

## Implementasi Algoritma *Fuzzy C-Means* untuk Mengkategorikan Tingkat Penjualan Produk pada Data Transaksi Swalayan

Sagung Rani Ari Puspita<sup>1</sup>, Indwiarti<sup>2</sup>, Fhira Nhita<sup>3</sup>

<sup>1,2,3</sup>Fakultas Informatika, Universitas Telkom, Bandung<sup>4</sup>

<sup>1</sup>sagungraniap@students.telkomuniversity.ac.id, <sup>2</sup>indwiarti@telkomuniversity.ac.id,

<sup>3</sup>fhiranhita@telkomuniversity.ac.id

---

### Abstrak

Swalayan setiap harinya mencatat transaksi penjualan yang sangat banyak, hal ini akan berdampak pada pertumbuhan jumlah data yang sangat banyak dan menimbulkan tumpukan data yang berjumlah besar. Dari data transaksi tersebut dapat digali informasi-informasi baru yang berguna untuk menunjang proses bisnis swalayan. Misalnya pentingnya jenis barang apa yang menjadi prioritas utama yang harus distok swalayan. Pada penelitian ini digunakan algoritma *Fuzzy c-means* pada data penjualan swalayan untuk mendapatkan tingkat penjualan produk. Dari hasil pengujian yang didapat, untuk produk susu yang termasuk kategori rendah adalah jenis produk susu bayi, dan untuk produk air mineral yang termasuk kategori rendah adalah jenis produk air mineral 240ml, 330ml dan 600ml, dan produk yang termasuk kategori tinggi adalah jenis produk mie goreng. Pada pengujian *cluster* menggunakan metode *Modified Partition Coefficient* (MPC) hasilnya yaitu validitas *cluster* tertinggi yang terbaik adalah pada saat menggunakan 5 *cluster*.

**Kata kunci :** Swalayan, Data Penjualan, *Fuzzy C-Means*, *Clustering*, MPC

---

### Abstract

Supermarkets every day record a lot of sales transactions, this will have an impact on the growth of very large amounts of data and generate large amounts of data. From the transaction data new information can be extracted that is useful to support the self-service business process. For example the importance of what kind of goods are the top priority that must be self-service. In this study, *Fuzzy c-means* algorithm is used in self-service sales data to obtain the level of product sales. babies, and for mineral water products that are categorized as low are types of mineral water products 240ml, 330ml and 600ml, and products that are in the high category are types of fried noodle products. In cluster testing using the *Modified Partition Coefficient* (MPC) method the result is that the highest cluster validity is best when using 5 clusters.

**Keywords:** *Supermarket*, *Sales Data*, *Fuzzy C-Means*, *Clustering*, MPC

---

### 1. Pendahuluan

Swalayan setiap harinya mencatat transaksi penjualan yang sangat banyak. Hal ini berdampak pada pertumbuhan jumlah data yang sangat pesat [1]. Jika hal ini dibiarkan, maka data-data transaksi tersebut akan menjadi tumpukan sampah yang merugikan karena membutuhkan media penyimpanan/database yang semakin besar. Dari data-data transaksi yang ada pada sebuah swalayan dapat digali informasi-informasi baru yang berguna untuk menunjang proses bisnis swalayan. Data-data tersebut digali dengan metode yang disebut dengan *data mining*[1].

Penerapan *datamining* dapat membantu proses analisa data yang diperoleh dari *record-record* transaksi pada Swalayan Sastra Mas, sehingga dapat menggali informasi yang dapat dijadikan pengetahuan baru untuk proses identifikasi produk di Swalayan Sastra Mas. Misalnya pentingnya jenis barang apa yang menjadi prioritas utama yang harus distok Swalayan[2]. Pada kasus penjualan di Swalayan Sastra Mas ini diharapkan dapat memudahkan swalayan dalam mengidentifikasi produk-produk terbaik mereka.

Oleh karena itu, penulis akan mengambil judul Implementasi Algoritma *Fuzzy C- Means* untuk Mengkategorikan Tingkat Penjualan Produk pada Data Transaksi Swalayan. Algoritma ini dipilih karena dengan algoritma ini, data-data beserta parameter-parameternya dapat dikelompokkan dalam *cluster-cluster* sesuai dengan kecenderungannya, selain itu dengan algoritma ini bisa ditentukan jumlah *cluster* yang akan dibentuk. Dengan penentuan jumlah *cluster* diawal, bisa diatur keragaman nilai akhir sesuai dengan *cluster*-nya. Kelebihan algoritma ini adalah penempatan pusat *cluster* yang lebih tepat dibandingkan dengan metode *cluster* lain. Caranya adalah dengan memperbaiki pusat *cluster* secara berulang, maka akan dapat dilihat bahwa pusat *cluster* akan bergerak menuju lokasi yang tepat [1].

Pada penulisan Tugas Akhir ini akan membahas tentang penggalian informasi data pada Swalayan Sastra Mas dengan menggunakan algoritma *Fuzzy C- Means* untuk mengidentifikasi tingkat penjualan produk.

