

## **Kombinasi K-Means dan Fuzzy C-Means untuk Clustering Transaksi PPOB Berdasarkan Validitas Cluster**

**Nanda Setiawan<sup>1\*</sup>, Heru Fredi<sup>2</sup>, Bualazatulo Laia<sup>3</sup>, Yiska Dayanti Zagoto<sup>4</sup>, Johan<sup>5</sup>, Andreas Jorghy Parapat<sup>6</sup> Wahyu Saptha Negro<sup>7</sup>**

<sup>1,2,3,4,5,6,7</sup>Fakultas Teknik dan Ilmu Komputer, Universitas Potensi Utama, Medan, Indonesia

Email: <sup>1\*</sup>nanda.setiawansn27@gmail.com, <sup>2</sup>herufredi85@gmail.com, <sup>3</sup>buallalaia57@gmail.com, <sup>4</sup>zagotoyiska@gmail.com, <sup>5</sup>johanst1984@gmail.com, <sup>6</sup>parapatandreas@gmail.com <sup>7</sup>wahyusaptha17-7@gmail.com  
(\*Email Corresponding Author: nanda.setiawansn27@gmail.com)

Received: 27 Januari 2026 | Revision: 7 Februari 2026 | Accepted: 7 Februari 2026

### **Abstrak**

Perkembangan layanan *Payment Point Online Bank* (PPOB) telah mendorong peningkatan signifikan pada volume dan kompleksitas data transaksi digital yang dihasilkan. Algoritma *K-Means* merupakan salah satu metode *clustering* yang paling banyak digunakan karena kesederhanaan, efisiensi komputasi, dan kemampuannya dalam menangani data berskala besar. Tujuan penelitian adalah mengelompokkan data transaksi PPOB secara optimal menggunakan kombinasi algoritma *K-Means* dan *Fuzzy C-Means* (FCM) serta mengevaluasi kualitas *cluster* berdasarkan validitas *cluster*. Data yang digunakan dalam penelitian ini merupakan data transaksi *Payment Point Online Bank* (PPOB) periode Januari 2024 yang diolah menggunakan Google Colaboratory (Google Colab). Data tersimpan dalam format CSV dan berisi informasi transaksi yang dilakukan oleh berbagai loket PPOB dengan jumlah data: 498.853 data transaksi. Penerapan metode *Fuzzy C-Means* memberikan hasil yang lebih sesuai karena mampu merepresentasikan derajat keanggotaan ganda pada loket-loket yang berada di zona transisi antar *cluster*. Keberadaan zona transisi tersebut membuktikan bahwa pendekatan *Fuzzy* lebih tepat digunakan dalam konteks bisnis PPOB yang dinamis, di mana performa loket dapat berubah seiring waktu dan tidak selalu berada pada kategori yang bersifat mutlak.

**Kata Kunci:** Kombinasi *K-Means*, *Fuzzy C-Means*, *Clustering* Transaksi, Validitas *Cluster*.

### **Abstract**

The development of *Payment Point Online Bank* (PPOB) services has driven a significant increase in the volume and complexity of digital transaction data generated. The *K-Means* algorithm is one of the most widely used clustering methods due to its efficiency, computational efficiency, and ability to handle large data. The purpose of this study is to optimally cluster PPOB transaction data using a combination of the *K-Means* and *Fuzzy C-Means* (FCM) algorithms and produce cluster quality based on cluster validity. The data used in this study are *Payment Point Online Bank* (PPOB) transaction data for the January 2024 period processed using Google Colaboratory (Google Colab). The data is stored in CSV format and contains transaction information carried out by various PPOB counters with a total of 498,853 transaction data. The application of the *Fuzzy C-Means* method provides more appropriate results because it is able to represent the degree of double thickness at counters located in the transition zone between clusters. The existence of this transition zone proves that the *Fuzzy* approach is more appropriate for use in the dynamic context of the PPOB business, where counter performance can change over time and does not always fall into absolute categories.

