

repository.ub.ac.id

# IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING DALAM PEMBANGKITAN ATURAN FUZZY PADA PERENCANAAN KONSUMSI PANGAN HARIAN

Novinda Fiqih Caesandria<sup>1</sup>, Candra Dewi, S.Kom, M.Sc<sup>2</sup>, Edy Santoso, S.Si, M.Kom<sup>2</sup>

<sup>1</sup>Mahasiswa<sup>2</sup>Dosen Pembimbing

Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya, Malang 65145, Indonesia

Email : [novcsndr42@gmail.com](mailto:novcsndr42@gmail.com)<sup>1</sup>, [dewi\\_candra@ub.ac.id](mailto:dewi_candra@ub.ac.id)<sup>2</sup>, [edy144@ub.ac.id](mailto:edy144@ub.ac.id)<sup>2</sup>

## ABSTRAK

Gizi seimbang dapat dicapai melalui perencanaan konsumsi pangan harian yang dapat dilakukan oleh seorang ahli gizi namun saat ini keberadaan tenaga ahli gizi masih belum merata. Terdapat studi yang bisa mengatasi permasalahan tersebut dan studi tersebut adalah sistem pakar penalaran *fuzzy*. Pembuatan sistem pakar tidak lepas dari pengetahuan pakar yang dalam sistem berupa aturan. Aturan dapat dibangkitkan secara otomatis dengan metode *clustering*. Metode yang digunakan untuk membangkitkan aturan *fuzzy* adalah *K-means clustering* yang dalam sistem menjadi proses pelatihan untuk membentuk aturan *fuzzy* sedangkan *Fuzzy Takagi Sugeno Kang* menjadi mesin inferensi. Hasil *Mean Absolute Percentage Error* (MAPE) terkecil pada penelitian ini adalah pada laki-laki 22,55% sedangkan pada perempuan sebesar 11,49%. Hasil penelitian ini menunjukkan bahwa *cluster* ideal atau *cluster* dengan nilai varian terkecil belum tentu menghasilkan nilai terbaik.

**Kata Kunci :** *K-means Clustering, fuzzy-Takagi Sugeno Kang, pembangkitan aturan.*

## ABSTRACT

*Balanced nutrition can be achieved through planning daily food intake and can be done by a nutritionist, but now the existence of nutrition experts are still not evenly distributed. There is study that can overcome these problem and the study is expert system fuzzy reasoning. Making expert systems can not be separated from the knowledge of experts and in the system it represent by rules. The current rules can be generated automatically with a clustering method. The method that used to generate rule is a K-means clustering that used to be the training process to establish rules while Fuzzy Takagi Sugeno Kang became inference engine. In this study the smallest Mean Absolute Percentage Error (MAPE) results are male 22,55%, while in women by 11,49%. Result of this study shows that cluster ideal or cluster with smallest varians value not necessarily result a best value.*

**Keywords:** *K-means Clustering, fuzzy-Takagi Sugeno Kang, rule generate.*

## 1. PENDAHULUAN

Program Indonesia Sehat bertujuan meningkatkan derajat kesehatan dan status gizi masyarakat Indonesia dimana salah satu strategi untuk mencapainya dengan mengupayakan peningkatan tenaga kesehatan di 5.600 puskesmas yang salah satunya yaitu tenaga ahli gizi. (Sekretariat Jenderal Kementerian Kesehatan RI, 2015). Peningkatan status gizi sangat berkaitan dengan Angka Kecukupan Gizi (AKG) yang didasarkan pada golongan umur, jenis kelamin, ukuran tubuh dan aktivitas tubuh. Gizi seimbang

dapat dicapai melalui perencanaan konsumsi pangan harian yang disusun oleh seorang ahli gizi agar tidak terjadi kesalahan.

Terdapat studi untuk mengatasi masalah gizi seimbang yaitu dengan menggunakan sistem pakar penalaran *fuzzy* yang dimana pengetahuan pakar dapat diadopsi ke dalam sistem. Dalam sistem pakar penalaran *fuzzy* diperlukan suatu aturan yang merupakan representasi pengetahuan pakar yang didapat melalui proses akuisisi pengetahuan. Terdapat perbedaan pengetahuan antara satu pakar



dengan lainnya, sehingga dapat menyebabkan kemungkinan pengetahuan yang diimplementasikan dalam sistem tidak lengkap untuk itu saat ini muncul suatu metode untuk membangkitkan aturan secara otomatis berdasarkan data yang sudah ada. Adanya hal tersebut membuat proses pembentukan aturan sangat efisien dibandingkan melakukan akuisisi pengetahuan secara manual.

Pada penelitian berjudul “A Clustering based Genetic Fuzzy expert system for electrical energy demand prediction” membahas masalah sistem pakar peramalan permintaan energi listrik. Pada penelitian tersebut digunakan metode *K-means Clustering* sebagai pembangkit aturan dan *Genetic Fuzzy System* (GFS) sebagai mesin inferensinya. Pada penelitian ini menghasilkan hasil yang lebih akurat dibanding dengan metode *Adaptive Neuro-Fuzzy Inference Systems* (ANFIS) dan metode yang hanya mengimplementasikan GFS tanpa pembangkitan aturan otomatis (Ghanbari dkk, 2010). Terdapat penelitian lain berjudul “Implementasi Metode K-Means Clustering untuk Pembangkitan Aturan Fuzzy pada Klasifikasi Ketahanan Hidup Penderita Kanker Payudara” yang membahas masalah klasifikasi ketahanan hidup penderita kanker payudara. Pada penelitian ini diimplementasi metode *K-means Clustering* yang bertidak sebagai pembangkitan aturan dan metode *Fuzzy Inference System Takagi Sugeno Kang* berguna untuk mengklasifikasi. Hasil dari penelitian ini adalah pembangkitan aturan berhasil dilakukan dengan nilai akurasi tertinggi adalah 83% (Sholeh, 2013).

