunan fuzzy yang diperoleh dari komposisi aturan-aturan fuzzy, sedangkan output yang dihasilkan merupakan suatu bilangan pada domain himpunan fuzzy tersebut.

**3) Fuzzy Clustering**

*Fuzzy clustering* adalah salah satu teknik untuk menentukan *cluster* optimal dalam suatu ruang vektor yang didasarkan pada bentuk norma Euclid untuk jarak antar vektor.

Suatu ukuran fuzzy menunjukkan derajat ke-fuzzy-an dari himpunan fuzzy. Secara umum ukuran ke-fuzzy-an dapat ditulis sebagai suatu fungsi:

$$f : P(x) \rightarrow R$$

dengan  $P(x)$  adalah himpunan semua subset dari  $X$ .  $f(A)$  adalah suatu fungsi yang memetakan subset  $A$  ke karakteristik derajat ke-fuzzy-annya.

Dalam mengukur nilai ke-fuzzy-an, fungsi  $f$  harus mengikuti hal-hal sebagai berikut:

1.  $f(A) = 0$  jika dan hanya jika  $A$  adalah himpunan crisp.

2. Jika  $A \subset B$ , maka  $f(A) \leq f(B)$ . Disini,  $A \subset B$  berarti  $B$  lebih fuzzy dibanding  $A$ . Relasi ketajaman  $A \subset B$  didefinisikan dengan:

$$\mu_A[x] \leq \mu_B[x], \text{ jika } \mu_A[x] \leq 0.5$$

$$\mu_A[x] \geq \mu_B[x], \text{ jika } \mu_B[x] \geq 0.5$$

3.  $f(A)$  akan mencapai maksimum jika dan hanya jika  $A$  benar-benar fuzzy secara maksimum. Nilai fuzzy maksimal biasanya terjadi pada saat  $\mu_A[x] = 0.5$  untuk setiap  $x$ .

Pada aplikasi fuzzy clustering, biasanya perlu dilakukan *pre-processing* terlebih dahulu. Dengan demikian kita perlu melakukan normalisasi untuk suatu nilai  $u$ , menjadi  $u$  normal ( $\bar{u}$ ) dengan rumus:

$$\bar{u} = \frac{u - u_{\min}}{u_{\max} - u_{\min}}$$

dengan  $u_{\min}$  adalah nilai terkecil yang terukur dan  $u_{\max}$  adalah nilai terbesar yang terukur.

Setelah melakukan preprocessing, variabel-variabel yang relevan dapat segera dipilih. Untuk sekumpulan data  $u = [u_1, u_2, \dots, u_N]$  dapat dicari:

- mean  $m = \frac{1}{N} \sum_{i=1}^N u_i$
- variansi  $v = \frac{1}{N-1} \sum_{i=1}^N (u_i - m)^2$
- deviansi standar  $\sigma = \sqrt{v}$
- range  $s = u_{\max} - u_{\min}$
- koefisien korelasi  $r = \frac{\sum_{i=1}^N (x_i - m_1)(y_i - m_2)}{\sqrt{\sum_{i=1}^N (x_i - m_1)^2 \sum_{i=1}^N (y_i - m_2)^2}}$

Dengan  $m_1$  adalah mean dari  $X$  dan  $m_2$  adalah mean dari  $Y$ .

**4) Analisis Cluster**

Tujuan dari Analisis cluster adalah pengklasifikasian obyek-obyek berdasar-

kan similaritas diantaranya dan menghimpun data menjadi beberapa kelompok.

Data yang diberikan adalah data kuantitatif yang diperoleh dari hasil pengukuran berupa data runtun waktu (time series). Setiap pengamatan terdiri dari  $n$  variabel, dikelompokkan dalam vektor baris pada dimensi- $n$

$$x_k = [x_{k1}, x_{k2}, \dots, x_{kn}]^T, \quad x_k \in \mathbf{R}^n$$

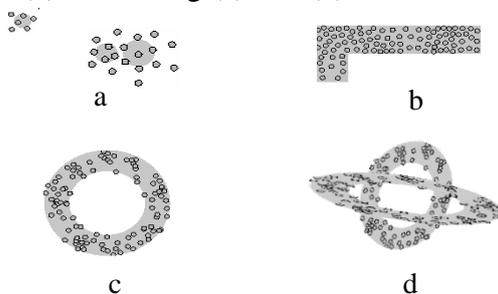
himpunan pada  $N$  pengamatan di-notasikan oleh  $\mathbf{X} = \{x_k \mid k = 1, 2, \dots, N\}$

dan disajikan dalam matriks  $N \times n$ :

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{Nn} \end{bmatrix} \quad (2.1)$$

baris dari  $\mathbf{X}$  disebut pola atau obyek, sedangkan kolomnya disebut *features* atau *attributes* dan  $\mathbf{X}$  sendiri disebut matrik pola atau matrik data.

Data dapat dinyatakan sebagai cluster pada ruang geometri yang berbeda, ukuran dan kepadatannya, yang ditunjukkan pada Gambar 2.1. Cluster adalah kumpulan dari benda-benda yang memiliki banyak kesamaan satu sama lain daripada dengan anggota cluster yang lain. Cluster dapat berbentuk bola (a), garis panjang atau lurus (b), berlubang (c) dan (d).