**Keywords:** Combination of *K-Means*, *Fuzzy C-Means*, *Transaction Clustering*, *Cluster Validity*.

## **1. PENDAHULUAN**

Perkembangan layanan *Payment Point Online Bank* (PPOB) telah mendorong peningkatan signifikan pada volume dan kompleksitas data transaksi digital yang dihasilkan. Setiap transaksi PPOB merepresentasikan perilaku pengguna melalui pola frekuensi, nominal, dan waktu transaksi, sehingga data tersebut memiliki potensi besar untuk dianalisis dalam rangka memahami perilaku pengguna serta mendukung pengambilan keputusan strategis berbasis data [1] menunjukkan bahwa analisis perilaku berbasis data transaksi lebih mampu merepresentasikan karakteristik pengguna dibandingkan atribut demografis konvensional [2]. menegaskan bahwa frekuensi dan nominal transaksi merupakan indikator utama dalam mengidentifikasi segmen pengguna bernilai tinggi pada data transaksi keuangan. Namun, besarnya volume data serta sifat data transaksi yang heterogen dan saling tumpang tindih menyebabkan analisis konvensional menjadi kurang efektif, sehingga diperlukan pendekatan *data mining* berbasis *unsupervised learning*, khususnya teknik *clustering*.

Algoritma *K-Means* merupakan salah satu metode *clustering* yang paling banyak digunakan karena kesederhanaan, efisiensi komputasi, dan kemampuannya dalam menangani data berskala besar. Penerapan *K-Means* pada data transaksi digital terbukti mampu menghasilkan *cluster* yang kompak, terpisah dengan jelas, serta mudah diinterpretasikan. [3] membuktikan bahwa *K-Means* efektif dalam mengelompokkan data transaksi payment aggregator berdasarkan kontribusi pendapatan. Pada konteks PPOB, penelitian oleh [4] menunjukkan bahwa *K-Means* mampu mengelompokkan produk PPOB berdasarkan tingkat transaksi ke dalam beberapa kategori yang informatif. Meskipun

demikian, pendekatan ini masih terbatas karena *K-Means* bersifat *hard clustering*, sehingga belum sepenuhnya mampu menangkap karakteristik transaksi PPOB yang cenderung tumpang tindih antar segmen pengguna.

Sebagai alternatif, algoritma *Fuzzy C-Means* (FCM) menawarkan pendekatan *soft clustering* yang memungkinkan setiap data memiliki derajat keanggotaan pada lebih dari satu *cluster*. Pendekatan ini dinilai lebih fleksibel dalam merepresentasikan perilaku transaksi yang kompleks dan tidak selalu terpisah secara tegas. Aditya dan Fitriana (2021) menunjukkan bahwa FCM mampu menghasilkan segmentasi pelanggan yang adaptif melalui pemanfaatan derajat keanggotaan *Fuzzy*. [5] juga membuktikan bahwa penerapan FCM pada data transaksi berbasis model RFM menghasilkan segmentasi pelanggan dengan kualitas *cluster* yang baik. Namun, pada konteks transaksi PPOB, pemanfaatan FCM masih jarang dikaji secara mendalam dan umumnya digunakan secara terpisah tanpa dikombinasikan dengan metode *hard clustering*.

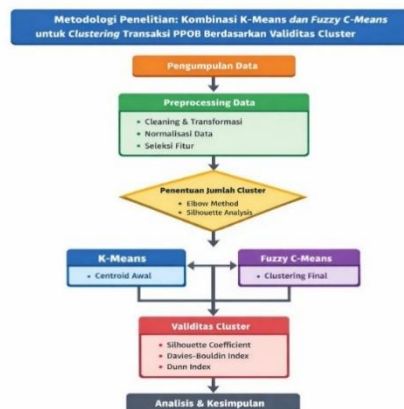
Perbedaan karakteristik antara *K-Means* dan *Fuzzy C-Means* menunjukkan adanya peluang pengembangan pendekatan yang menggabungkan keunggulan kedua metode tersebut. [6] menyatakan bahwa *K-Means* unggul dari sisi efisiensi dan kemudahan interpretasi, sedangkan *Fuzzy C-Means* lebih fleksibel dalam menangani data dengan karakteristik tumpang tindih. [7] juga menegaskan bahwa masing-masing metode memiliki keunggulan dan keterbatasan yang saling melengkapi. Namun, sebagian besar penelitian terdahulu masih berfokus pada studi komparatif, sehingga terdapat *research gap* berupa minimnya penelitian yang mengintegrasikan *K-Means* dan *Fuzzy C-Means* dalam satu kerangka *clustering* yang sistematis, khususnya pada data transaksi PPOB. Selain pemilihan algoritma, aspek validitas *cluster* juga masih menjadi celah penelitian. Sebagian penelitian menggunakan metrik validitas *cluster* seperti *Silhouette Coefficient* dan *Davies-Bouldin Index* hanya sebagai alat evaluasi hasil akhir, bukan sebagai dasar utama dalam menentukan pendekatan *clustering* yang optimal. [8] menunjukkan pentingnya evaluasi validitas *cluster* dalam menilai kualitas segmentasi, sementara [9] menegaskan bahwa perbandingan nilai validitas *cluster* dapat digunakan untuk menentukan metode *clustering* terbaik. Namun, hingga saat ini masih terdapat keterbatasan penelitian yang menjadikan validitas *cluster* sebagai landasan dalam menggabungkan *K-Means* dan *Fuzzy C-Means* pada data transaksi PPOB.