## Rumusan dan Batasannya

Pada penulisan tugas akhir ini adapun rumusan masalah yang ada yakni :

1. Bagaimana penggunaan algoritma *Fuzzy C-Means* untuk *clustering* terhadap data transaksi Swalayan?
2. Bagaimana mendapatkan hasil performansi yang baik dengan metode *Modified Coefficient Partition* (MPC)?

Dengan batasan masalah yang ada sebagai berikut :

1. Data yang digunakan adalah data transaksi penjualan yang diperoleh dari Toko Swalayan Sastra Mas Estetika Bali, data selama tiga tahun terakhir (2015-2017).
2. Pada penelitian ini meneliti 3 kategori produk secara umum, dikategorikan berdasarkan jenis dari produk yakni : Susu, Mie dan Air.

## Tujuan

Menentukan nilai derajat keanggotaan dan tingkat penjualan produk pada data transaksi Swalayan dengan menggunakan Algoritma *Fuzzy C-Means* berdasarkan *clustering* yang terdiri dari 3 kategori yaitu rendah, sedang dan tinggi. Serta menentukan tingkat performansi yang baik pada data transaksi menggunakan Metode MPC (*Modified Coefficient Partition*).

## 2. Studi Terkait

### 2.1 Data Mining

*Data mining* merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakrit dari berbagai *database* besar/*Data Warehouse* (Turban, dkk. 2005) Keluaran dari *data mining* bisa dipakai untuk memperbaiki pengambilan keputusan dimasa depan (Budi Santosa, 2007). Adapun beberapa metode yang terdapat pada data mining ialah *Classification*, *Association rule*, *Clustering*, dan *Anomali*. kemudian beberapa algoritma yang terdapat dalam data mining ialah *Naive bayes*, *Random forest*, *Apriori*, dan sebagainya.

*Data mining*, sering juga disebut sebagai *Knowledge Discovery in Database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola hubungan dalam set data berukuran besar [9].

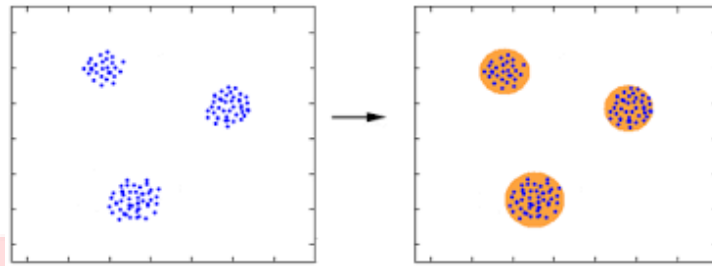
Ada 3 tahapan dalam *Knowledge Discovery in Database* (KDD) :

- 1) *Data Preprocessing*  
Proses ini bertujuan untuk mentransformasikan data *inout* ke dalam format yang sesuai untuk kemudian dianalisa. Dalam tahap ini dilakukan proses penggabungan data dari berbagai sumber, pembersihan data untuk menghilangkan *noise data* dan data ganda, serta memilih atribut data yang diperlukan bagi proses *data mining*.
- 2) *Data Mining*  
Proses ini bertujuan untuk mendapatkan pola-pola informasi yang tersembunyi di dalam basis data. Ada beberapa teknik yang dapat digunakan dalam *data mining* untuk mendapatkan pola-pola dan informasi tersembunyi, yaitu *classification*, *neural network*, *decision tree*, *genetic algorithm*, *clustering*, *OLAP* (*Online Analytical Processing*), dan *association rules*.
- 3) *Postprocessing*  
Proses ini bertujuan untuk memastikan hanya hasil yang valid dan berguna yang dapat digunakan oleh pihak yang berkepentingan. Contoh dari proses ini adalah proses visualisasi, yaitu proses untuk menganalisa dan mengeksplorasi data dan hasil dari proses *data mining* dari berbagai sudut pandang [5].

### 2.2 Clustering

Pengelompokan (*clustering*) merupakan teknik yang sudah cukup dikenal dan banyak digunakan untuk mengelompokkan data/objek ke dalam kelompok data (*cluster*) sehingga setiap *cluster* memiliki data yang mirip dan berbeda dengan data yang berada dalam *cluster* lain. Garcia – Molina et al. [10] menyatakan *clustering* adalah mengelompokkan item data ke dalam sejumlah kecil grup sedemikian sehingga masing-masing grup mempunyai sesuatu persamaan yang esensial.

Jika diberikan himpunan data yang berjumlah terhingga, yaitu  $X$ , maka permasalahan *clustering* dalam  $X$  adalah mencari beberapa pusat *cluster* yang dapat memberikan ciri kepada masing-masing *cluster* dalam  $X$ . Kriteria kemiripan yang digunakan dalam kasus ini adalah jarak (dalam kasus ini jarak geometris). Proses ini disebut *distance-based clustering*. Cara lain untuk melakukan *clustering* adalah *conceptual clustering*. Dalam *conceptual clustering*, objek dikelompokkan berdasarkan kecocokannya menurut konsep deskriptif [5].



