

## BAB II. LANDASAN TEORI

Pada bab ini diuraikan kajian pustaka dan dasar teori yang mendukung laporan penelitian. Dasar teori tersebut diperoleh dari referensi yang relevan dengan topik yang diangkat dalam laporan penelitian ini. Dalam bab ini dijelaskan kajian pustaka dan metode *Fuzzy C-Means*.

### 2.1 Tinjauan Pustaka

Tinjauan pustaka merupakan bagian yang membahas tentang penyelesaian masalah yang akan memberikan jalan keluarnya. Dalam hal ini dikemukakan beberapa teori-teori yang berkaitan dengan masalah yang diangkat.

#### 2.1.1 Penelitian Terdahulu

Studi literatur merupakan bagian yang akan membahas beberapa penelitian yang dijadikan rujukan dalam penelitian ini adalah sebagai berikut:

Penelitian dengan judul “Implementasi *Fuzzy C-Means* Untuk Prediksi Perilaku Mahasiswa Berdasarkan Jumlah Ketidakhadiran” disusun Oleh Ridwan Riswanto, Imam Fahrur Rozi dan Arief Prasetyo pada tahun. Penelitian ini menyebutkan bahwa hasil dari klasifikasi dengan *Fuzzy C-Means* digunakan untuk memprediksi perilaku mahasiswa di semester berikutnya dengan menjadi 2 kelompok diantaranya kelompok 1 adalah kelompok yang tidak perlu dilakukan pengawasan, dan kelompok 2 adalah kelompok yang diperlukan pengawasan (Rismanto et al., 2017).

Penelitian dengan judul “Kernel Based *Fuzzy C-Means Clustering* untuk Klasifikasi Sinusitis Kronis” disusun oleh Rezki Aulia Putri, Zuherman Rustam, Jacub Pandelaki pada tahun 2019. Menyebutkan bahwa untuk mengklasifikasikan sinusitis digunakan Kernel Based *Fuzzy C-Means Clustering* Algorithm yang merupakan pengembangan dari *Fuzzy C-Means* (FCM). Dengan menggunakan kernel alat sebagai pengganti jarak Euclidian yang akan memetakan sampel ke dalam ruang yang berdimensi tinggi sehingga meningkatkan perbedaan antara pusat *Cluster* yang dimana memperbaiki dari kekurangan *Fuzzy C-Means* itu sendiri dan di dapatkan hasil akurasi sebesar 100% (Putri et al., 2019).

Penelitian dengan judul “Implementasi Metode *Fuzzy C-Means* Untuk Klasifikasi Status Gizi Pada Balita Berdasarkan Indeks Antropometri” oleh Cahyati Resty Ardiantipada tahun 2019. Penelitian ini menjelaskan dengan menggunakan Sistem Implementasi Metode *Fuzzy C-Means*

untuk Klasifikasi Status Gizi Pada Balita Berdasarkan Indeks Antropometri yang dapat mengetahui status gizi balita secara cepat dan tepat. Dengan data uji yang dinormalisasikan dengan menggunakan normalisasi Max-Min, klasifikasi status gizi dihitung menggunakan metode *Fuzzy C-Means*. Didapatkan dari perhitungan *Fuzzy C-Means* menghasilkan hasil akurasi sebesar 73% (Ardianti, 2019).

Penelitian dengan judul “*Clustering* Kinerja Pegawai PT Gresik Cipta Sejahtera Menggunakan *Fuzzy C-Means*” disusun oleh Adi Pranoto, Ariadi Retno Tri Hayati Ririd, Rizky Ardiansyah pada tahun 2018. Penelitian ini menyebutkan bahwa Nilai kriteria tersebut diolah dengan menggunakan metode *Fuzzy C-Means* (FCM) yang menghasilkan sejumlah kelompok karyawan yang memiliki standar penilaian yang bersifat dinamis. Uji validitas hasil *Clustering* dengan menggunakan perhitungan Partition Coefficient Index (PCI) dan diperoleh 0.814 dari hasil uji validitas maka penelitian ini dapat dikatakan bahwa hasil *Clustering* tergolong dalam kategori baik (Pranoto et al., 2018).

Penelitian “Implementasi Metode K-Means *Clustering* pada Kasus *Stunting* di Indonesia” oleh Nailul Izza, A.Md. S.KM, Dr. Windhu Purnomo M.S. dan Dr. Mahmudah, Ir.M.Kes., pada tahun 2019. Penelitian dengan K-Means *Clustering* menghasilkan hasil pengklasteran yang optimal yang membentuk 4 *Cluster*, dimana kelompok 1 merupakan klaster yang provinsinya memiliki status gizi baik. Dan juga menghasilkan 4 klaster prioritas perhatian karena nilai rata-rata *stunting* bernilai tinggi (Izza, Nailul A.Md. S.KM et al., 2019).