Berdasarkan uraian tersebut, terdapat beberapa *research gap* utama dalam penelitian ini, yaitu keterbatasan pendekatan integratif antara metode *hard* dan *soft clustering*, dominasi penggunaan satu algoritma *clustering* pada data transaksi PPOB, minimnya pemanfaatan *Fuzzy C-Means* dalam konteks PPOB, belum optimalnya penggunaan validitas *cluster* sebagai dasar pemilihan metode *clustering*, serta kurangnya kajian yang mengaitkan hasil *clustering* gabungan dengan karakteristik kompleks dan tumpang tindih pada data transaksi digital. Oleh karena itu, penelitian ini bertujuan untuk mengusulkan kombinasi algoritma *K-Means* dan *Fuzzy C-Means* dalam melakukan *clustering* data transaksi PPOB berdasarkan validitas *cluster*. Pendekatan ini diharapkan mampu menghasilkan segmentasi transaksi PPOB yang lebih akurat, informatif, dan relevan, serta memberikan kontribusi akademik dan praktis dalam pengembangan analisis data transaksi digital berbasis *unsupervised learning*. Dalam penelitian ini, algoritma *K-Means* digunakan sebagai tahap awal pengelompokan (inisialisasi), sedangkan *Fuzzy C-Means* digunakan untuk menyempurnakan hasil *clustering* dengan mempertimbangkan derajat keanggotaan data, yang selanjutnya dievaluasi menggunakan validitas *cluster*.

## 2. METODOLOGI PENELITIAN

### 2.1 Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan metode data mining, khususnya teknik *clustering*. Tujuan penelitian adalah mengelompokkan data transaksi PPOB secara optimal menggunakan kombinasi algoritma *K-Means* dan *Fuzzy C-Means* (FCM) serta mengevaluasi kualitas *cluster* berdasarkan validitas *cluster*. Alur penelitian pada jurnal ini adalah terlihat pada gambar berikut :



**Gambar 1.** Alur Metodologi Penelitian

### 2.2 Sumber dan Karakteristik Data

### 2.2.1 Sumber Data

Jenis data yang digunakan dalam penelitian ini adalah data sekunder, berupa data transaksi PPOB yang telah tersimpan pada sistem informasi penyedia layanan PPOB.

### 2.2.2 Atribut Data

Kemudian Sumber data diperoleh dari database transaksi PPOB milik [nama instansi / perusahaan / aplikasi PPOB], yang mencakup informasi transaksi selama periode tertentu, seperti:

1. ID Transaksi
2. Jenis Layanan (pulsa, listrik, PDAM )
3. Nominal Transaksi
4. Waktu Transaksi
5. Frekuensi Transaksi Pelanggan
6. Total Nilai Transaksi

## 2.3. Tahapan Penelitian

Metodologi penelitian dilakukan melalui beberapa tahapan utama sebagai berikut:

### 2.3.1 Pengumpulan Data

Pengumpulan data adalah proses sistematis dalam penelitian yang bertujuan untuk mengumpulkan informasi dari berbagai sumber yang relevan untuk menjawab pertanyaan penelitian yang telah ditetapkan [10], seperti observasi, wawancara, kuesioner, dan dokumentasi. Tahap ini pengumpulan data adalah tahap krusial dalam setiap studi karena memungkinkan peneliti menjawab pertanyaan penelitian, menguji hipotesis, dan mencapai tujuan studi [11]. Adapun metode yang digunakan dalam proses pengumpulan data ini yaitu seperti berikut :

#### a. Data Primer

Data primer adalah data yang diperoleh langsung dari sumber pertama oleh peneliti melalui proses pengumpulan data secara langsung di lapangan; data ini bersifat asli, sesuai kebutuhan penelitian, dan belum diolah oleh pihak lain [12]. Data yang digunakan data Transaksi PPOB bulan januari 2024 berjumlah 498.853 data.