**Gambar 2.1. Cluster dengan bentuk dan dimensi yang berbeda dalam  $\mathcal{R}^2$ .**

Pada hard clustering, setiap obyek (data) hanya bisa menjadi anggota tepat satu cluster. Sedangkan pada fuzzy clustering obyek-obyek dapat menjadi

anggota dari lebih satu cluster dengan derajat keanggotaan yang berbeda. Dengan demikian  $X$  mempunyai  $c$  himpunan bagian fuzzy.

Struktur dari partisi matrik  $U = [\mu_{ik}]$ :

$$U = \begin{bmatrix} \mu_{1,1} & \mu_{1,2} & \dots & \mu_{1,c} \\ \mu_{2,1} & \mu_{2,2} & \dots & \mu_{2,c} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{N,1} & \mu_{N,2} & \dots & \mu_{N,c} \end{bmatrix}$$

### 1. Hard partition

Hard partisi dapat diartikan sebagai keluarga dari himpunan bagian  $\{A_i \subset X \mid 1 \leq i \leq c, A_i \neq \emptyset\}$ , yang memenuhi:

$$\bigcup_{i=1}^c A_i = X \quad (2.2)$$

$$A_i \cap A_j = \emptyset, 1 \leq i \neq j \leq c, \quad (2.3)$$

dimana,

$A_i$  = Himpunan bagian pada cluster ke- $i$ .

$A_j$  = Himpunan bagian pada cluster ke- $j$ .

$X$  = Semesta pembicaraan.

$c$  = Banyaknya himpunan bagian cluster.

Bentuk dari fungsi keanggotaannya:

$$\bigvee_{i=1}^c \mu_{A_i} = 1, \quad (2.4)$$

$$\mu_{A_i} \vee \mu_{A_j} = 1, \quad 1 \leq i \neq j \leq c \quad (2.5)$$

$$0 < \mu_{A_i} < 1, \quad 1 \leq i \leq c \quad (2.6)$$

dimana,

$\mu_{A_i}$  = fungsi karakteristik dari himpunan bagian  $A_i$  yang bernilai 0 dan 1.

$\mu_{A_j}$  = fungsi karakteristik dari himpunan bagian  $A_j$  yang bernilai 0 dan 1.

$c$  = Banyaknya himpunan bagian cluster.

Matrik  $N \times c$ ,  $U = [\mu_{ik}]$  menggambarkan hard partisi jika hanya jika elemennya memenuhi:

$$\mu_{ij} \in \{0, 1\}, 1 \leq i \leq N, 1 \leq j \leq c \quad (2.7)$$

$$\sum_{k=1}^c \mu_{ik} = 1, \quad 1 \leq i \leq N, \quad (2.8)$$

$$0 < \sum_{i=1}^N \mu_{ik} < N, \quad 1 \leq k \leq c \quad (2.9)$$

dimana,

$\mu_{ij}$  = fungsi karakteristik pada himpunan bagian  $A_i$  dan  $A_j$ .

$\mu_{ik}$  = Fungsi karakteristik pada  $x_k$ .

$N$  = Banyaknya cluster.  
 $c$  = Banyaknya himpunan bagian cluster.

**Definisi 2.1 (Hard Partition Space)**

Misal  $X = [x_1, x_2, \dots, x_N]$  adalah himpunan berhingga dan misal  $c \in \{2, 3, \dots, N\}$ . *Hard Partition Space* pada  $X$  adalah himpunan

$$M_{hc} = \{U \in \mathcal{R}^{N \times c} \mid \mu_{ik} \in [0,1], \forall i, k, \sum_{k=1}^c \mu_{ik} = 1, \forall i; 0 < \sum_{i=1}^N \mu_{ik} < N, \forall k\} \quad (2.10)$$

2. Fuzzy partition

Pada fuzzy partisi  $\mu_{ik}$  mencapai nilai riil  $[0, 1]$ . Matriks  $N \times c$ ,  $U = [\mu_{ik}]$  digambarkan sebagai fuzzy partisi. Kondisi ini diberikan oleh:

$$\mu_{ij} \in [0,1], 1 \leq i \leq N, \quad 1 \leq k \leq c \quad (2.11)$$

$$\sum_{k=1}^c \mu_{ik} = 1, \quad 1 \leq i \leq N, \quad (2.12)$$

$$0 < \sum_{i=1}^N \mu_{ik} < N, 1 \leq k \leq c \quad (2.13)$$

dimana,

$\mu_{ij}$  = fungsi karakteristik pada himpunan bagian  $A_i$  dan  $A_j$ .

$\mu_{ik}$  = Fungsi karakteristik pada  $x_k$ .

$N$  = Banyaknya cluster.

$c$  = Banyaknya himpunan bagian cluster.

**Definisi 2.2 (Fuzzy Partition Space)**

Misal  $X = [x_1, x_2, \dots, x_N]$  adalah himpunan berhingga dan misal  $c \in \{2, 3, \dots, N\}$ . *Fuzzy Partition Space* pada  $X$  adalah himpunan

$$M_{fc} = \{U \in \mathcal{R}^{N \times c} \mid \mu_{ik} \in [0,1], \forall i, k; \sum_{k=1}^c \mu_{ik} = 1, \forall i; 0 < \sum_{i=1}^N \mu_{ik} < N, \forall k\} \quad (2.14)$$

**5) Peramalan**

Sering terdapat senjang waktu (time lag) antara kesadaran akan peristiwa atau kebutuhan mendatang dengan peristiwa itu sendiri. Adanya waktu tenggang (time lead) ini merupakan alasan utama bagi perencanaan dan peramalan.