Berdasarkan penelitian diatas dibuat suatu perencanaan konsumsi pangan harian yang berfokus pada kebutuhan energi atau kalori tiap individu. Pada penelitian ini diimplementasikan metode *Fuzzy Inference System Takagi Sugeno Kang* dan *K-means Clustering* sebagai pembangkitan aturan fuzzy secara otomatis. Pada penelitian ini perhitungannya didasarkan pada atribut jenis kelamin, usia, berat badan, tinggi badan, dan aktivitas fisik.

## 2. TINJAUAN PUSTAKA

### 2.1 Gizi

Gizi berasal dari kata *ghidza* dalam bahasa arab yang berarti makanan. Sekitar tahun 1952-1955 istilah gizi mulai dikenal Indonesia. Sebenarnya gizi merupakan zat dalam makanan yang berguna bagi kesehatan makhluk hidup. Penelitian terdahulu membuktikan bahwa dengan mengkonsumsi makanan secara teratur dan dalam jumlah yang cukup maka akan meningkatkan kemampuan fisik dan mempertahankan kesehatan tubuh. Susunan konsumsi pangan dikatakan cukup bagi seseorang bila jumlah zat gizi yang diperoleh dari pangan yang dikonsumsi tersebut memenuhi kecukupan tubuh akan masing-masing zat gizi. Diperlukan kegiatan perencanaan konsumsi pangan untuk mengetahui susunan konsumsi pangan yang memenuhi

kebutuhan dan kecukupan gizi (Briawan dan Hardiyansyah, 1990).

Faktor yang mempengaruhi dalam perencanaan konsumsi pangan:

1. Jenis Kelamin  
Jenis kelamin laki-laki pada umumnya memiliki kebutuhan energi yang lebih besar dibandingkan perempuan. Hal tersebut disebabkan oleh proses metabolisme laki-laki lebih tinggi dibanding perempuan.
2. Usia  
Seiring bertambahnya usia proses metabolisme dan kinerja organ tubuh menurun oleh karenanya nilai Energi Metabolisme Basal (EMB) akan semakin berkurang.
3. Indeks Masa Tubuh (IMT)  
IMT digunakan sebagai pengukuran lemak tubuh, bila semakin rendah nilai IMT seseorang maka orang tersebut membutuhkan energi yang lebih banyak.
4. Aktivitas Fisik  
Energi yang digunakan untuk aktivitas fisik bergantung pada jenis, intensitas dan durasi dalam melakukan aktivitas fisik (Direktorat Jenderal Bina Gizi dan Kesehatan Ibu dan Anak, 2011). Range mengenai aktivitas akan dijelaskan pada Tabel 1.

Tabel 1. Nilai Aktivitas Fisik

Keterangan	Nilai Aktivitas Harian
Tidak ada olahraga	1.2
Olahraga ringan (1–3 hari per minggu)	1.375
Olahraga Sedang (3–5 hari per minggu)	1.55
Olahraga berat (6–7 hari per minggu)	1.725
Olahraga sangat berat (2 kali dalam sehari di per minggu)	1.9

Sumber: Wawancara Pakar

### 2.2 Fuzzy Inference System Sugeno

Penalaran metode *Sugeno* memiliki kemiripan dengan metode Mamdani, perbedaannya terletak pada *output* (konsekuen) sistem yang tidak berupa himpunan *Fuzzy*, melainkan berupa konstanta atau persamaan linear. Metode *Sugeno* yang digunakan adalah model *Fuzzy Sugeno* orde-satu yang secara umum berbentuk sebagai Persamaan (1).

$$IF(x_1 \text{ is } A_1) \circ \dots \circ (x_n \text{ is } A_n) \text{ THEN } z = p_1 * x_1 + \dots + p_n * x_n + q \quad (1)$$

Dengan  $A_i$  adalah himpunan *Fuzzy* ke- $i$  sebagai anteseden dan  $p_i$  adalah suatu konstanta (tegas) ke- $i$



dan  $q$  juga merupakan konstanta dalam konsekuen (Kusumadewi dan Purnomo, 2010).

### 2.3 K-Means Clustering

Algoritma *k-means* merupakan salah satu yang paling sering digunakan karena algoritmanya sangat sederhana dan efisien. Secara garis besar tahapan awal dari algoritma *k-means clustering* ini adalah membagi data menjadi beberapa *cluster*, dimana pemilihan *cluster* bisa secara acak. Membagi data dan menetapkan pusat data *cluster* (*centroid*), Menghitung ulang *centroid cluster*, kemudian tahapan awal dan kedua diulang-ulang agar dapat mencapai *cluster* yang optimal (Fielding, 2007).

Metode *k-means* dikembangkan pada tahun 1967 oleh Mac Queen. *K-Means Clustering* memiliki kemampuan untuk mengkluster data yang besar dan kompleksitas waktunya linear. Pada *K-Means*  $n$  adalah jumlah data,  $k$  adalah jumlah kluster, dan  $t$  adalah jumlah iterasi. *K-means* merupakan metode pengklusteran secara *partitioning* yang berarti memisahkan data ke dalam kelompok yang berbeda. Oleh karenanya, *K-Means* mampu meminimalkan rata-rata jarak setiap data ke setiap *cluster*. Untuk lebih jelasnya lagi berikut tahapan klusterisasi *K-Means* (Sholeh, 2013).