### 2.3.2 Preprocessing Data

Data *preprocessing* mencakup pembersihan data, transformasi, reduksi, dan integrasi untuk menyiapkan data mentah menjadi data berkualitas tinggi sebelum analisis lebih lanjut. Tahapan ini diperlukan karena data mentah sering memiliki nilai hilang, noise, inkonsistensi, dan redundansi yang harus diatasi [13]. Berikut proses yang dilakukan yang terdapat pada tahapan *Preprocessing* data:

#### 1. *Cleaning* Data (Pembersihan Data)

Data *cleaning* adalah bagian penting dari data *preprocessing* yang digunakan untuk meningkatkan kualitas data dengan mengatasi nilai hilang, noise, dan inkonsistensi sebelum dilakukan analisis atau algoritma seperti *clustering* [14]. Data transaksi PPOB yang digunakan sering kali mengandung data tidak lengkap, duplikat, atau tidak konsisten sehingga perlu dilakukan pembersihan terlebih dahulu.

#### 2. Transformasi Data

Dalam sejumlah penelitian data mining dengan algoritma *K-Means*, *transformasi data* disebut sebagai bagian utama dari *preprocessing data* untuk menyiapkan dataset agar sesuai dengan kebutuhan analisis dan pengelompokan. Transformasi dilakukan setelah tahap pembersihan dan seleksi data sebelum proses *clustering* [15]. Pada tahap ini, data mentah yang masih bersifat kategorikal, temporal, atau tidak seragam diubah menjadi data numerik dan terstruktur.

#### 3. Normalisasi Data

Normalisasi data merupakan tahap *preprocessing* yang bertujuan untuk menyamakan skala nilai antar atribut sehingga tidak terjadi dominasi atribut tertentu terhadap hasil *clustering*. Algoritma *K-Means* dan *Fuzzy C-Means* menggunakan perhitungan jarak (*distance-based*), sehingga perbedaan skala data dapat menyebabkan hasil pengelompokan menjadi tidak akurat. Menunjukkan Tujuan dari normalisasi data adalah menyamakan rentang nilai dari setiap atribut, mengurangi bias akibat perbedaan skala, meningkatkan kualitas dan stabilitas hasil *cluster*.

### 2.3.3 Proses *Clustering*

Proses *clustering* merupakan tahapan utama dalam penelitian ini yang bertujuan untuk mengelompokkan data transaksi PPOB ke dalam beberapa *cluster* berdasarkan tingkat kemiripan karakteristik data. Proses ini dilakukan menggunakan algoritma *K-Means* dan *Fuzzy C-Means* (FCM) untuk memperoleh hasil pengelompokan yang optimal.

#### a. *K-Means Clustering*

*K-Means Clustering* merupakan salah satu metode *unsupervised learning* yang digunakan untuk mengelompokkan data ke dalam beberapa *cluster* berdasarkan tingkat kemiripan. Algoritma ini mengelompokkan data dengan cara meminimalkan jarak antara data dan pusat *cluster* (*centroid*). Menunjukkan pengelompokan data transaksi PPOB berdasarkan karakteristik yang serupa dan menyederhanakan data menjadi beberapa kelompok yang mudah dianalisis yang menjadi dasar inialisasi pada metode *Fuzzy C-Means*. Adapun langkah-langkah algoritma *K-Means* adalah sebagai berikut :

1. **Menentukan jumlah cluster (k)**

Nilai k ditentukan terlebih dahulu berdasarkan metode evaluasi cluster.

2. **Inisialisasi centroid awal**

Centroid dipilih secara acak dari data atau menggunakan metode tertentu.

3. **Perhitungan jarak**

Jarak antara data dan centroid dihitung menggunakan *Euclidean Distance*:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

4. **Pengelompokan data**

Setiap data dimasukkan ke cluster dengan jarak terdekat.

5. **Pembaruan centroid**

Centroid baru dihitung dari rata-rata data dalam setiap cluster :

$$C_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i \quad (2)$$

6. **Iterasi hingga konvergen**

Proses diulang sampai tidak ada perubahan cluster atau centroid.

Berikut contoh perhitungan manual *K-Means Clustering* yang sederhana dan sering dipakai dengan menggunakan data transaksi PPOB.