Perihal mendasar yang perlu diperhatikan dalam peramalan adalah

bagaimana mengukur kesesuaian suatu metode peramalan tertentu untuk suatu kumpulan data yang diberikan. Berbagai ketepatan ukuran peramalan (pemodelan) didefinisikan dengan mempertimbangkan ukuran-ukuran yang ada. Ukuran-ukuran yang dimaksudkan diantaranya yaitu :

1. Ukuran statistik.

Jika  $X_t$  merupakan data aktual untuk periode  $t$  dan  $F_t$  merupakan ramalan (atau nilai kecocokan/fitted value) untuk periode yang sama, maka kesalahannya didefinisikan sebagai berikut :

$$e_t = X_t - F_t \quad (2.15)$$

Jika terdapat nilai pengamatan ramalan untuk  $n$  periode waktu, maka akan terdapat  $n$  buah kesalahan dan ukuran statistik standar berikut dapat didefinisikan:

Nilai Tengah Kesalahan (Mean Error)

$$ME = \sum_{t=1}^n e_t / n \quad (2.16)$$

Nilai Tengah Absolute (Mean Absolute Error)

$$MAE = \sum_{t=1}^n |e_t| / n \quad (2.17)$$

Jumlah Kuadrat Kesalahan (Sum of Squared Error)

$$SUM = \sum_{t=1}^n e_t^2 \quad (2.18)$$

Nilai Tengah Kesalahan Kuadrat (Mean Squared Error)

$$MSE = \sum_{t=1}^n e_t^2 / n \quad (2.19)$$

Deviasi Standar Kesalahan (Standard Deviation of Error)

$$SDE = \sqrt{\sum_{t=1}^n e_t^2 / (n-1)} \quad (2.20)$$

2. Ukuran-ukuran relatif.

Tiga ukuran yang sering digunakan adalah :

Kesalahan Persentase (Percentage Error)

$$PE_t = \left( \frac{X_t - F_t}{X_t} \right) \quad (2.21)$$

Nilai Tengah Kesalahan Persentase (*Mean Percentage Error*)

$$MPE = \frac{\sum_{i=1}^n PE_i}{n} \quad (2.22)$$

Nilai Tengah Kesalahan Persentase Absolut (*Mean Absolute Percentage Error*)

$$MAPE = \frac{\sum_{i=1}^n |PE_i|}{n} \quad (2.23)$$

### 3. Metode Penelitian

Tahap-tahap yang digunakan dalam penelitian ini agar dapat mencapai tujuan penelitian adalah:

#### 1. Studi literatur

Pada tahap ini meliputi pencarian informasi dan pemahaman teoritis dari penerapan metode K-means clustering dan Gath-Geva clustering untuk memprediksi volume ekspor non migas pada pakaian jadi.

#### 2. Pengambilan data

Data yang diolah adalah data sekunder yang diambil dari Badan Pusat Statistik (BPS) mulai dari bulan Januari 1987 sampai dengan bulan Desember 2005, yaitu data ekspor non migas pada pakaian jadi dalam satuan berat (ton).

#### 3. Penyusunan Algoritma K-means Clustering

Metode K-means clustering dapat dirumuskan sebagai berikut:

$$\sum_{i=1}^c \sum_{k \in A_i} \|x_k - v_i\|^2 \quad (3.1)$$

dimana,

$x_k$  = titik data ke-k pada cluster,

$A_i$  = himpunan titik data dalam cluster ke-i,

$v_i$  = mean titik data pada data cluster ke-i,

dalam algoritma K-means,  $v_i$  disebut sebagai *cluster prototype*, yaitu pusat cluster

$$v_i = \frac{\sum_{k=1}^{N_i} x_k}{N_i}, x_k \in A_i, \quad (3.2)$$

dimana,

$x_k$  = titik data ke-k pada cluster,

$A_i$  = himpunan titik data dalam cluster ke-i,

$N_i$  = banyaknya titik-titik data pada cluster ke-i,

Hasil pengklasteran yang diperoleh selanjutnya dipergunakan pada model peramalan.

#### 4. Penyusunan Algoritma Gath-Geva Clustering

Metode Gath-Geva dapat dirumuskan sebagai berikut:

$$D_{ik}(x_k, v_i) = \frac{\sqrt{\det(F_{oi})}}{\alpha_i} \times \exp\left(\frac{1}{2}(x_k - v_i^{(i)})^T F_{oi}^{-1}(x_k - v_i^{(i)})\right) \quad (3.5)$$

dimana,

$x_k$  = titik data ke-k pada cluster,

$v_i$  = mean titik data pada data cluster ke-i,

$F_{oi}$  = matriks kovariansi fuzzy dari cluster ke-i,

$\alpha_i$  = probabilitas prior dari cluster terpilih,

Dengan matriks kovariansi fuzzy

$$F_{oi} = \frac{\sum_{k=1}^N (\mu_{ik}^\omega (x_k - v_i)(x_k - v_i)^T)}{\sum_{k=1}^N (\mu_{ik}^\omega)} \quad 1 \leq i \leq c \quad (3.6)$$

dan

$$\alpha_i = \frac{1}{N} \sum_{k=1}^N \mu_{ik} \quad (3.7)$$

Derajat keanggotaan  $\mu_{ik}$  diinterpretasikan sebagai probabilitas posterior dari cluster ke-i yang terpilih. Hasil pengklasteran yang diperoleh selanjutnya juga akan dipergunakan pada model peramalan.