1. Tentukan nilai  $k$  sebagai jumlah kluster yang ingin dibentuk.
2. Inisialisasi  $k$  *centroid* (titik pusat kluster) awal secara random.
3. Hitung jarak setiap data ke masing-masing *centroid* menggunakan rumus korelasi antar dua objek yaitu *Euclidean Distance* dan kesamaan *Cosine*.
4. Berikut Persamaan (2) adalah Rumus *Euclidean Distance* untuk menghitung jarak dari titik  $x$  dan  $y$ . Titik  $x$  menyatakan data *centroid* ke  $I$  dan  $y$  menyatakan data latih ke  $y$ .  

$$d(x, y) = ||x - y||^2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$
5. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroid*nya.
6. Tentukan posisi *centroid* baru ( $k$ ) dengan cara menghitung nilai rata-rata dari data yang ada pada *centroid* yang sama menggunakan Persamaan (3) dimana  $\mu$  = rata-rata (*mean*),  $x_n$  = data ke- $n$ , dan  $n$  = jumlah data.  

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \quad (3)$$
7. Kembali ke langkah 3 jika posisi kelompok data pada *centroid* baru dengan *centroid* lama tidak sama.

### 2.4 Analisis Varian

*Cluster* yang baik adalah *cluster* yang memiliki homogenitas dan heterogenitas yang tinggi. Dari hal tersebut maka bisa dilakukan pendekatan untuk mengetahui baik atau tidaknya suatu *cluster* berdasarkan nilai varian (penyebaran dari data). Varian dalam *clustering* ada dua, yaitu varian dalam *cluster* (*variance within cluster/  $V_w$* ) yang merepresentasikan internal homogeneity dan varian

antar *cluster* (*variance between cluster/  $V_b$* ) yang merepresentasikan external homogeneity (Larose, 2015). Kedua nilai varian tersebut dapat menentukan kepadatan suatu *cluster* (Ilham, 2011).

*Variance cluster* ( $V_c$ ) ditentukan dengan Persamaan (4). dimana  $V_c^2 = V_i^2$  atau varian pada *cluster*,  $c = 1 \dots k$  ( $k$  = jumlah *cluster*),  $n_c$  = jumlah data pada *cluster*,  $d_i$  = data ke- $i$  pada suatu *cluster*, dan  $\bar{d}_c$  = rata-rata data pada suatu *cluster*.

$$V_c^2 = \frac{1}{n_c - 1} \sum_{i=1}^{n_c} (d_i - \bar{d}_c)^2 \quad (4)$$

Mencari kerapatan suatu *cluster* dengan menghitung nilai varian dalam *cluster* atau *variance within cluster* ( $V_w$ ) yang dapat dihitung dengan Persamaan (5) dimana  $V_w$  = *variance within cluster*,  $N$  = jumlah semua data,  $k$  = jumlah *cluster*,  $n_i$  = jumlah data pada *cluster* ke- $i$ , dan  $V_i^2$  = varian pada *cluster* ke- $i$

$$V_w = \frac{1}{N - k} \sum_{i=1}^k (n_i - 1) V_i^2 \quad (5)$$

Varian diantara *cluster* atau *variance between cluster* ( $V_b$ ) dapat dihitung dengan Persamaan (6) dimana  $\bar{d}_i$  = rata-rata data pada *cluster* ke- $i$  dan  $\bar{d}$  = rata-rata dari  $d_i$ .

$$V_b = \frac{1}{k - 1} \sum_{i=1}^k n_i (\bar{d}_i - \bar{d})^2 \quad (6)$$

*Cluster* yang ideal ditentukan dengan nilai batasan varian yang terkecil, semakin kecil nilai varian, maka semakin ideal *cluster* tersebut. Nilai batasan varian yang baik mempunyai  $V_w$  minimum dan  $V_b$  maksimum, Nilai batasan varian ( $V$ ) dapat dihitung dengan Persamaan (7).

$$V = \frac{V_w}{V_b} \quad (7)$$

Menghitung standar deviasi masing-masing *cluster* menggunakan Persamaan (8) dimana  $k$  = jumlah *cluster* 1... $k$ ,  $i$  = atribut 1.. $i$ ,  $n$  = jumlah data,  $x_j$  = data ke  $j$ ,  $\mu$  = rata-rata *cluster*.

$$\sigma_{ki} = \sqrt{\frac{\sum_{j=1}^n (x_j - \mu_{ki})^2}{n - 1}} \quad (8)$$

### 2.5 Fungsi Keanggotaan

Fungsi keanggotaan atau *membership function* merupakan kurva yang menunjukkan pemetaan titik-titik *input* data ke dalam nilai keanggotaannya (derajat keanggotaan) yang memiliki nilai interval antara 0 sampai 1. Salah satu cara yang dapat digunakan untuk mendapatkan nilai keanggotaan adalah dengan melalui pendekatan fungsi yang salah satunya adalah Fungsi keanggotaan *Gaussian* yang menggunakan parameter  $\{c, \sigma\}$  dengan Persamaan (9) dimana  $x$  = data,  $c$  = pusat *cluster* dan  $\sigma$  = standar deviasi.

$$\text{gaussian}(x; c, \sigma) = e^{-\frac{1}{2} \left(\frac{x-c}{\sigma}\right)^2} \quad (9)$$

### 2.6 Least Square Estimation

Mencari nilai *antecedent* dan *consequent* dalam inferensi *Takagi Sugeno Kang* dilakukan dengan menggunakan metode kuadrat terkecil. Metode *least square estimator* merupakan metode



untuk menentukan metode kuadrat terkecil. Langkah pertama yang dilakukan yaitu identifikasi struktur agar dapat menerapkan pengetahuan tentang target sistem yang bertujuan untuk menentukan kelas yang paling cocok (Jang dkk, 1997).

### 2.7 Ekstraksi Aturan Fuzzy

Ekstraksi aturan fuzzy menggunakan metode inferensi fuzzy Takagi Sugeno Kang orde-satu. Sebelum mengekstraksi, data pada variabel input dipisah dengan data pada variabel output. Misal jumlah variabel input adalah  $m$  dan variabel output adalah 1, sehingga akan diperoleh kumpulan aturan yang berbentuk (Kusumadewi dan Purnomo, 2010) pada Persamaan (10).