## 2.2 *Stunting*

Definisi *Stunting* (kerdil) adalah kondisi dimana balita memiliki panjang atau tinggi badan yang kurang jika dibandingkan dengan umur. Kondisi ini diukur dengan panjang atau tinggi badan yang lebih dari minus dua standar deviasi median standar pertumbuhan anak dari WHO. Balita *stunting* termasuk masalah gizi kronik yang disebabkan oleh banyak faktor seperti kondisi sosial ekonomi, gizi ibu saat hamil, kesakitan pada balita, dan kurangnya asupan gizi pada balita. Balita *stunting* di masa yang akan datang akan mengalami kesulitan dalam mencapai perkembangan fisik dan kognitif yang optimal (Kemenkes RI, 2018).

## 2.3 *Clustering*

*Clustering* adalah termasuk metode pengelompokan data. *Clustering* adalah suatu proses untuk mengelompokkan data ke dalam beberapa *cluster* atau kelompok data sehingga dalam suatu cluster memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki tingkat kemiripan yang sangat minimum atau sangat jauh (Tan et al., n.d.). *Clustering* juga dikenal sebagai data segmentasi karena *clustering* mempartisi banyak data set ke dalam banyak *group* berdasarkan kesamaannya. Selain itu *clustering* juga bisa sebagai *outlier detection* (IRWANSYAH, n.d.).

*Clustering* memiliki manfaat diantara lain:

- a) *Clustering* merupakan metode segmentasi data yang sangat berguna dalam prediksi dan analisa masalah bisnis tertentu. Misalnya Segmentasi pasar, marketing dan pemetaan zonasi wilayah.
- b) Identifikasi obyek dalam bidang berbagai bidang seperti computer vision dan image processing.

### 2.3.1 *Fuzzy C-Means*

Hubungan dari clustering Metode ini ditemukan pertama kali oleh Dunn pada tahun 1973 kemudian dikembangkan lagi oleh Bezdek pada tahun 1981. Ide dasar dari metode ini mirip dengan metode *K-Means*. *FCM* didasarkan pada logika fuzzy, setiap titik data dimasukkan ke suatu kelompok berdasarkan nilai keanggotaannya pada kelompok tersebut (Haqiqi & Kurniawan, 2015). *Fuzzy C-Means Clustering (FCM)*, atau dikenal juga sebagai *Fuzzy Isodata* merupakan salah satu metode *Clustering* yang merupakan bagian dari metode *Hard K-Means*. *FCM* menggunakan model pengklasteran *fuzzy* sehingga data dapat menjadi anggota dari semua kelas atau *Cluster* terbentuk dengan derajat atau tingkat keanggotaan yang berbeda antara 0 hingga 1. Tingkat keberadaan data dalam suatu kelas atau *Cluster* ditentukan oleh derajat keanggotaannya. Konsep dasar *FCM*, pertama kali adalah menentukan pusat *Cluster* yang akan menandai lokasi rata-rata untuk tiap-tiap *Cluster*. Pada kondisi awal pusat *Cluster* ini masih belum akurat. Tiap-tiap data memiliki derajat keanggotaan untuk tiap-tiap *Cluster*. Dengan cara memperbaiki pusat *Cluster* dan nilai keanggotaan tiap-tiap data secara berulang, maka dapat dilihat bahwa pusat *Cluster* akan menuju lokasi yang tepat. Perulangan ini didasarkan pada minimasi fungsi obyektif yang menggambarkan jarak dari titik data yang diberikan ke pusat *Cluster* yang berbobot oleh

derajat keanggotaan titik data tersebut implementasi (Ilmu Matematika Dan Terapan & Maret, 2017).

Konsep dasar *FCM*, pertama kali adalah menentukan pusat *Cluster*, yang akan menandai lokasi rata-rata untuk tiap-tiap *Cluster*. Pada kondisi awal, pusat *Cluster* ini masih belum akurat. Tiap-tiap titik data memiliki derajat keanggotaan untuk tiap-tiap *Cluster*. Dengan cara memperbaiki pusat *Cluster* dan derajat keanggotaan tiap-tiap titik data secara berulang, maka akan dapat dilihat bahwa pusat *Cluster* akan bergerak menuju lokasi yang tepat. Perulangan ini didasarkan pada minimasi fungsi objektif yang menggambarkan jarak dari titik data yang diberikan ke pusat *Cluster* yang terbobot oleh derajat keanggotaan titik data tersebut (Ardianti, 2019).

### 2.3.2 Algoritma *Fuzzy C-Means*

Langkah-langkah perhitungan Algoritma *Fuzzy C-Means* adalah sebagai berikut:

- a) Menginput data yang akan di kelompokkan menjadi klaster  $X$ , berupa matriks berukuran  $n \times m$ .  $X_{ij}$  adalah data sampel ke- $i$  ( $i = 1, 2, \dots, n$ ), dan atribut ke- $j$  ( $j = 1, 2, \dots, m$ ).

Keterangan:

$n$  = jumlah sampel data.