**Gambar 1. Contoh Proses Clustering**

Klasterisasi atau *clustering* adalah proses pengelompokan himpunan data ke dalam beberapa grup atau kluster sedemikian hingga objek-objek dalam suatu kluster memiliki kemiripan yang tinggi, namun sangat berbeda (memiliki ketidakmiripan yang tinggi) dengan objek-objek di kluster-kluster lainnya (J Han et al.2012). Kemiripan (*similarities*) dan ketidakmiripan (*dissimilarities*) dihitung berdasarkan nilai-nilai atribut yang menggambarkan objek-objek tersebut dan seringkali melibatkan ukuran jarak.

Klasterisasi banyak digunakan dalam berbagai bidang dengan beragam aplikasi vital yang sangat penting, diantaranya adalah : riset pasar, dimana klasterisasi digunakan untuk segmentasi dan profiling pelanggan yang membantu dalam merancang strategi-strategi untuk mengimplementasikan *customer relationship management* (CRM) yang efektif : sistem rekomendasi produk dalam sistem jual beli online yang biasanya menggunakan pendekatan *collaborative filtering*; *business intelligence*; sistem keamanan; mesin pencarian di internet (*search engine*), dan sebagainya [3].

### 2.3 Fuzzy C-Means

*Fuzzy C-Means* merupakan salah satu metode *clustering* yang menggunakan model pengelompokan *fuzzy* sehingga data dapat menjadi anggota dari semua kelas atau *cluster* terbentuk dengan derajat atau tingkat keanggotaan yang berbeda antara 0 dan 1 [4].

*Fuzzy C-Means* adalah salah satu metode yang *optimizing partitioned cluster*. Kelebihan metode *Fuzzy C-Means* adalah penempatan pusat *cluster* yang lebih tepat dibandingkan dengan metode *cluster* lain. Caranya adalah dengan memperbaiki pusat *cluster* secara berulang maka akan dapat dilihat bahwa pusat *cluster* akan bergerak menuju lokasi yang tepat [1].

*Fuzzy C-Means* mencoba untuk mempartisi kumpulan elemen yang terbatas  $X = \{x_1, x_2, \dots, x_n\}$  ke dalam kumpulan *cluster fuzzy c* sehubungan dengan beberapa kriteria yang diberikan. Pengelompokan *fuzzy* adalah metode yang diterapkan secara luas untuk mendapatkan model *fuzzy* dari data [8].

### 2.4 Modified Partition Coefficient (MPC)

Untuk pengujian hasil cluster, digunakan metode modified partition coefficient (MPC). MPC merupakan metode yang digunakan untuk menguji validitas jumlah cluster. MPC sendiri merupakan pengembangan dari metode partition coefficient (PC). Partition coefficient (PC) merupakan metode yang mengukur jumlah cluster yang mengalami overlap. Nilai PC berada dalam batas  $\frac{1}{c} \leq PC(c) \leq 1$ . Pada umumnya jumlah *cluster* yang paling optimal ditentukan dari nilai PC yang paling besar ( $\max_{2 \leq c \leq n-1} PC(c)$ ). Berikut adalah algoritma metode PC : [1]

$$PC(c) = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^N (\mu_{ij}^2) \quad (3)$$

Dimana :

C = jumlah *cluster*

N = jumlah data

$\mu_{ij}$  = derajat keanggotaan data k-j pada *cluster* ke-i

$PC(c)$  = nilai indeks PC pada *cluster* ke-c

*Partition coefficient* cenderung mengalami perubahan yang monoton terhadap beragam nilai c (jumlah *cluster*). Modifikasi dari indeks PC (*Modified Partition Coefficient MPC*) dapat mengurangi perubahan yang

monoton tersebut. Nilai MPC berada dalam batas  $0 \leq PC(c) \leq 1$ . Pada umumnya jumlah *cluster* yang paling optimal ditentukan dari nilai MPC yang paling besar ( $\max_{2 \leq c \leq n-1} PC(c)$ ) [1]. Berikut adalah algoritma metode MPC :

$$MPC(c) = 1 - \frac{c}{c-1} (1 - PC(c)) \quad (5)$$

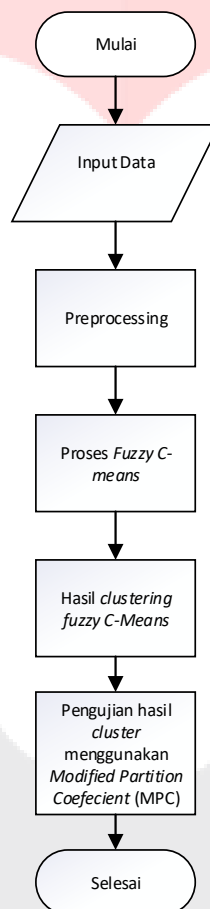
Dimana:

$c$  = jumlah *cluster*

MPC(c) = nilai indeks MPC pada *cluster* ke- $c$

### 3. Sistem yang Dibangun

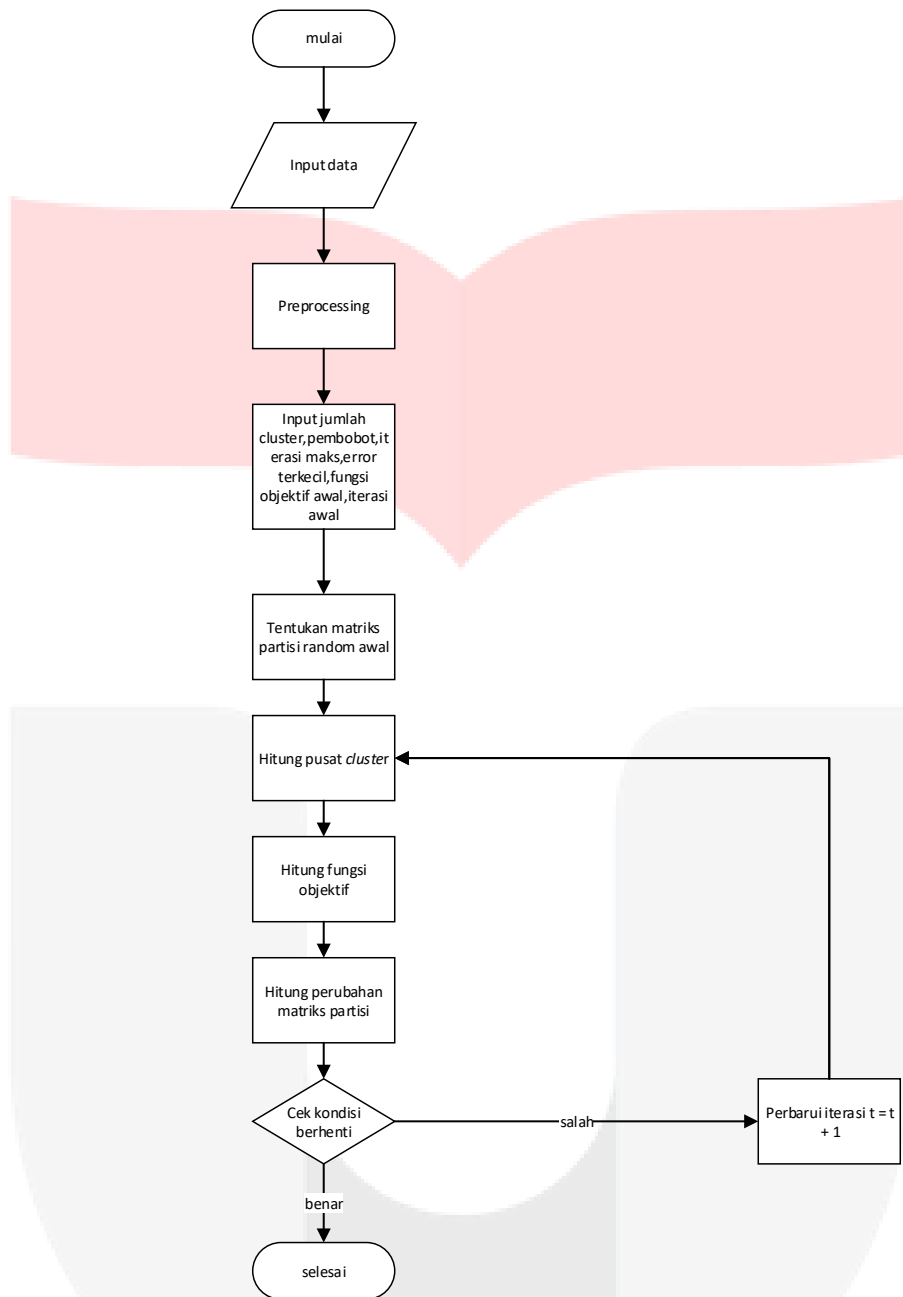
Pada bagian ini akan dijelaskan mengenai data yang digunakan dalam penelitian tugas akhir ini serta gambaran sistem yang akan dibangun.



**Gambar 2. Gambaran umum flowchart**

Berikut adalah gambaran umum flowchart yang dilakukan dalam penelitian ini. Pada langkah pertama yang dilakukan adalah mengambil data berupa data transaksi yang di dapatkan dari swalayan, Kemudian dari data tersebut dilakukan tahap preprocessing, tahapan ini merupakan tahap awal dalam proses pengolahan data. Kemudian data tersebut dilakukan proses *Fuzzy C-Means*. Setelah dilakukan proses tersebut dilanjutkan dengan penguian validitas *cluster* menggunakan metode *Modified Partition Coefficient* (MPC).