[R1] IF  $(x_1 \text{ is } A_{11}) \circ (x_2 \text{ is } A_{12}) \circ \dots \circ (x_n \text{ is } A_{1m})$   
THEN  $(z = k_{11}x_1 + \dots + k_{1m}x_m + k_{10})$ ;

[R2] IF  $(x_1 \text{ is } A_{21}) \circ (x_2 \text{ is } A_{22}) \circ \dots \circ (x_n \text{ is } A_{2m})$   
THEN  $(z = k_{21}x_1 + \dots + k_{2m}x_m + k_{20})$ ;

...

[Rr] IF  $(x_1 \text{ is } A_{r1}) \circ (x_2 \text{ is } A_{r2}) \circ \dots \circ (x_n \text{ is } A_{rm})$   
THEN  $(z = k_{r1}x_1 + \dots + k_{rm}x_m + k_{r0})$ ;  
(10)

Keterangan:

- $A_{ij}$  adalah himpunan fuzzy aturan ke- $i$  variabel ke- $j$  sebagai anteseden,
- $k_{ij}$  adalah koefisien persamaan output fuzzy aturan ke- $i$  variabel ke- $j$  ( $i = 1, 2, \dots, r$ ;  $j = 1, 2, \dots, m$ ), dan  $k_{i0}$  adalah konstanta persamaan output fuzzy aturan ke- $i$ ,
- o menunjukkan operator yang digunakan

Nilai koefisien output tiap-tiap aturan pada setiap variabel ( $k_{ij}$ ,  $i = 1, 2, \dots, r$ ; dan  $j = 1, 2, \dots, m + 1$ ) didapatkan dengan metode kuadrat terkecil dikarenakan matriks  $U$  bukan merupakan matriks bujursangkar. Dalam membentuk anteseden, setiap variabel input akan terbagi menjadi  $r$  himpunan fuzzy, dengan setiap himpunan memiliki fungsi keanggotaan Gaussian. Aturan yang diekstraksi ditunjukkan dengan Persamaan (11).

[R1]: IF  $(X_{i1} \text{ is } V1H1) \circ (X_{i2} \text{ is } V2H1) \circ \dots \circ (X_{im} \text{ is } VmH1)$   
THEN  $Y = Z_1$

[R2]: IF  $(X_{i1} \text{ is } V1H2) \circ (X_{i2} \text{ is } V2H2) \circ \dots \circ (X_{im} \text{ is } VmH2)$   
THEN  $Y = Z_2$

[R3]: IF  $(X_{i1} \text{ is } V1H3) \circ (X_{i2} \text{ is } V2H3) \circ \dots \circ (X_{im} \text{ is } VmH3)$   
THEN  $Y = Z_3$

...

[Rr]: IF  $(X_{i1} \text{ is } V1Hr) \circ (X_{i2} \text{ is } V2Hr) \circ \dots \circ (X_{im} \text{ is } VmHr)$

$$\text{THEN } Y = Z_r \quad (11)$$

$V_p H_q$  adalah variabel ke- $p$  himpunan ke- $q$ .

### 2.8 Mean Absolute Percentage Error (MAPE)

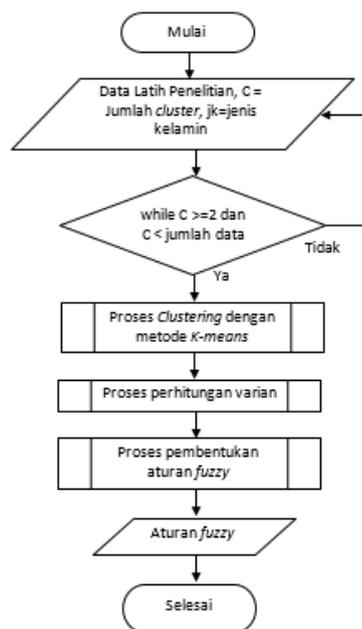
Salah satu metode statistika yang digunakan mengukur tingkat kesalahan adalah MAPE (*Mean Absolute Percentage Error*) yang dirumuskan pada Persamaan (12). Nilai  $N$  disini berarti jumlah data yang diobservasi Hasil yang paling baik adalah yang memiliki nilai persentase error terkecil (Albright, 2016).

$$\text{MAPE} = \frac{\sum_{t=1}^N \frac{|E_t|}{Y_t}}{N} \times 100\% \quad (12)$$

Terdapat target performa dalam MAPE dimana jika MAPE lebih kecil dari 20% maka hasil dapat dikatakan baik. Sedangkan jika MAPE lebih kecil dari 10% maka hasil dapat dikatakan sangat baik. Meskipun demikian hal tersebut tidak menjadi patokan utama karena dalam menghitung performa MAPE perlu dilihat juga target hasil peramalan dengan nilai asli apakah sesuai dengan yang diharapkan atau tidak (Gilliland, 2010).

## 3. METODE

Pada penelitian ini metode yang digunakan adalah *K-Means Clustering* dan FIS Sugeno serta terdapat dua buah proses besar yaitu pelatihan dan pengujian. Perancangan sistem yang menjabarkan proses pelatihan untuk perencanaan konsumsi pangan harian ditunjukkan pada Gambar 1.