$m$  = atribut data

- b) Tentukan:

Jumlah klaster =  $c$

Pangkat pembobot =  $w$

Maksimum iterasi =  $MaxIter$

Error terkecil diharapkan =  $\zeta$

Fungsi objektif awal =  $P_0 = 0$

Iterasi awal =  $t = 1$

- c) Membangkitkan bilangan random  $u_{ik}$  dengan  $i$  merupakan banyak data dan  $k$  merupakan banyak kelompok sebagai elemen-elemen awal matriks keanggotaan awal  $u_{ik}$ .  $u_{ik}$  disebut derajat keanggotaan sampel data ke  $i$  dan *Cluster* ke- $k$  dengan  $u_{ik}$ . Posisi dan nilai matriks dibangun secara random. Nilai keanggotaan terletak pada interval 0 sampai dengan 1. Pada posisi awal matriks partisi  $U$  masih belum akurat

begitu juga pusat klasternya. Sehingga kecenderungan data untuk masuk suatu kluster belum akurat.

Sehingga hitung jumlah setiap kolom atribut yang ada menggunakan persamaan 2.1.

$$Q_i = \sum_{k=1}^c (\mu_{ik}) \quad (2.1)$$

Keterangan:

$Q_i$  = jumlah nilai derajat keanggotaan per kolom = 1 dengan  $i = 1, 2, 3, \dots, n$

Setelah itu hitung derajat keanggotaan iterasi pertama  $Q_j$  dengan  $j = 1, 2, 3, \dots, m$  dengan persamaan 2.2

$$u_{ik} = \frac{u_{ik}}{Q_j} \quad (2.2)$$

- d) Menghitung pusat kluster ke- $k$  ( $V_{kj}$ ) dengan menggunakan persamaan 2.3, dengan  $k = 1, 2, 3, \dots, c$ ; dan  $j = 1, 2, 3, \dots, m$  :

$$V_{kj} = \frac{\sum_{k=1}^n (u_{ik})^w * X_{ij}}{\sum_{k=1}^n (u_{ik})^w} \quad (2.3)$$

Keterangan:

$u_{ik}$  = nilai keanggotaan objek ke- $k$  dengan pusat kelompok ke- $i$

$X_k$  = objek data ke- $k$

$n$  = banyaknya objek penelitian

$w$  = fuzzifier atau pembobot

- e) Menghitung fungsi objektif pada iterasi ke- $t$ ,  $P_t$  dengan persamaan 2.4 :

$$P_t = \sum_{i=1}^n \sum_{k=1}^c ([\sum_j^m (X_{ij} - V_{kj})^2] (\mu_{ik})^w) \quad (2.4)$$

Keterangan :

$c$  = banyak kelompok yang diinginkan

$n$  = banyak objek penelitian

$m$  = fuzzifier

$u_{ik}$  = nilai keanggotaan objek ke- $k$  pada kelompok ke- $i$  yang merupakan bagian dari matriks U

$V_{kj}$  = Pusat Cluster,

Fungsi objektif di gunakan sebagai syarat perulangan untuk mendapatkan error terkecil yang di inginkan.

- f) Menghitung perubahan matriks partisi keanggotaan dengan  $i = 1, 2, 3 \dots, n$ ; dan  $k = 1, 2, 3, \dots, c$ ; dengan persamaan 2.5 dibawah:

$$u_{ik} = \frac{[\sum_{j=1}^m (x_{kj} - v_{ij})^2]^{\frac{-1}{p-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (x_{kj} - v_{ij})^2]^{\frac{1}{p-1}}} \quad (2.5)$$

g) Cek kondisi jika berhenti

- Jika  $|J_t - J_{t-1}| < \zeta$  atau  $t > \text{MaxIter}$  maka berhenti;
- Jika tidak maka  $t = t + 1$ , ulangi dengan langkah 3 dengan persamaan 1

## 2.4 *Silhouette Coefficient*

Pengujian menggunakan *Silhouette Coefficient* yaitu menghitung jarak antar *cluster* yang dihitung menggunakan jarak *Euclidean*. *Silhouette Coefficient* merupakan salah satu ukuran validasi yang berbasis kriteria internal. *Silhouette index* akan mengevaluasi penempatan setiap objek dalam setiap kluster dengan membandingkan jarak rata-rata objek dalam satu kluster dan jarak antara objek dengan kluster yang berbeda (Nahdliyah et al., 2019).

Pengujian menggunakan *Silhouette Coefficient* dilihat nilai *cluster* terbaik masing – masing *cluster*. Dengan rumus persamaan dibawah ini :

1. Menghitung rata-rata jarak dari obyek i dengan seluruh obyek yang berada pada 1 *cluster*:

$$a(i) = \frac{1}{[A]-1} \sum_{j \in A, j \neq i} d(i, j) \quad (2.7)$$

2. Menghitung rata – rata jarak objek i dengan objek yang berada pada *cluster* lain, kemudian diambil nilai terkecilnya:

$$d(i, C) = \frac{1}{[A]} \sum_{j \in C} d(i, j) \quad (2.8)$$

3. Untuk menghitung nilai *Silhouette Coefficient*nya digunakan rumus:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2.9)$$