Data Awal

Misalkan digunakan 2 atribut:

1. X<sub>1</sub> = Total transaksi (juta rupiah)
2. X<sub>2</sub> = Jumlah transaksi

**Tabel 1.** Perhitungan manual K-Means

Data	Total Transaksi (X <sub>1</sub> )	Jumlah Transaksi (X <sub>2</sub> )
A	10	2
B	12	4
C	25	8
D	30	10

Jumlah cluster di tentukan K=2

Langkah 1: Menentukan Centroid Awal

Dipilih secara acak dari data:

1. **Centroid 1 (C1)** = Data A → (10, 2)
2. **Centroid 2 (C2)** = Data C → (25, 8)

Langkah 2: Menghitung Jarak (*Euclidean Distance*)

Rumus:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \quad (3)$$

Jarak Data ke Centroid

**Data A (10,2)**

1. ke C1:

$$\sqrt{(10 - 10)^2 + (2 - 2)^2} = 0 \quad (4)$$

2. ke C2 :

$$\sqrt{\{(10 - 25)^2 + (2 - 8)^2\}} = \sqrt{\{261\}} = 16.15 \quad (5)$$

→ Masuk *Cluster 1*

**Data B (12,4)**

1. ke C1:

$$\sqrt{\{(12 - 10)^2 + (4 - 2)^2\}} = \sqrt{\{8\}} = 2.83 \quad (6)$$

2. ke C2:

$$\sqrt{\{(12 - 25)^2 + (4 - 8)^2\}} = \sqrt{\{185\}} = 13.60 \quad (7)$$

→ Masuk *Cluster 1*

**Data C (25,8)**

1. ke C1:

$$\sqrt{\{(25 - 10)^2 + (8 - 2)^2\}} = 16.15 \quad (8)$$

2. ke C2:

$$\sqrt{\{(25 - 25)^2 + (8 - 8)^2\}} = 0 \quad (9)$$

→ Masuk *Cluster 2*

**Data D (30,10)**

1. ke C1:

$$\sqrt{\{(30 - 10)^2 + (10 - 2)^2\}} = \sqrt{\{464\}} = 21.54 \quad (10)$$

2. ke C2:

$$\sqrt{\{(30 - 25)^2 + (10 - 8)^2\}} = \sqrt{\{29\}} = 5.39 \quad (11)$$

→ Masuk *Cluster 2*

Langkah 3: Hasil *Cluster* Iterasi 1

1. **Cluster 1:** A, B
2. **Cluster 2:** C, D

Langkah 4: Menghitung *Centroid* Baru

**Centroid Baru Cluster 1**

$$C_1 \frac{10 + 12}{2}, \frac{2 + 4}{2} = (11,3) \quad (12)$$

**Centroid Baru Cluster 2**

$$C_2 \frac{25 + 30}{2}, \frac{8 + 10}{2} = (27.9,5) \quad (13)$$

### b. Fuzzy C-Means Clustering

*Fuzzy C-Means* (FCM) merupakan metode *unsupervised learning* yang digunakan untuk mengelompokkan data ke dalam beberapa *cluster* berdasarkan tingkat kemiripan dengan konsep *Fuzzy*. Berbeda dengan *K-Means*, FCM memungkinkan setiap data memiliki derajat keanggotaan pada lebih dari satu *cluster*. Tujuan penerapan metode ini adalah menghasilkan pengelompokan yang lebih fleksibel kemudian menangkan ambiguitas data transaksi PPOB dan menyempurnakan hasil *clustering* dari *K-Means*.

#### Konsep Dasar FCM

Setiap data  $x_i$  memiliki derajat keanggotaan  $\mu_{ik}$  terhadap *cluster* ke-K dengan nilai antara 0 dan 1, serta memenuhi:

$$\sum_{k=1}^c \mu_{ik} = 1 \quad (14)$$

#### Langkah-Langkah Algoritma Fuzzy C-Means

1. Menentukan parameter awal
  - a. Jumlah *cluster* ( $c$ )
  - b. Nilai *fuzziness* ( $m > 1$ , umumnya  $m=2$ )
  - c. Maksimum iterasi dan toleransi error
2. Inisialisasi matriks keanggotaan  
Nilai keanggotaan diinisialisasi secara acak atau menggunakan hasil *K-Means*.
3. **Menghitung pusat *cluster* (centroid)**

$$v_k = \frac{\sum_{i=1}^n (\mu_{ik})^m x_i}{\sum_{i=1}^n (\mu_{ik})^m} \quad (15)$$

$$\sum_i^n (\mu_{ik})^m = 1$$

4. Menghitung jarak data ke pusat *cluster*  
Menggunkan jarak *Euclidean*
5. Memperbaharui matriks ke anggotaan

$$\mu_{\{ik\}} = \frac{1}{\sum_j^c \left( \frac{d_{ik}}{d_{ij}} \right)^{\frac{2}{m-1}}} \quad (16)$$