Gambar 1. Diagram alir proses pelatihan

Pada proses pelatihan Gambar 1 terdapat sub proses yang dapat dijelaskan sebagai berikut:

1. Proses *K-Means Clustering*

- Pada iterasi pertama tentukan *centroid* awal sebanyak jumlah *cluster* atau aturan yang akan dibentuk.
  - Menghitung jarak *euclidean* tiap data dengan *centroid* awal
  - Mengelompokkan data ke dalam *cluster* sesuai dengan jarak *euclidean* terkecil.
  - Pada iterasi kedua menghitung nilai *centroid* baru yang ada tiap *cluster* per atribut dengan merata-ratakan semua nilai.
  - Dilakukan kembali perhitungan jarak *euclidean* tiap data dengan *centroid* baru.
  - Dilakukan pengelompokkan data kembali lalu dibandingkan apakah ada anggota *cluster* yang berubah. Iterasi dihentikan apabila tidak ada perubahan anggota cluster.
2. Proses Menghitung Varian
    - Perhitungan selanjutnya yaitu perhitungan nilai varian *cluster* ( $V_c^2$ ), varian *within* ( $V_w$ ), varian *between* ( $V_b$ ) dengan persamaan berikut.
    - Kemudian dari sana akan dihitung nilai varian yang dihitung sesuai dengan persamaan berikut.
  3. Proses Pembentukan Aturan Fuzzy
    - Menghitung standar deviasi disetiap *cluster* per atribut.
    - Hitung derajat keanggotaan tiap data di tiap *cluster* yang dibentuk dengan menggunakan Gaussian.
    - Menghitung koefisien output dengan Least Square Error (LSE).
    - Dilakukan ekstraksi aturan fuzzy sesuai bentuk FIS Takagi Sugeno Kang orde-satu.

1. Proses Pemilihan Varian Terkecil
  - Memilih aturan dengan nilai varian yang paling kecil dari seluruh aturan yang dibangkitkan.
2. Proses FIS Takagi Sugeno
  - Menghitung derajat keanggotaan input data yang dalam perhitungannya menggunakan nilai *centroid* dan juga nilai standar deviasi.
  - Selanjutnya dilanjutkan untuk menghitung  $\alpha$  predikat.
  - Menghitung nilai Z pada setiap aturan yang terbentuk dan dihitung menggunakan koefisien p yang terbentuk pada proses perhitungan koefisien *output*.
  - Menghitung nilai Z akhir (defuzzifikasi) dengan perhitungan nilai bobot dengan menggunakan weighted average.

#### 4. HASIL PENGUJIAN DAN ANALISIS

Berikut adalah analisis yang telah dilakukan dengan scenario pengujian pada tahap pertama yaitu *cluster* ideal pada setiap jenis kelamin di dalam tiga variasi data latih (50,60,70). Tahap kedua adalah pengujian MAPE pada setiap jenis kelamin di dalam tiga variasi data latih (50,60,70).

##### 4.1 Analisis Hasil Aturan

Pada analisis aturan diambil contoh hasil pada parameter data 50 di setiap jenis kelamin dimana *cluster* ideal yang dihasilkan pada laki-laki adalah 11 dan perempuan adalah 7. Pada setiap parameter dilakukan 3 kali percobaan yang menghasilkan aturan, nilai varian, dan MAPE yang berbeda-beda yang digambarkan sesuai dengan Tabel 2.

Proses pengujian digambarkan pada Gambar 2.



Gambar 2. Diagram alir proses pengujian

Pada proses pengujian Gambar 2 terdapat sub proses yang dapat dijelaskan sebagai berikut:

Tabel 2. Hasil Analisis Aturan Percobaan di Cluster Ideal

Aturan Percobaan ke-	Jenis Kelamin	Nilai Varian	MAPE
1	Laki-laki	0.0029569 69	33.1%
2	Laki-laki	0.0022591 76	30.08%
3	Laki-laki	0.0029814 79	25.5%
1	Perempuan	0.0054716 03	20.36%
2	Perempuan	0.0065963 37	19.83%
3	Perempuan	0.0048030 92	15.49%

Berdasarkan pada Tabel 2 didapat bahwa dari 3 kali percobaan di setiap parameter inputan yang



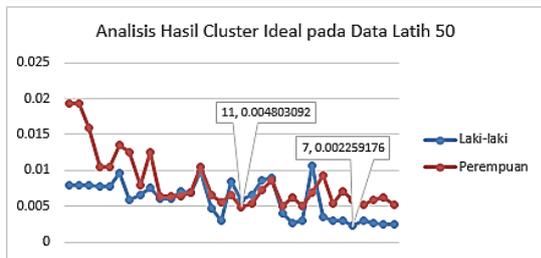
ada menghasilkan hasil yang berbeda baik dari segi aturan, batasan varian, dan hasil MAPE. Hal tersebut disebabkan oleh penentuan *centroid* awal dalam K-Means yang dilakukan secara random sehingga menyebabkan hasil aturan yang berbeda-beda.

#### 4.2 Analisis Hasil Cluster Ideal pada Data Latih

Hasil *cluster* ideal pada setiap jenis kelamin yang berkaitan dengan nilai varian terhadap data latih 50, 60, dan 70.

##### 4.2.1 Analisis Hasil Cluster Ideal pada Data Latih 50

Setelah melakukan pengujian *cluster* ideal, dianalisislah keseluruhan pengujian *cluster* ideal pada masing-masing jenis kelamin terhadap data latih 50 yang digambarkan pada Gambar 3.

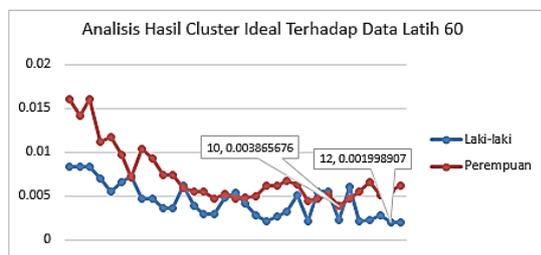


Gambar 3 Analisis hasil *cluster* ideal pada data latih 50

Pada Gambar 3 menunjukkan pergerakan batasan varian mulai dari jumlah *cluster* 2 hingga 12. Pada grafik tersebut, nilai batasan varian minimum untuk jenis kelamin laki-laki terjadi pada *cluster* 11 ( $V=0.002259176$ ), sedangkan untuk jenis kelamin perempuan terjadi pada *cluster* 7 ( $V=0.004803092$ ).