#### 5. Validasi

Perbandingan terhadap unjuk kerja kedua algoritma pengklasteran di atas dilakukan dengan mengukur indeks validasi masing-masing algoritma:

(a). Koefisien partisi (PC): mengukur nilai overlapping antara klaster dengan

$$PC(c) = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^N (\mu_{ij})^2 \quad (3.12)$$

dimana  $\mu_{ij}$  adalah anggota dari titik data  $j$  dalam cluster ke- $i$ . Optimal jumlah cluster adalah nilai maksimum dari PC.

(b). Entropi Klasifikasi (CE): mengukur ke-fuzzy-an partisi cluster dengan

$$CE(c) = -\frac{1}{N} \sum_{i=1}^c \sum_{j=1}^N \mu_{ij} \log(\mu_{ij}) \quad (3.13)$$

(c). Partisi indeks (SC): merupakan rasio antara kekompakan dan separasi cluster-cluster, dengan

$$SC(c) = \frac{\sum_{i=1}^c \sum_{j=1}^N (\mu_{ij})^m \|x_j - v_i\|^2}{N_i \sum_{k=1}^c \|v_k - v_i\|^2} \quad (3.14)$$

Nilai paling rendah dari SC menunjukkan partisi terbaik.

(d). Separasi indeks (S): separasi indeks menggunakan separasi jarak minimum untuk validasi partisinya, dengan

$$S(c) = \frac{\sum_{i=1}^c \sum_{j=1}^N (\mu_{ij})^2 \|x_j - v_i\|^2}{N \min_{i,k} \|v_k - v_i\|^2} \quad (3.15)$$

(e). Jumlah kuadrat kesalahan (SSE): jumlah simpangan kuadrat antara realita dengan hasil prediksi (peramalan).

## 6. Pengolahan Data.

Langkah-langkahnya sebagai berikut:

- Penerapan metode pengklasteran pada data observasi.
- Penerapan masing-masing algoritma pada data yang diberikan untuk menentukan banyaknya cluster.
- Penerapan metode pengklasteran dalam peramalan pada data volume ekspor non migas pakaian jadi.
- Menghitung masing-masing MSE, dengan menggunakan rumus:

$$MSE = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \quad (3.16)$$

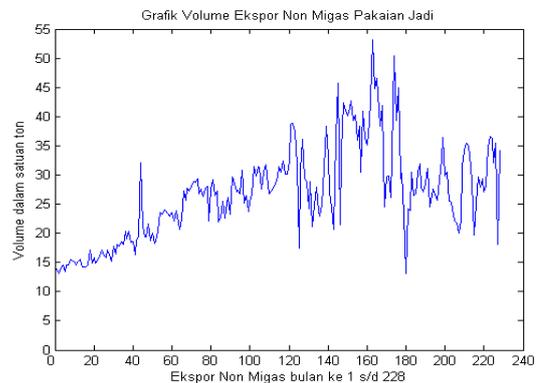
(e). Menghitung RMSE, dengan menarik akar persamaan (3.16) sehingga didapat rumus untuk menghitung RMSE sebagai berikut:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (3.17)$$

## 4. Pembahasan Dan Hasil

### 4.1 Penyajian Data

Untuk menerapkan konsep tersebut ke dalam data real, diambil data volume ekspor non migas pada pakaian jadi yang dimulai bulan Januari tahun 1987 sampai dengan bulan Desember tahun 2005. Perubahan waktu (interval waktu) data tersebut dalam satu bulan sehingga jumlah keseluruhannya adalah sejumlah 228 bulan. Perubahan secara grafis volume ekspor non migas pakaian jadi dapat dilihat pada gambar sebagai berikut:



Gambar 4.1 Volume Ekspor Nonmigas Pakaian Jadi

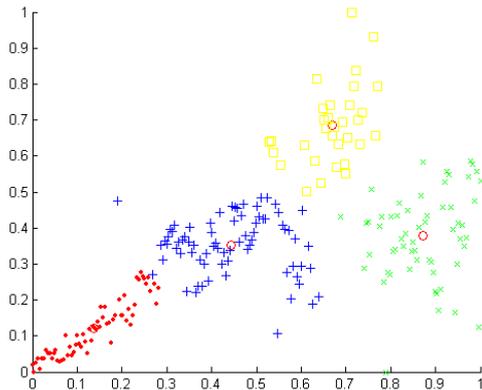
### 4.2 Penentuan Cluster

Ada beberapa cara yang dilakukan untuk menentukan cluster, diantaranya yaitu:

- Hard clustering menggunakan algoritma kmeans clustering.

Metode k-means clustering didasarkan pada pengalokasian setiap titik data kedalam satu dari  $c$  cluster yang meminimumkan jumlah kuadrat dalam cluster.