6. Menghitung fungsi objektif

$$J = \sum_{i=1}^n \sum_{k=1}^c (\mu_{ik})^m m d_{ik}^2 \quad (17)$$

#### 2.3.4 Analisis dan Interpretasi Hasil

Analisis dan interpretasi hasil bertujuan untuk memahami makna dari *cluster* yang terbentuk serta mengevaluasi kualitas pengelompokan data transaksi PPOB. Tahap ini dilakukan dengan menganalisis karakteristik setiap *cluster* dan membandingkan hasil *clustering* menggunakan metode *K-Means* dan *Fuzzy C-Means*. Perbandingan Hasil *K-Means* dan *Fuzzy C-Means* dapat dilakukan berdasarkan:

1. Nilai centroid
2. Distribusi anggota *cluster*
3. Fleksibilitas pengelompokan
4. Nilai validitas *cluster*

Secara umum, *Fuzzy C-Means* memberikan hasil *clustering* yang lebih halus, sedangkan *K-Means* lebih sederhana dan cepat. Sehingga hasil analisis menunjukkan bahwa kombinasi *K-Means* dan *Fuzzy C-Means* mampu mengelompokkan transaksi PPOB secara efektif dan memberikan informasi yang bermanfaat dalam memahami perilaku pelanggan.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Deskripsi Data Penelitian (Dataset)

Data yang digunakan dalam penelitian ini merupakan data transaksi *Payment Point Online Bank* (PPOB) periode Januari 2024 yang diolah menggunakan *Google Colaboratory* (*Google Colab*). Data tersimpan dalam format CSV dan berisi informasi transaksi yang dilakukan oleh berbagai loket PPOB dengan jumlah data: 498.853 data transaksi. Unit analisis dalam penelitian ini adalah loket PPOB, bukan transaksi individual, sehingga data transaksi terlebih dahulu dilakukan proses agregasi berdasarkan identitas loket.

##### 3.1.1 Atribut Data

Atribut utama yang digunakan dalam proses *clustering* meliputi:

**Tabel 2.** Atribut Data

Atribut	Keterangan
Total_transaksi	Jumlah transaksi per loket
Total_nominal	Total nilai transaksi per loket
Rata_rata_transaksi	Rata-rata nilai transaksi
Total_admin	Total pendapatan admin loket
Jumlah_pelanggan	Jumlah pelanggan unik
Jumlah_produk	Variasi produk PPOB

#### 3.2 Pra-Pemrosesan Data

Tahapan pra-pemrosesan dilakukan di *Google Colab* menggunakan bahasa pemrograman *Python*, dengan langkah-langkah sebagai berikut:

##### 3.2.1 Pembersihan Data (*Data Cleaning*)

Menghapus spasi pada nama kolom. Memastikan tidak terdapat nilai kosong (*missing value*) pada atribut utama

```

# Menghilangkan spasi pada nama kolom
df.columns = df.columns.str.strip()

# Mengecek missing value
df.isnull().sum()

***
id_transaksi  0
periode      0
loket        0
id_pelanggan 0
nama_produk  0
tgl_transaksi 0
lembar       0
rptag        0
rpadmin      0
dtype: int64

```

**Gambar 2.** Pembersihan Data

##### 3.2.2 Agregasi Data

Data transaksi dikelompokkan berdasarkan identitas loket. Setiap loket direpresentasikan oleh nilai agregat dari seluruh transaksi

```
data_loket = df.groupby('loket').agg(
    total_transaksi = ('id_transaksi', 'count'),
    total_nominal = ('rptag', 'sum'),
    rata_rata_transaksi = ('rptag', 'mean'),
    total_admin = ('rpadmin', 'sum'),
    jumlah_pelanggan = ('id_pelanggan', 'nunique'),
    jumlah_produk = ('nama_produk', 'nunique')
).reset_index()

# Menampilkan hasil agregasi
data_loket.head()
```

	loket	total_transaksi	total_nominal	rata_rata_transaksi	total_admin	jumlah_pelanggan	jumlah_produk
0	01ALFIAN	1	3000	3000.000000	1000	1	1
1	06MULIADI	431	34197017	79343.426914	2160300	427	3
2	110AMR01	309	85726273	277431.304207	929500	301	11
3	110ZZCELL01	65	5894964	90691.753846	266500	65	1
4	120BAT01	156	22177954	142166.371795	416400	150	10