### 3.1 Fuzzy C-Means



**Gambar 3. Alur tahapan Fuzzy c-means**

1. Data pada Tugas akhir ini menggunakan data transaksi swalayan dalam 3 tahun terakhir (2015-2017) dengan jumlah data dalam satu tahun rata-rata terdiri dari 92677 record, dan data yang akan digunakan terhitung dari tanggal 1 Januari 2015 hingga 31 Desember 2017.
2. Melakukan *Preprocessing* Data  
Pada tahap ini dilakukan pembersihan data yaitu cek missing value, menentukan variable, menghapus dan mengambil variable yang diperlukan dan melakukan pengolahan data.
3. Setelah dilakukan *Preprocessing* data, maka dapat dilakukan input data.
4. Pada tahap ini dilakukan inisialisasi data, dengan menentukan parameter awal, parameter yang diperlukan seperti jumlah *cluster* yang diinginkan, pangkat, maksimum iterasi, kriteria penghentian, fungsi objektif awal, dan iterasi awal.
5. Setelah menetapkan nilai parameter awal dan data siap di *cluster* selanjutnya akan dilakukan proses *clustering* dengan menggunakan langkah-langkah dari algoritma *Fuzzy C-Means* yaitu menentukan matriks partisi random awal.

6. Setelah menentukan matriks partisi random awal,selanjutnya dapat dilakukan menghitung pusat *cluster* dengan menggunakan persamaan (1).
7. Pada tahap ini dilanjutkan dengan menghitung fungsi obyektif, fungsi obyektif digunakan sebagai syarat perulangan untuk mendapatkan pusat *cluster* yang tepat dengan menggunakan persamaan (2).
8. Selanjutnya menghitung perubahan matriks partisi dengan menggunakan persamaan (3).
9. Lalu lakukan cek kondisi berhenti.

### 3.2 Data Set

Data set yang digunakan pada penelitian ini menggunakan data transaksi swalayan Sastra mas bali dalam 3 tahun terakhir (2015-2017). Pada data penjualan yang di dapatkan ini memiliki 6 variabel dengan memiliki jumlah record 272.637 record dari keseluruhan data. Berikut contoh sampel data penjualan. Penulis disini mengambil 3 variabel yang akan digunakan yaitu Tanggal, Nama barang, dan Jumlah produk yang terjual.

**Tabel 1. Data transaksi swalayan**

No	Faktur	Tanggal	Kode Barang	Nama Barang	Jumlah	Harga
1	0115/FJ/0001-1	2-Jan-15	8.99296E+12	LG Singa Choco mete	1	1500
2	0115/FJ/0001-1	2-Jan-15	8.99703E+12	Purnama Roti Manis	1	3000
3	0115/FJ/0002-1	2-Jan-15	15018	Mandala Roti Kombinasi	2	9000
4	0115/FJ/0002-1	2-Jan-15	7.6223E+12	Oreo Golden	1	7000
5	0115/FJ/0002-1	2-Jan-15	8.88817E+12	Peanut crackers biscuits 475 gr	1	13600
...	...	...	...	...	...	...
272.637	1217/FJ/2350-0	31-Dec-17	3.0112E+13	Air Reffil Nonmin	1	19000

Sumber : Database penjualan Swalayan Sastra Mas

Pada swalayan terdapat sangat banyak merek-merek produk yang tidak bisa disebutkan satu-persatu, mulai dari jenisnya sampai kegunaannya yang bermacam-macam, oleh karena itu disini penulis akan mengelompokkan produk berdasarkan kategori produk tersebut. Berikut kategori produk yang dipilih :

1. Susu
2. Mie
3. Air Mineral

Dari pemilihan kategori produk diatas, dilihat berdasarkan data dari kategori produk-produk yang selalu dibeli konsumen setiap harinya.

Setelah 3 kategori produk diatas terpilih ,ketiga kategori produk tersebut dikelompokkan kembali menjadi beberapa jenis produk, disini penulis mengubah nama produk untuk mempermudah penelitian yakni :

1. Kategori produk susu terbagi atas :
  - a. Bayi 1-12 bulan menjadi produk P1.
  - b. Balita menjadi produk P2.
  - c. Anak – Dewasa menjadi produk P3.
  - d. Manula menjadi produk P4.
  - e. Ibu hamil menjadi produk P5.
2. Kategori produk mie terbagi atas :
  - a. Mie goreng menjadi produk P6.
  - b. Mie kuah menjadi produk P7.
3. Kategori produk air mineral terbagi atas :
  - a. Air mineral 240 ml menjadi produk P8.
  - b. Air mineral 330 ml menjadi produk P9.
  - c. Air mineral 600 ml menjadi produk P10.

- d. Air mineral 1500 ml menjadi produk P11.

Setelah dilakukan pengelompokan produk dan mengubah nama produk, penulis melakukan pemrosesan data menjadi jumlah penjualan dalam perhari dan perbulan, agar mempermudah proses pada saat tahap algoritma.