Berikut ini hasil dari plotting algoritma kmeans:



**Gambar 4.2 Kmeans Volume Ekspor Nonmigis Pakaian Jadi**

Dari Gambar 4.2 diatas diperoleh jumlah cluster = 4, dengan pusat cluster sebagai berikut:

$$C = \begin{pmatrix} 0.1382 & 0.1200 \\ 0.4438 & 0.3536 \\ 0.6693 & 0.6845 \\ 0.8722 & 0.3792 \end{pmatrix}$$

Dari hasil coba dengan merubah parameter c (jumlah cluster) maka didapatkan hasil sebagai berikut:

**Tabel 4.1: Hasil Validasi Kmeans**

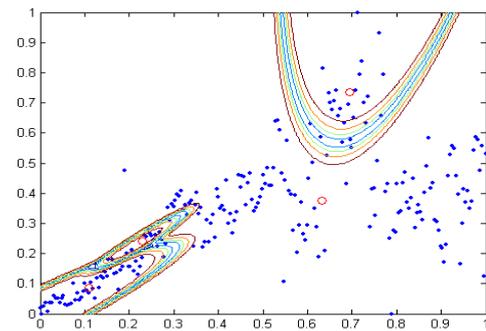
C	PC (Partition Coeffisient)	CE (Classification Entropy)	SC (Partition Index)	S (Separation Index)
3	1	NaN	0.7599	0.0056
4	1	NaN	0.5672	0.0040
5	1	NaN	0.5418	0.0040

Untuk menentukan cluster optimal diperoleh saat PC bernilai maksimum dan SC bernilai minimum, karena PC yang bernilai tetap dan nilai SC yang se-makin berkurang maka tidak dapat ditentukan cluster optimal dari nilai PC dan SC. Oleh karena itu dibutuhkan nilai validasi yang lainnya yaitu S yang ber-nilai tetap pada saat jumlah cluster = 4 se-hingga dapat disimpulkan bahwa cluster optimal terjadi pada saat jumlah cluster =4.

## 2. algoritma gath-geva clustering

Metode Gath-Geva didasarkan pada fungsi jarak dengan *fuzzy maximum likelihood estimates*.

Berikut ini hasil dari plotting Algoritma GGclust



**Gambar 4.3 GGclust Volume Ekspor Nonmigis Pakaian Jadi**

Dari Gambar 4.3 diatas diperoleh jumlah cluster = 4, dengan pusat cluster sebagai berikut:

$$C = \begin{pmatrix} 0.1101 & 0.0869 \\ 0.2295 & 0.2425 \\ 0.6323 & 0.3761 \\ 0.6950 & 0.7358 \end{pmatrix}$$

Dari hasil coba dengan merubah parameter c (jumlah cluster) maka didapatkan hasil sebagai berikut:

**Tabel 4.2: Hasil Validasi GGclust**

C	PC (Part. Coeffisient)	CE (Classification Entropy)	SC (Partition Index)	S (Separation Index)
3	0.9583	0.0755	0.0038	2.5524 e-005
4	0.9607	NaN	NaN	NaN
5	0.9459	NaN	NaN	NaN

Untuk menentukan cluster optimal diperoleh saat PC bernilai maksimum dan SC bernilai minimum. Dari tabel 4.2 diatas cluster optimalnya diperoleh pada saat c = 4 karena nilai PC pada saat jumlah cluster = 4 bernilai maksimum dan nilai SC bernilai minimum walaupun tidak dapat ditentukan nilainya.

## 4.3 Perhitungan Prediksi Peramalan dengan Metode K-means Clustering

### 1. Data-data yang digunakan

Data yang akan digunakan untuk memprediksi besar volume ekspor non

migas pada pakaian jadi berdasarkan periode waktu dua belas bulan dengan metode *moving average* atau metode rata-rata bergerak. Input sistemnya adalah permintaan ke- (t-11), (t-10), (t-9), (t-8), (t-7), (t-6), (t-5), (t-4), (t-3), (t-2), (t-1), (t), sedangkan output sistemnya adalah permintaan ke- (t+1), dengan demikian terdapat rentang data dari 12 hingga 227 (216 data). Dari data tersebut akan digunakan sebanyak 172 data untuk dicluster, sedangkan sisanya sebanyak 44 data akan digunakan sebagai data cek.

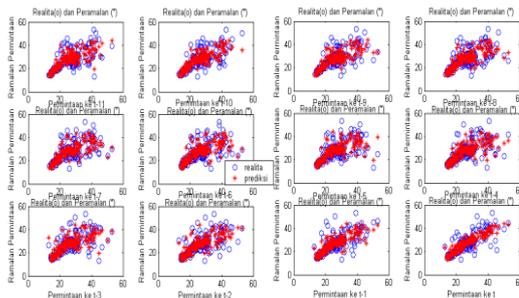
2. Hasil clustering

Jika dilakukan dengan menggunakan influence range 0,50 maka akan diperoleh 4 cluster dengan matriks pusat cluster sebagai berikut:

$$C = \begin{pmatrix} 0.1382 & 0.1200 \\ 0.4438 & 0.3536 \\ 0.6693 & 0.6845 \\ 0.8722 & 0.3792 \end{pmatrix}$$