##### 4.2.2 Analisis Hasil Cluster Ideal pada Data Latih 60

Setelah melakukan pengujian *cluster* ideal, pada bagian ini dianalisislah keseluruhan pengujian *cluster* ideal pada masing-masing jenis kelamin terhadap data latih 60, yang dapat digambarkan pada Gambar 4.



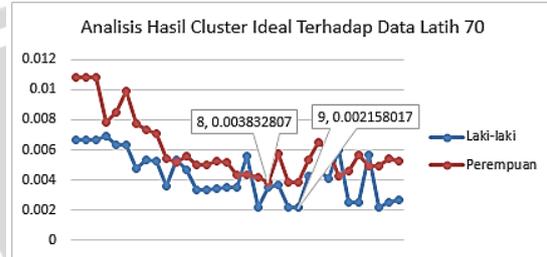
Gambar 4. Analisis hasil *cluster* ideal terhadap data latih 60

Pada Gambar 4 menunjukkan pergerakan batasan varian mulai dari jumlah *cluster* 2 hingga 12. Pada grafik tersebut, nilai batasan varian minimum untuk jenis kelamin laki-laki terjadi pada *cluster* 12 ( $V=0.002259176$ ), sedangkan untuk jenis

kelamin perempuan terjadi pada *cluster* 10 ( $V=0.003865676$ ).

##### 4.2.3 Analisis Hasil Cluster Ideal pada Data Latih 70

Setelah melakukan pengujian *cluster* ideal, dianalisislah keseluruhan pengujian *cluster* ideal pada masing-masing jenis kelamin terhadap data latih 70, yang dapat digambarkan pada Gambar 5.



Gambar 5 Analisis hasil *cluster* ideal terhadap data latih 70

Pada Gambar 5 menunjukkan pergerakan batasan varian mulai dari jumlah *cluster* 2 hingga 12. Pada grafik tersebut, nilai batasan varian minimum untuk jenis kelamin laki-laki terjadi pada *cluster* 9 ( $V=0.00215802$ ), sedangkan untuk jenis kelamin perempuan terjadi pada *cluster* 8 ( $V=0.003832807$ ).

#### 4.3 Analisis Hasil MAPE

Analisis hasil MAPE berguna untuk melihat persentase rata-rata *error* pada setiap laki-laki dan perempuan. Dianalisislah parameter (*cluster* dan jumlah data) yang memiliki nilai MAPE paling kecil yang dalam hal ini berarti aturan tersebut memiliki hasil yang lebih optimal.

##### 4.3.1 Analisis Hasil MAPE pada Laki-laki

Analisis hasil MAPE pada laki-laki dilakukan pada seluruh jumlah data latih untuk mengetahui nilai persentase rata-rata *error* paling kecil dari pengujian yang telah dilakukan sebelumnya. Hasil MAPE pada laki-laki di tampilkan pada Tabel 3.

Tabel 3 Hasil MAPE pada Laki-laki

50 Data latih		60 Data latih		70 data Latih	
Jumlah cluster	Rata-rata MAPE	Jumlah cluster	Rata-rata MAPE	Jumlah cluster	Rata-rata MAPE
8	34.19%	8	28.18%	7	33.40%
9	27.64%	9	29.57%	8	32.57%
10	28.24%	10	27.37%	9	27.56%
11	30.08%	11	24.64%	10	33.47%
12	30.38%	12	33.11%	11	31.57%
Rata-rata	30.11%	Rata-rata	28.57%	Rata-rata	31.71%
MIN	27.64%	MIN	24.64%	MIN	27.56%



Berdasarkan Tabel 3 terlihat bahwa pada jumlah data latih 50 nilai persentase MAPE terkecil tidak berada pada jumlah *cluster* ideal yaitu 11 yang menghasilkan persentase 30.8%. Nilai MAPE terkecil berada pada jumlah *cluster* 9 yang persentasenya 27.64%. Pada jumlah data latih 60 nilai MAPE terkecil tidak berada pada jumlah *cluster* ideal yaitu 12 yang menghasilkan persentase 33.11%. Nilai MAPE terkecil berada pada jumlah *cluster* 11 yang persentasenya 24.64%. Pada jumlah data latih 70 nilai MAPE terkecil berada pada jumlah *cluster* ideal yaitu 9 yang menghasilkan persentase 27.56%. Dari adanya hal tersebut membuktikan bahwa penentuan *cluster* ideal dengan menggunakan analisis varian(batas varian terkecil) belum tentu menghasilkan nilai yang baik untuk outputnya.

#### 4.3.2 Analisis Hasil MAPE pada Perempuan

Analisis hasil MAPE pada perempuan dilakukan pada seluruh jumlah data latih untuk mengetahui nilai persentase rata-rata *error* paling kecil dari pengujian yang telah dilakukan sebelumnya. Hasil MAPE pada perempuan di tampilkan pada Tabel 4.

Tabel 4 Hasil MAPE pada Perempuan

50 Data latih		60 Data latih		70 data Latih	
Jumlah <i>cluster</i>	Rata-rata MAPE	Jumlah <i>cluster</i>	Rata-rata MAPE	Jumlah <i>cluster</i>	Rata-rata MAPE
5	20.28%	8	20.19%	6	25.1%
6	24.15%	9	24.8%	7	16.04%
7	15.49%	10	14.18%	8	16.81%
8	22.64%	11	16.42%	9	25.37%
9	18.69%	12	17.98%	10	22.3%
Rata-rata	20.25%	Rata-rata	18.71%	Rata-rata	21.12%
MIN	15.49%	MIN	14.18%	MIN	16.04%

Berdasarkan Tabel 4 terlihat bahwa pada jumlah data latih 50 nilai persentase MAPE terkecil berada pada jumlah *cluster* ideal yaitu 7 yang menghasilkan persentase 15.49%. Pada jumlah data latih 60 nilai MAPE terkecil berada pada jumlah *cluster* ideal yaitu 10 yang menghasilkan persentase 14.18%. Pada jumlah data latih 70 nilai MAPE terkecil tidak berada pada jumlah *cluster* ideal yaitu 8 yang menghasilkan persentase 16.81%. Nilai MAPE terkecil berada pada jumlah *cluster* 7 yang persentasenya 16.04%.