Gambar 3. Agregasi Data

### 3.2.3 Normalisasi Data

Normalisasi dilakukan menggunakan *Standard Scaler*. Bertujuan menyamakan skala antar variabel agar tidak terjadi bias dalam proses *clustering*. Normalisasi penting karena nilai total nominal dan jumlah transaksi memiliki rentang yang sangat berbeda.

```
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# konversi ke DataFrame
X_scaled = pd.DataFrame(X_scaled, columns=X.columns)

X_scaled.head()
```

	total_transaksi	total_nominal	rata_rata_transaksi	total_admin	jumlah_pelanggan	jumlah_produk
0	-0.586437	-0.408723	-0.083546	-0.651802	-0.578219	-1.056765
1	0.510691	0.120303	-0.059054	0.983114	0.516684	-0.544444
2	0.199413	0.917528	0.004495	0.051213	0.192840	1.504837
3	-0.423143	-0.317567	-0.055414	-0.450778	-0.413727	-1.056765
4	-0.190960	-0.065648	-0.038900	-0.337281	-0.195260	1.248677

Gambar 4. Normalisasi Data

Hasil Akhir Pra-pemrosesan Data

```
print("Jumlah Locket:", X_scaled.shape[0])
print("Jumlah Variabel:", X_scaled.shape[1])

*** Jumlah Locket: 2161
    Jumlah Variabel: 6
```

Gambar 5. Hasil Akhir Pra-pemrosesan Data

## 3.3 Proses Clustering Menggunakan K-Means

### 3.3.1 Penentuan Jumlah Cluster

Jumlah *cluster* ditetapkan sebanyak **3 cluster**, dengan pertimbangan: Karakteristik bisnis loket PPOB. Kemudahan interpretasi hasil. Penelitian-penelitian sejenis yang mengelompokkan agen/loket ke dalam kategori kecil, menengah, dan besar.

```

1) from sklearn.cluster import KMeans
2) from sklearn.metrics import davies_bouldin_score
3) k=3
4) # Inisialisasi model K-Means
5) kmeans = KMeans(
6)     n_clusters=k,
7)     random_state=42,
8)     n_init=10
9) )
10) # Proses clustering
11) cluster_kmeans = kmeans.fit_predict(X_scaled)
12)
13) # Menambahkan hasil cluster ke data loket
14) data_loket['cluster_kmeans'] = cluster_kmeans
15)
16) # Menampilkan hasil clustering
17) data_loket.head()

```

	loket	total_transaksi	total_nominal	rata_rata_transaksi	total_admin	jumlah_pelanggan	jumlah_produk	cluster_kmeans
0	01ALFIAN	1	3000	3000.000000	1000	1	1	1
1	06MULIADI	431	34197017	79343.426914	2160300	427	3	1
2	110AMR01	309	85726273	277431.304207	929500	301	11	1
3	110ZZCELL01	65	5894964	90691.753846	266500	65	1	1
4	120BAT01	156	22177954	142166.371795	416400	150	10	1

Gambar 6. Penentuan Jumlah Cluster

### 3.3.2 Hasil Clustering K-Means

Hasil *clustering* K-Means membagi loket PPOB ke dalam tiga kelompok utama:

*Cluster* Karakteristik Umum

*Cluster* 0 Aktivitas rendah, nominal kecil

*Cluster* 1 Aktivitas sedang, transaksi stabil

*Cluster* 2 Aktivitas tinggi, omzet besar

```

# Jumlah loket pada setiap cluster
data_loket['cluster_kmeans'].value_counts().sort_index()

```

cluster_kmeans	count
0	319
1	1841
2	1

dtype: int64

Gambar 7. Hasil Clustering K-Means

### 3.3.3 Evaluasi Validitas Cluster

Evaluasi kualitas *cluster* dilakukan menggunakan *Davies-Bouldin Index* (DBI). Nilai DBI yang diperoleh menunjukkan bahwa *cluster* yang terbentuk memiliki tingkat pemisahan yang cukup baik, namun masih terdapat tumpang tindih karakteristik antar *cluster*, khususnya antara *cluster* menengah dan *cluster* tinggi. Nilai *Davies-Bouldin Index* (K-Means): 0.6010510082164627.

```

# Mengambil centroid dalam bentuk DataFrame
centroid_kmeans = pd.DataFrame(
    kmeans.cluster_centers_,
    columns=X.columns
)

centroid_kmeans

```

	total_transaksi	total_nominal	rata_rata_transaksi	total_admin	jumlah_pelanggan	jumlah_produk
0	1.625049	1.026687	-0.034919	1.742735	1.616182	0.900459
1	-0.282298	-0.179531	-0.018747	-0.302731	-0.280761	-0.156027
2	-0.508437	1.797714	45.669197	-0.650666	-0.578219	-1.056785

```

# Menghitung Davies-Bouldin Index
dbi_kmeans = davies_bouldin_score(X_scaled, cluster_kmeans)

print("Nilai Davies-Bouldin Index (K-Means):", dbi_kmeans)