3. Hasil inferensi

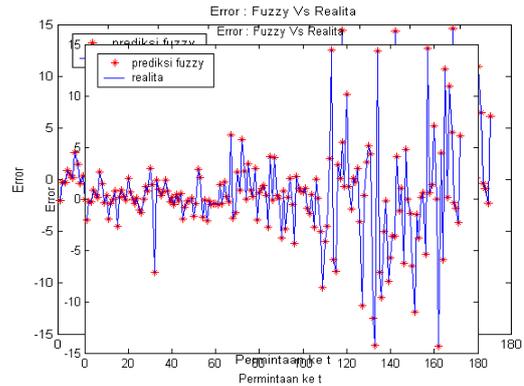
Hasil inferensi dengan jari-jari = 0.50, *accept ratio* = 0.5 dan *reject ratio* = 0.15, seperti terlihat pada berikut ini:



**Gambar 4.4: Plot hasil inferensi peramalan terhadap realita**

Tampilan error secara grafis terlihat pada Gambar 4.5. Error terbesar terjadi pada data inferensi yang ke-157 yaitu sebesar 14.6175, sedangkan error terkecil terjadi pada data inferensi yang ke-162 yaitu sebesar -14.3349.

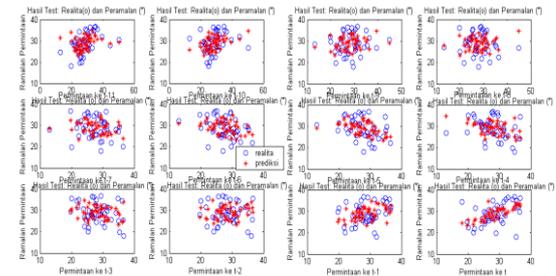
Berdasarkan tabel 1 pada lampiran 3 didapatkan hasil validasi jumlah kuadrat kesalahan (SSE) untuk data inferensi yaitu 3080.886294.



**Gambar 4.5: Grafik error hasil inferensi peramalan terhadap realita**

4. Hasil pengetesan

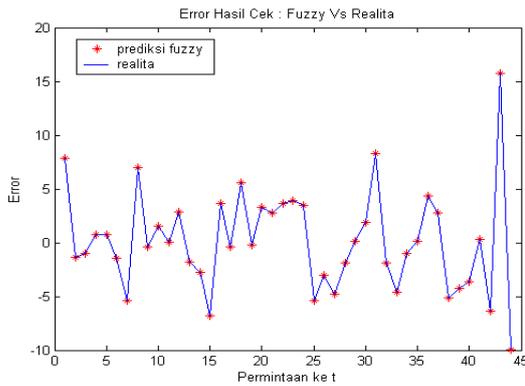
Pengetesan sistem fuzzy dengan data pengecekan berupa peramalan permintaan untuk data ke : 173 s/d 216. Hasil pengetesan dengan jari-jari = 0.50, *accept ratio* = 0.5 dan *reject ratio* = 0.15, seperti terlihat pada berikut ini:



**Gambar 4.6: Plot hasil pengetesan peramalan terhadap realita**

Tampilan error secara grafis terlihat pada Gambar 4.7 Error terbesar terjadi pada data pengetesan yang ke-43 yaitu sebesar 15.7584, sedangkan error terkecil terjadi pada data pengetesan yang ke-44 yaitu sebesar -9.9222.

Berdasarkan tabel 2 pada lampiran 3 didapatkan hasil validasi jumlah kuadrat kesalahan (SSE) untuk data pemgetesan yaitu 563.9616241.



**Gambar 4.7: Grafik error hasil cek peramalan terhadap realita**

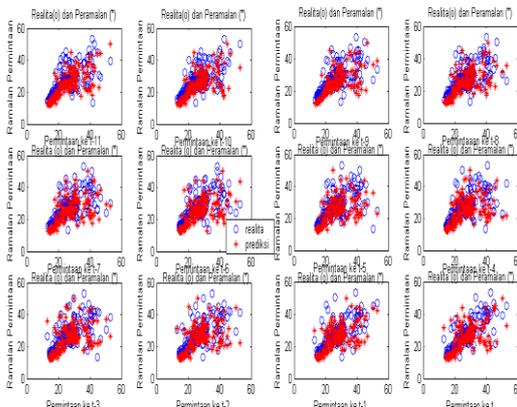
#### 4.4 Perhitungan Prediksi Peramalan dengan Metode Gath-Geva Clustering

Jika digunakan metode Gath-Geva Clustering dilakukan dengan menggunakan influence range 0.50 maka akan diperoleh 4 cluster dengan matriks pusat cluster sebagai berikut:

$$C = \begin{pmatrix} 0.1101 & 0.0869 \\ 0.2295 & 0.2425 \\ 0.6323 & 0.3761 \\ 0.6950 & 0.7358 \end{pmatrix}$$

##### 1. Hasil inferensi

Hasil inferensi dengan jari-jari = 0.50, *accept ratio* = 0.5 dan *reject ratio* = 0.15, (dapat dilihat pada tabel 3 lampiran 3) seperti terlihat pada berikut ini:

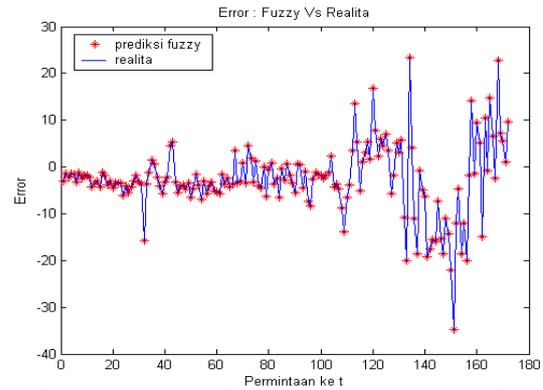


**Gambar 4.8: Plot hasil inferensi peramalan terhadap realita**

Tampilan error secara grafis terlihat pada Gambar 4.9. Error terbesar terjadi pada data inferensi yang ke-134 yaitu sebesar 23.3592, sedangkan error terkecil

terjadi pada data inferensi yang ke-151 yaitu sebesar -34.8103.

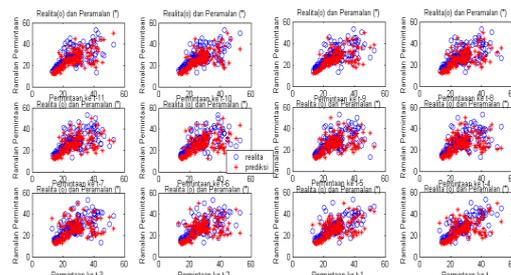
Berdasarkan tabel 3 pada lampiran 3 didapatkan hasil validasi jumlah kuadrat kesalahan (SSE) untuk data inferensi yaitu 10869.45645.



**Gambar 4.9: Grafik error hasil inferensi peramalan terhadap realita**

##### 2. Hasil pengetesan

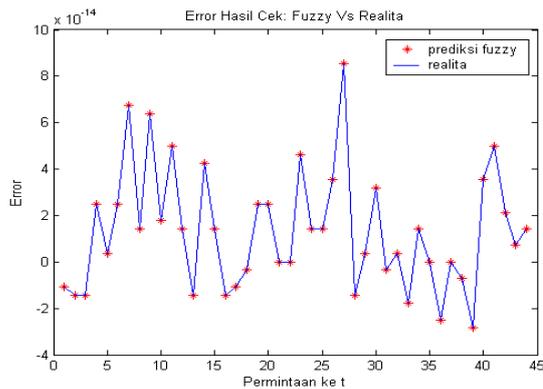
Pengetesan sistem fuzzy dengan data pengecekan yang terdapat dalam tabel 2 (Lampiran 1) berupa peramalan permintaan untuk data ke : 173 s/d 216. Hasil pengetesan dengan jari-jari = 0.50, *accept ratio* = 0.5 dan *reject ratio* = 0.15, seperti terlihat pada berikut ini:



**Gambar 4.10: Plot hasil pengetesan peramalan terhadap realita**

Tampilan error secara grafis terlihat pada Gambar 4.11. Error terbesar terjadi pada data pengetesan yang ke-134 yaitu sebesar 0.852e-013, sedangkan error terkecil terjadi pada data pengetesan yang ke-39 yaitu sebesar -0.2842 e-013.

Berdasarkan tabel 4 pada lampiran 3 didapatkan hasil validasi jumlah kuadrat kesalahan (SSE) untuk data pengetesan yaitu 3.616627719e-013.



**Gambar 4.11: Grafik error hasil cek peramalan terhadap realita**

Berdasarkan hasil diatas k-means clustering pada data inferensi menghasilkan nilai RMSE = 4.2624, error maksimum = 14.6175, dan error minimum = -14.3349, sedangkan pada data pengetesan kmeans clustering menghasilkan error maksimum = 15.7584, dan error minimum = -9.9222. Kmeans clustering termasuk hard clustering, oleh karena itu perlu dilakukan normalisasi untuk membandingkan hasil tersebut dengan hasil gath-geva clustering yang termasuk fuzzy clustering sehingga tidak perlu dilakukan normalisasi.

Rumus penormalisasian berdasar-kan dasar teori diatas yaitu: 
$$\bar{u} = \frac{u - u_{\min}}{u_{\max} - u_{\min}}$$

dengan  $\bar{u}$  adalah data hasil normalisasi,  $u$  adalah data sebelum dinormalisasikan,  $u_{\min}$  adalah nilai terkecil yang terukur dan  $u_{\max}$  adalah nilai terbesar yang terukur. Diperoleh hasil normalisasi dari hasil kmeans clustering adalah:

- Hasil inferensi

$$error_{\max} = \frac{14.6175 - 13.1}{53.1 - 13.1} = \frac{1.5175}{40} = 0.0379375$$

$$error_{\min} = \frac{-14.3349 - 13.1}{53.1 - 13.1} = \frac{-27.4349}{40} = -0.6858725$$

- Hasil pengetesan

$$error_{\max} = \frac{15.7584 - 13.1}{53.1 - 13.1} = \frac{2.6584}{40} = 0.06646$$

$$error_{\min} = \frac{-9.9222 - 13.1}{53.1 - 13.1} = \frac{-23.0222}{40} = -0.575555$$

## 5. Simpulan dan Saran

### 5.1 Simpulan

Berdasarkan hasil analisa dan pembahasan sebelumnya, maka dapat disimpulkan beberapa hal yaitu:

1. Berdasarkan nilai validasi K-means clustering dihasilkan cluster optimal terjadi pada saat cluster = 4 dengan pusat cluster adalah:

$$C = \begin{pmatrix} 0.1382 & 0.1200 \\ 0.4438 & 0.3536 \\ 0.6693 & 0.6845 \\ 0.8722 & 0.3792 \end{pmatrix}$$

dan hasil validasi pada saat cluster = 4, diperoleh nilai PC (Partition Coeffisient) = 1, CE (Classification Entropy) = NaN (bernilai sangat kecil sekali sehingga tidak dituliskan nilainya), SC (Partition Index) = 0.5672, S (Separation Index) = 0.0040 dan SSE (Jumlah Kuadrat Kesalahan) pada hasil inferensi = 20.09905621. sedangkan SSE pada hasil pengetesan = 4.620844103.

2. Berdasarkan nilai validasi Gath-Geva (GG) clustering dihasilkan cluster optimal terjadi pada saat cluster = 4 dengan pusat cluster adalah:

$$C = \begin{pmatrix} 0.1101 & 0.0869 \\ 0.2295 & 0.2425 \\ 0.6323 & 0.3761 \\ 0.6950 & 0.7358 \end{pmatrix}$$

dan hasil validasi pada saat cluster = 4, diperoleh nilai PC (Partition Coeffisient) = 0.9607, CE (Classification Entropy) = NaN (bernilai sangat kecil sekali sehingga tidak dituliskan nilainya), SC (Partition

- Index) = NaN, S (Separation Index) = NaN dan SSE (Jumlah Kuadrat Kesalahan) pada hasil inferensi = 10869.45645. sedangkan SSE pada hasil pengetesan = 3.61662719e-013.
3. Dengan jari-jari 0.5 dan jumlah cluster = 4, K-means Clustering menghasilkan RMSE = 4.2624, sedangkan Gath-Geva (GG) Clustering menghasilkan RMSE = 2.8668e-14.
  4. Dengan jumlah jari-jari dan cluster yang sama, K-means Clustering menghasilkan error cek maksimum = 0.06646 dan error cek minimum = -0.575555, sedangkan Gath-Geva (GG) Clustering menghasilkan error cek maksimum = 0.8527e-013, dan error cek minimum = -0.2842e-013.
  5. Berdasarkan hasil RMSE, error cek maksimum dan error cek minimum, dapat disimpulkan bahwa Gath-Geva (GG) Clustering lebih teliti dibandingkan dengan metode K-means clustering.
- 5.2 Saran**
- Berdasarkan dari kesimpulan di atas maka dapat diberikan saran-saran sebagai berikut: 1) Algoritma GG Clustering dapat digunakan dalam pengklasteran data time series. 2) Ada beberapa metode clustering yang lainnya yang masih perlu dianalisis untuk mengetahui tingkat ketelitannya. 3) Masih perlu dianalisis untuk mengetahui tingkat ketelitiannya diantara kedua metode tersebut.
- Referensi**
- [1] Babuska, R.; P,J van der Veen; U. Kaymak; 2002, *Improved Covariance Estimation for Gustafson-Kessel Clustering*, Prodeccings of 2002 IEEE International Conference on Fuzzy System Honolulu, Hawaii.
  - [2] Babuska, R., -, *Fuzzy Clustering with Applications in Patter Recognition and Data-Driven Modeling*; Delft Center for System and Control, www.dscs.tudelft.nl/~babuska, 5 April 2006.
  - [3] Balasko, B.; J. Abonyi; B. Feil; -; *Fuzzy Clustering and Data Analysis Toolbox*, Departement of Process Engineering Univ. of Veszprem, Hungary, [www.fmt.vein.hu](http://www.fmt.vein.hu), 5 April 2006.
  - [4] Biro Pusat Statistik Indonesia, - , *Indikator Ekonomi Januari 1987-Desember 2005*, Jakarta, Indonesia.
  - [5] Biro Pusat Statistik Indonesia, 2001, *Ekspor Indonesia menurut kode ISIC*, CV Nario Sari, Jakarta, Indonesia.
  - [6] Chen, A; Efstrations Nikolaidis, 1999, *Comparison of Probabilistic and fuzzy set Methods for Designing under Uncertainty*, American Institute of Aeronautics and Astronaitics, Vol. 99.
  - [7] Gunawan; 2006, *Perbandingan Antara Metode Fuzzy Subtractive Clustering dengan Metode Gustafson Kessel Clustering. (Studi kasus data time series volume pinjaman kredit gadai)*, Tesis, Matematika, FMIPA-ITS.
  - [8] Kusumadewi, Sri; 2002, *Analisis Desain Sistem Fuzzy Menggunakan Toolbox MATLAB*, Graha Ilmu, edisi pertama, cetakan pertama, Yogyakarta.
  - [9] Makridas.S, Steven C.W and Victor E.M, 1999, *Metode dan Aplikasi Peramalan*, Erlangga, edisi kedua, jilid 1, cetakan keenam, Jakarta.
  - [10] Yuniarti, D, W; (2005), *Prediksi Volume Ekspor Non-Migas dengan Metode Subtractive Clustering*, Skripsi, Jurusan Matematika, FMIPA-ITS.