#### 4.3.3 Analisis Hasil MAPE pada Laki-laki dan Perempuan

Setelah melakukan 6 kali skenario percobaan pada setiap jumlah data latih di setiap jenis kelamin didapat persentase MAPE pada *cluster* ideal dan pada *cluster* sekitarnya. Melalui hasil tersebut dapat dilihat terdapat beberapa nilai

persentase MAPE terkecil tidak berada pada *cluster* ideal (varian terkecil). Hal tersebut digambarkan pada Tabel 5 dan 6.

Tabel 5 Hasil MAPE pada Laki-laki

Laki-laki				
J	J <sub>A</sub>	E <sub>A</sub>	J <sub>B</sub>	E <sub>B</sub>
1	11	30.08%	9	27.64%
2	12	33.11%	11	24.64%
3	9	27.56%	9	27.56%

Tabel 6 Hasil MAPE pada Perempuan

Perempuan				
J	J <sub>A</sub>	E <sub>A</sub>	J <sub>B</sub>	E <sub>B</sub>
1	7	15.49%	7	15.49%
2	10	14.18%	10	14.18%
3	8	16.81%	7	16.04%

J = Jumlah data (1 = 50 data, 2 = 60 data, dan 3= 70 data)

J<sub>A</sub>= Jumlah *cluster* hasil MAPE terbaik *cluster* ideal

E<sub>A</sub>= Rata-rata hasil MAPE terbaik *cluster* ideal

J<sub>B</sub>= Jumlah *cluster* hasil MAPE terkecil dari percobaan disekitar *cluster* ideal

E<sub>B</sub>= Rata-rata hasil MAPE terkecil dari percobaan disekitar *cluster* ideal

Berdasarkan Tabel 5 dan 6 dari 6 skenario percobaan MAPE disekitar *cluster* ideal didapatkan bahwa nilai *cluster* ideal yang menghasilkan nilai persentase terkecil berdasarkan pengujian disekitar *cluster* ideal ada 3 dari 6 percobaan. Hal ini membuktikan bahwa walaupun *cluster* ideal tidak selamanya menghasilkan persentase hasil yang baik namun teori tersebut cukup bisa dijadikan acuan dalam mekanisme pengujian.

Setelah beberapa tahapan pengujian dilakukan perbandingan hasil MAPE terbaik menurut penentuan *cluster* ideal dengan hasil MAPE terbaik menurut seluruh aturan yang dibangkitkan (189), penjabarannya ditunjukkan dalam Tabel 7 dan 8.

Tabel 7 Perbandingan Hasil MAPE Laki-laki

Laki-laki				
J	J <sub>A</sub>	E <sub>A</sub>	J <sub>B</sub>	E <sub>B</sub>
1	11	30.08%	4	22.55%
2	12	33.11%	11	24.64%
3	9	27.56%	6	24.14%
	Terkecil	27.56%	Terkecil	22.55%

Tabel 8. Perbandingan Hasil MAPE Perempuan

Perempuan				
J	J <sub>A</sub>	E <sub>A</sub>	J <sub>B</sub>	E <sub>B</sub>
1	7	15.49%	7	15.49%
2	10	14.18%	10	14.18%
3	8	16.81%	5	11.49%



Terkecil 14.18% Terkecil 11.49%

$J$ = Jumlah data (1 = 50 data, 2 = 60 data, dan 3= 70 data)

$J_A$ = Jumlah *cluster* hasil MAPE terbaik *cluster* ideal

$E_A$ = Rata-rata hasil MAPE terbaik *cluster* ideal

$J_B$ = Jumlah *cluster* hasil MAPE terbaik seluruh aturan

$E_B$ = Rata-rata hasil MAPE terbaik seluruh aturan

Berdasarkan Tabel 7 dan 8 tersebut didapatkan nilai jumlah *cluster* terbaik di setiap data latih yang berbeda antara analisis varian (*cluster* ideal) dengan perhitungan seluruh aturan yang dibangkitkan. Pada laki-laki jumlah *cluster* terbaik berdasarkan analisis varian (*cluster* ideal) adalah 11, 12, dan 9 sedangkan menurut perhitungan seluruh aturan yang dibangkitkan jumlah *cluster* terbaik adalah 4, 11, dan 6. Pada perempuan jumlah *cluster* terbaik berdasarkan analisis varian (*cluster* ideal) adalah 7, 8, dan 10 sedangkan menurut perhitungan seluruh aturan yang dibangkitkan jumlah *cluster* terbaik adalah 7, 10, dan 5.

Pada tabel 7 dan 8 tersebut telah dijabarkan hasil persentase *error* di ketiga jumlah data latih dimana menurut analisis varian (*cluster* ideal) nilai terkecil pada laki-laki terletak di jumlah *cluster* 9 (70 data latih) dengan persentase MAPE 27.56%. Sedangkan pada hasil seluruh aturan yang dibangkitkan MAPE terkecil terdapat pada jumlah *cluster* 4 (50 data latih) dengan persentase 22.55%. Pada perempuan hasil persentase MAPE terkecil menurut analisis varian (*cluster* ideal) berada pada jumlah *cluster* 10 (60 data latih) dengan persentase sebesar 14.18% sedangkan menurut perhitungan seluruh aturan, MAPE terkecil berada di jumlah *cluster* 5 (70 data latih) dengan persentase sebesar 11.49%. Pada beberapa uji coba antar nilai varian dengan perhitungan seluruh aturan yang dibangkitkan didapat bahwa hasil jumlah *cluster* dengan nilai varian terkecil belum tentu memberikan hasil yang baik. Dan dengan jumlah data yang banyak belum tentu juga menghasilkan hasil yang baik.