Berikut data yang akan digunakan adalah data penjualan per hari dari 11 jenis kategori produk yang dibeli konsumen selama 3 tahun yakni 2015, 2016, 2017. Data memiliki 1096 record data penjualan berdasarkan hari seperti tabel 2

**Tabel 2. Jumlah penjualan per hari**

No	Periode Tanggal	Jumlah barang terjual										
		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
1	01-jan-15	0	0	0	0	0	0	0	0	0	0	0
2	02-jan-15	1	2	8	3	0	6	14	0	0	7	1
3	03-jan-15	0	0	14	0	1	18	7	0	0	3	5
4	04-jan-15	2	2	11	2	0	13	21	0	0	9	5
5	05-jan-15	0	0	10	0	0	0	2	0	0	4	4
.	...	...	...	...	...	...	...	...	...	...	...	...
.	...	...	...	...	...	...	...	...	...	...	...	...
1094	30-Des-17	0	0	0	0	0	0	0	0	0	0	0
1095	31-Des-17	0	0	12	0	0	8	8	0	1	4	2

Setelah data diolah menjadi data per hari, maka untuk mempermudah proses dalam algoritma, data diolah kembali menjadi data penjualan setiap bulan dalam 3 tahun. Seperti tabel 3

**Tabel 3 Jumlah penjualan per bulan**

No	Periode Bulan	Jumlah barang terjual										
		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
1	Jan-2015	10	16	305	31	6	223	233	9	0	244	130
2	Feb-2015	17	11	262	23	8	192	223	0	1	212	148
3	Maret-2015	33	17	398	26	13	221	251	27	3	359	155
4	April-2015	26	16	347	12	6	193	151	25	17	269	170
5	Mei-2015	23	14	323	21	4	221	181	17	24	266	162
.	...	...	...	...	...	...	...	...	...	...	...	...
.	...	...	...	...	...	...	...	...	...	...	...	...
35	November-2017	0	10	348	11	4	132	162	27	46	257	207
36	Desember-2017	1	8	366	6	2	146	165	53	15	355	188

Dan terakhir digabungkan data harian menjadi data bulanan dengan jumlah 36 record dari keseluruhan data yakni data penjualan per bulan dalam 3 tahun.

#### 4. Evaluasi

##### 4.1 Hasil Pengujian

Setelah data sudah siap untuk diklaster, akan dilakukan proses clustering menggunakan langkah-langkah dari Algoritma Fuzzy C-Means (FCM).

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

1. Input data yang akan dikelompokkan, yaitu  $X$  berupa matriks berukuran  $n \times m$  ( $n$  = jumlah sampel data,  $m$  = atribut setiap data.).  $X_{ij}$  data sampel ke- $i$  ( $i=1,2,\dots,n$ ), atribut ke- $j$  ( $j=1,2,\dots,m$ ).
2. Tentukan jumlah *cluster* ( $c$ ), pangkat untuk matriks partisi ( $w$ ), maksimum iterasi ( $MaxIter$ ), error terkecil yang diharapkan ( $\xi$ ), fungsi objektif awal ( $P_0=0$ ), dan iterasi awal ( $t=1$ ).

**Tabel 4. Nilai Parameter Awal**

Parameter	Nilai
Cluster (c)	3
Pangkat (w)	2
Maksimum Iterasi (MaxIter)	100
Kriteria Penghentian ( $\xi$ )	$10^{-5}$
Fungsi Objektif Awal ( $P_0$ )	0
Iterasi Awal (t)	1

- Bangkitkan bilangan random  $\mu_{ik}$   $i=1,2,\dots,n$ ;  $k=1,2,\dots,c$  sebagai elemen matriks partisi awal U.
- Hitung pusat *cluster* ke-k:  $V_{kj}$ , dengan  $k=1,2,\dots,c$ ; dan  $j=1,2,\dots,m$ , pusat *cluster* digunakan untuk mengkategorikan setiap cluster. menggunakan persamaan berikut (Yan,1994) :

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

dengan :

$V_{kj}$  = pusat *cluster* ke-k untuk atribut ke-j

$\mu_{ik}$  = derajat keanggotaan untuk data sampel ke-i pada *cluster* ke-k

$X_{ij}$  = data ke-i, atribut ke-j

- Hitung fungsi objektif pada iterasi ke-t, fungsi objektif dilakukan sebagai syarat perulangan untuk mendapatkan pusat *cluster* yang tepat. menggunakan persamaan berikut (Yan,1994) :

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

dengan :

$V_{kj}$  = pusat *cluster* ke-k untuk atribut ke-j

$\mu_{ik}$  = derajat keanggotaan untuk data sampel ke-i pada *cluster* ke-k

$X_{ij}$  = data ke-i, atribut ke-j

$P_t$  = fungsi objektif pada iterasi ke-t

- Hitung perubahan matriks partisi, perhitungan ini dilakukan untuk menghasilkan derajat keanggotaan yang baru. menggunakan persamaan berikut (Yan ,1994) :

$$\mu_{ik} = \frac{\left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{1}{w-1}}}{\sum_{k=1}^c \left[ (X_{ij} - V_{kj})^2 \right]^{\frac{1}{w-1}}} \quad (3)$$

dengan  $i=1,2,\dots,n$ ; dan  $k=1,2,\dots,c$

Dimana :



$V_{kj}$  = pusat *cluster* ke-k untuk atribut ke-j

$X_{ij}$  = data ke-i, atribut ke-j

$\mu_{ik}$  = derajat keanggotaan untuk data sampel ke-i pada *cluster* ke-k

7. Cek kondisi berhenti :

Jika :  $(|P_t - P_{t-1}| < \xi)$  atau  $(t > MaxIter)$  maka berhenti. Jika tidak :  $t = t + 1$ , ulangi langkah ke 4.