## 5. PENUTUP

### 5.1 Kesimpulan

Setelah dilakukannya proses perancangan, pengimplementasian, dan pengujian dapat diambil kesimpulan sebagai berikut :

1. Berdasarkan hasil pengujian yang telah dilakukan dapat diambil kesimpulan bahwa nilai *cluster* ideal belum tentu memiliki hasil MAPE terbaik..
2. Penerapan metode K-means *Clustering* untuk perencanaan konsumsi pangan harian bisa dikatakan belum optimal di karenakan hasil nilai MAPE yang besar dan tidak konstan disetiap percobaannya. Hasil MAPE yang besar dapat dilihat berdasarkan nilai MAPE terbaik dari pengujian seluruh aturan yang dibangkitkan dimana pada laki-laki

menghasilkan nilai MAPE yang cukup besar yaitu lebih dari 20%, sedangkan pada perempuan memiliki hasil MAPE yang baik yaitu dibawah 20% sehingga bisa disimpulkan bahwa MAPE yang dihasilkan memiliki nilai yang besar. Hasil kalori yang dihasilkan sistem tersebut juga tidak sesuai dengan yang diharapkan dimana harapannya selisih nilai hasil perhitungan dan nilai sebenarnya berkisar puluhan atau belasan. Hasil MAPE yang tidak konstan dapat dilihat berdasarkan pada seluruh aturan yang dibangkitkan dimana di setiap percobaan *cluster* dengan parameter (jenis kelamin, jumlah data, dan jumlah *cluster*) yang sama, memiliki hasil aturan yang berbeda dan secara otomatis menghasilkan nilai MAPE yang berbeda juga. Hal tersebut disebabkan oleh inisiasi *centroid* awal secara random saat penentuan aturan.

### 5.2 Saran

Berikut saran untuk menelitian selanjutnya :

1. Pada skenario uji hendaknya dilakukan lebih banyak lagi dan bisa juga dilakukan pengujian dengan model rasio data latih dan data uji agar pola data dapat lebih terlihat.
2. Pada algoritma *k-means clustering* dapat diimplementasikan suatu metode optimasi untuk pengoptimalan dalam penentuan pusat *cluster* awal atau *centroid* awal agar menemukan hasil yang lebih optimal dan memperkecil jumlah skenario uji
3. Pada perhitungan koefisien *output* disetiap aturan dapat diimplementasikan suatu metode optimasi agar menghasilkan nilai Z yang lebih baik.
4. Pada pembangunan sistem otomatisasi aturan *fuzzy* hendaknya kedepannya dibangun juga sistem yang sama namun dengan aturan yang didapat dari akuisisi pengetahuan pakar sehingga bisa dibandingkan dan diketahui optimal tidaknya hasil sistem yang menerapkan otomatisasi aturan *fuzzy* tersebut.

### DAFTAR PUSTAKA

- ALBRIGHT, S. C. DAN WINSTON, W. L. 2016. *Business Analytics: Data Analysis & Decision Making*. United States of America: Cengage Learning.
- BRIAWAN, D. DAN HARDIYANSYAH. 1990. *Penilaian dan Perencanaan Konsumsi Pangan*. Bogor: Institut Pertanian Bogor.
- DIREKTORAT JENDERAL BINA GIZI DAN KESEHATAN IBU DAN ANAK., 2011. *Strategi Nasional Penerapan Pola Konsumsi Makanan dan Aktivitas Fisik Untuk Mencegah Penyakit Tidak Menular*. Kementerian Kesehatan Republik Indonesia.

FIELDING, ALAN H. 2007. *cluster and Classification Techniques for the Biosciences*. Cambridge University Press.

GHANBARI, A. ALI AZADEH, M. GHADERI, S F. 2010. A Clustering based Genetic Fuzzy expert system for electrical energy demand prediction. *Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on Volume:5*.

GILLILAND, M., 2010. *The Bussiness Forecasting Deal: Exposing Myths, Eliminating Bad Practices, Providiing Practical Solutions*. New Jersey: John Wiley & Sons, Inc.

ILHAM, B. P. 2011. Implementasi Metode Single Linkage untuk Menentukan Kinerja Agen pada Call Centre Berbasis Asterisk for Java. *Paper and Presentation of Telecommunication Engineering, 006.42*.

JANG, J.S., SUN, C.T. DAN MIZUTANI, E. 1997. *Neuro Fuzzy and Soft Computing*. New York: Prentice Hall.

KUSUMADEWI, S. DAN PURNOMO, H. 2010. *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*. Edisi Kedua. Yogyakarta: Graha Ilmu.

LAROSE, DANIEL T. 2005. *Discovering Knowledge in Data: an Introduction to Data Mining*. New Jersey: John & Wiley & Sons, Inc.

PERSATUAN AHLI GIZI INDONESIA (PERSAGI). 2009. *Tabel Komposisi Pangan Indonesia (TKPI)*. Jakarta: PT Elex Media Komputindo.

SEKRETARIAT JENDERAL KEMENTERIAN KESEHATAN RI. 2015. *Rencana Strategis Kementerian Kesehatan Tahun 2015-2019*. Kementerian Kesehatan Republik Indonesia.

SHOLEH, KHOIRUL. 2013. *Implementasi Metode K-Means Clustering untuk Pembangkitan Aturan Fuzzy pada Klasifikasi Ketahanan Hidup Penderita Kanker Payudara*. S1. Universitas Brawijaya.