Maka setelah dilakukan proses Algoritma Fuzzy C-Means, dari hasil perbaikan matriks partisi baru pada iterasi terakhir inilah yang akan menentukan produk tersebut di setiap bulannya mana yang termasuk penjualan tinggi, sedang atau rendah. Berikut adalah contoh proses menentukan kategori untuk sampel produk susu bayi

**Tabel 5. Nilai Derajat Keanggotaan Baru pada Iterasi Terakhir Produk Susu Bayi dalam 3 Tahun**

Bulan	Penjualan	Cluster 1	Cluster 2	Cluster 3	Status Cluster
1	10	0.901901	0.011512	0.086587	Sedang
2	17	0.636042	0.252837	0.111121	Sedang
3	33	0.107718	0.832718	0.059564	Tinggi
4	26	0.00206	0.997041	0.000899	Tinggi
5	23	0.041123	0.943788	0.015089	Tinggi
6	21	0.171076	0.77561	0.053314	Tinggi
7	17	0.636042	0.252837	0.111121	Sedang
8	10	0.901901	0.011512	0.086587	Sedang
9	15	0.837192	0.083317	0.079491	Sedang
10	12	0.998385	0.000395	0.00122	Sedang
11	6	0.071582	0.006286	0.922132	Rendah
12	6	0.071582	0.006286	0.922132	Rendah
13	5	0.007809	0.00085	0.991336	Rendah
14	6	0.071582	0.006286	0.922132	Rendah
15	6	0.071582	0.006286	0.922132	Rendah
16	1	0.089697	0.017429	0.892874	Sedang
17	6	0.071582	0.006286	0.922132	Rendah
18	9	0.723758	0.02024	0.256002	Sedang
19	7	0.227883	0.015167	0.75695	Rendah
20	9	0.723758	0.02024	0.256002	Sedang
21	4	0.002697	0.000354	0.996949	Rendah
22	3	0.025032	0.003823	0.971145	Rendah
23	6	0.071582	0.006286	0.922132	Rendah
24	5	0.007809	0.000855	0.991336	Rendah
25	2	0.056839	0.010171	0.020992	Rendah
26	13	0.968836	0.010171	0.020992	Sedang
27	8	0.470861	0.02181	0.507329	Rendah
28	5	0.007809	0.000855	0.991336	Rendah
29	11	0.985212	0.00258	0.012208	Sedang
30	7	0.227883	0.015167	0.75695	Rendah
31	0	0.120203	0.025753	0.854044	Rendah
32	6	0.071582	0.006286	0.922132	Rendah
33	2	0.056839	0.009877	0.933284	Rendah
34	6	0.071582	0.006286	0.922132	Rendah
35	0	0.120203	0.025753	0.854044	Rendah
36	1	0.089697	0.017429	0.892874	Rendah
Pusat cluster		11.73447	25.35157	4.402297	

Berdasarkan tabel 5 dapat disimpulkan bahwa susu bayi termasuk *cluster* rendah (*cluster* 1), karena dilihat dari jumlah status *cluster* yang dominan adalah rendah

Untuk menentukan *cluster* pada 10 produk lainnya, cara penentuannya yaitu sama seperti melakukan proses pada penentuan *cluster* untuk produk susu bayi yang bisa dilihat pada tabel 5. Berikut adalah hasil penentuan *cluster* untuk setiap produk, seperti pada tabel 6.

**Tabel 6. Hasil Penentuan Cluster**

Produk	Cluster
Susu Bayi	1
Susu Balita	2
Susu Anak – Dewasa	2
Susu Manula	2
Susu Ibu Hamil	2
Mie Goreng	3
Mie Kuah	2
Air 240ml	1
Air 330ml	1
Air 600ml	1
Air 1500ml	2

Keterangan:

Cluster 1 merupakan kategori rendah.

Cluster 2 merupakan kategori sedang.

Cluster 3 merupakan kategori tinggi.

Berdasarkan tabel 6 dapat dilihat bahwa produk yang tingkat penjualannya rendah adalah susu bayi, air mineral 240ml, air mineral 330ml dan air mineral 600ml. Produk yang tingkat penjualannya sedang adalah susu balita, susu dewasa, susu manula, susu ibu hamil, mie kuah dan air mineral 1500ml. Sedangkan produk yang tingkat penjualannya tinggi adalah mie goreng.

#### 4.1.1 Evaluasi validitas *cluster*

Pengujian validitas *cluster* menggunakan metode *Modiefied Partition Coefficient* (MPC). Tiap periode dataset diuji tingkat validitasnya. Pada ketetapan awal penulis menentukan 3 *cluster*. Disini penulis mencoba menguji dengan 2 sampai 5 *cluster* untuk sebagai pembandingan.

**Tabel 7. Hasil Validitas Cluster**

Jumlah Cluster	Nilai Validitas
2 cluster	0.628381
3 cluster	0.721286
4 cluster	0.751829
5 cluster	0.767736

Dari tabel diatas dapat dilihat bahwa nilai rata-rata tertinggi validitas didapat pada saat menggunakan 5 *cluster*. Kesimpulan dari tabel diatas adalah bahwa 5 *cluster* memperoleh nilai yang tertinggi dibanding *cluster* lainnya, pada umumnya jumlah *cluster* optimal ditentukan dengan mengambil *cluster* yang memiliki nilai MPC yang paling besar, yang berarti pada data transaksi swalayan ini cocok di uji dengan 5 *cluster*.

#### 4.2 Analisis hasil pengujian

Dari hasil yang telah diuji maka dapat disimpulkan bahwa :

Hasil dalam data transaksi swalayan keseluruhan (3 tahun) yang di proses oleh algoritma fuzzy c-means yaitu :

- a. Pada produk susu yang termasuk kategori rendah adalah jenis produk susu bayi ,dan yang termasuk kategori sedang adalah susu balita,dewasa,manula,dan ibu hamil.
- b. Pada produk mie yang termasuk kategori tinggi adalah jenis produk mie goreng dan yang termasuk kategori sedang adalah jenis produk mie kuah.
- c. Pada produk air mineral yang termasuk kategori rendah adalah jenis produk air mineral 240ml, 330ml dan 600ml dan yang termasuk kategori sedang adalah jenis produk air mineral 600ml dan 1500ml.

Dan dari hasil yang diuji dapat diketahui bahwa jumlah *cluster* juga mempengaruhi ukuran matriks u pada tahap membangkitkan bilangan random dan banyaknya jumlah *cluster* yang digunakan akan mempengaruhi nilai validitasnya, semakin banyak jumlah *cluster* semakin tinggi tingkat validitasnya.

### 5 Kesimpulan

Berdasarkan hasil yang telah diuji maka dapat disimpulkan :

1. Jumlah *clustering* akan mempengaruhi hasil dalam menentukan tingkat kategori rendah, sedang, dan tinggi pada data transaksi swalayan.
2. Proses *clustering* pada data penjualan swalayan menggunakan algoritma Fuzzy C-Means pada penelitian ini cocok karena dapat menghasilkan output berupa tingkat penjualan produk.
3. Pada pengujian validitas *cluster* dengan menggunakan MPC, 5 *cluster* memiliki jumlah nilai yang lebih tinggi dibandingkan dengan *cluster* lainnya. Maka hasil performansi yang baik dimiliki oleh 5 *cluster*. Yakni data transaksi swalayan ini cocok diuji dengan 5 *cluster*.
4. Jumlah *cluster* dapat mempengaruhi beberapa hal yaitu ukuran matriks u pada tahap membangkitkan nilai random, dan semakin banyak jumlah *cluster* yang digunakan maka semakin tinggi tingkat validitasnya.

Saran untuk penelitian selanjutnya :

1. Disarankan untuk memperbanyak jenis kategori, tidak hanya 3 jenis kategori produk. Dan akan lebih baik jika menggunakan data produk berdasarkan merk.

## Daftar Pustaka

- [1] Cakra Ramadhana, Yohana Dewi Lulu W, Kartina Diah K. W., Data Mining dengan Algoritma Fuzzy C-Means Clustering Dalam Kasus Penjualan di UD Subur Baru ,Artikel 2014.
- [2] Kennedy Tampubolon,Hoga Saragih,Bobby Reza (2013). Majalah Ilmiah Informasi dan Teknologi Ilmiah (INTI). Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat kesehatan Volume I,Nomor I, Oktober 2013.
- [3] Suyanto, Data Mining: Untuk Klasifikasi Dan Klasterisasi Data.,Bandung: Informatika,2017
- [4] Sandhi Yudha Charezita, Suyanto , Clustering Dokumen Bahasa Indonesia Dengan Menggunakan Fuzzy C-Means,Tugas Akhir 2012.
- [5] Titus Kristanto, Rachman Arief, Analisa Data Mining Metode Fuzzy Untuk Costumer Relationship Management Pada Perusahaan Tour & Travel,Seminar Nasional Sistem Informasi Indonesia 2013.
- [6] Catur Sugeng Pribadi,Sistem Pendukung Keputusan Pengadaan Buku Perpustakaan STIKOM Surabaya Menggunakan fuzzy C-Means Clustering,Tugas Akhir 2012.
- [7] Fitria Febrianti , Moh. Hafiyusholeh , Ahmad Hanif Asyhar,Perbandingan Pengklusteran Data IRIS Menggunakan Metode K-Means Dan Fuzzy C-Means,Jurnal Matematika"Mantik",Volume 02,Nomor 01 Oktober 2016.
- [8] Yohana Nugraheni, Data Mining Using Fuzzy Method For Costumer Relationship Management In Retail Industry,Lontar Komputer Vol .4 No.1 April 2013.
- [9] Kusrini, & Luthfi, e. t. (2013). Algoritma Data mining. Bogor.
- [10] Garcia-Molina, Hector; Ullman, JD., & Widom, Jennifer. 2002. Database systems the complete book, International edition. New Jersey, Prentice Hall.