```

Nilai Davies-Bouldin Index (K-Means): 0.6010510082164627

Gambar 8. Evaluasi Validitas Cluster

### 3.4 Proses Clustering Menggunakan Fuzzy C-Means

Berbeda dengan K-Means yang bersifat tegas (crisp), Fuzzy C-Means (FCM) memungkinkan satu loket memiliki derajat keanggotaan pada lebih dari satu cluster. Penggunaan FCM didasarkan pada karakteristik data loket PPOB yang tidak bersifat hitam-putih, memiliki banyak loket di zona transisi, mengalami fluktuasi transaksi. Artinya dalam realitas PPOB loket bisa ramai, kadang sedang tidak mutlak Aktivitas kecil atau besar [16].

#### 3.4.1 Penentuan Parameter dan Penentuan Cluster Domain Fuzzy C-Means

Parameter yang digunakan dalam penelitian ini adalah:

**Tabel 3.** Cluster Domain Fuzzy C-Means

Parameter	Nilai	Keterangan
c	3	Jumlah cluster
m	2.0	Fuzziness coefficient
error	0.005	Toleransi error
maxiter	1000	Iterasi maksimum

```

c, u, ub, d, jm, p, fpc = fuzz.cluster.cmeans(
    X_fm,
    c=n_cluster,
    m=2.0,
    error=0.005,
    maxiter=1000,
    init=None
)
# Cluster dominan berdasarkan nilai keanggotaan tertinggi
cluster_fm = np.argmax(u, axis=0)
# Menambahkan hasil cluster FCM ke data loket
data_loket['cluster_fm'] = cluster_fm
data_loket.head()

```

loket	total_transaksi	total_nominal	rata_rata_transaksi	total_admin	jumlah_pelanggan	jumlah_produk	cluster_kmeans	cluster_fm
0	01ALFIAN	1	3000	3000.000000	1000	1	1	0
1	06MULLIADI	431	34197017	79343.426914	2160300	427	3	1
2	110AMR01	309	85726273	277431.304207	929500	301	11	1
3	110ZZCELL01	65	5894964	90691.753846	266500	65	1	0
4	120BAT01	156	22177954	142166.371795	416400	150	10	1

**Gambar 9.** Cluster Domain Fuzzy C-Means

#### 3.4.2 Derajat Keanggotaan Loket

Menunjukkan tingkat kedekatan loket terhadap setiap cluster.

```

# Membuat DataFrame derajat keanggotaan
membership = pd.DataFrame(
    u.T,
    columns=[f'cluster_{i}' for i in range(n_cluster)]
)
membership.head()

```

	cluster_0	cluster_1	cluster_2
0	0.914875	0.069977	0.015148
1	0.332455	0.484637	0.182908
2	0.144615	0.758044	0.097341
3	0.931365	0.056816	0.011819
4	0.172862	0.788616	0.038521

**Gambar 10.** Keanggotaan Loket

#### 3.4.3 Evaluasi Kualitas Cluster (Fuzzy Partition Coefficient)

Nilai FPC mendekati 1 cluster semakin baik, Nilai FPC rendah : cluster tumpang tindih

```

print("Fuzzy Partition Coefficient (FPC):", fpc)

```

```

Fuzzy Partition Coefficient (FPC): 0.69719568193722

```

**Gambar 11.** Fuzzy Partition Coefficient

### 3.4.4 Davies–Bouldin Index untuk FCM

Walaupun FCM bersifat *Fuzzy*, DBI dapat dihitung menggunakan *cluster* dominan. Artinya FCM menghasilkan keanggotaan ganda dan DBI membutuhkan *cluster* tunggal maka dipakai *cluster* domain sebagai pendekatan. Nilai DBI 1.0450578626244225

```

from sklearn.metrics import davies_bouldin_score

dbi_fcm = davies_bouldin_score(X_scaled, cluster_fcm)
print("Davies-Bouldin Index (FCM):", dbi_fcm)

```

\*\*\* Davies-Bouldin Index (FCM): 1.0450578626244225

**Gambar 12.** Davies–Bouldin Index untuk FCM

### 3.4.5 Statistik Deskriptif Tiap Cluster (FCM)

Statistik deskriptif tiap *cluster* digunakan untuk menggambarkan karakteristik rata-rata loket PPOB pada masing-masing *cluster* hasil *Fuzzy C-Means*

```

cluster_summary_fcm = data_loket.groupby('cluster_fcm').mean(numeric_only=True)
cluster_summary_fcm

```

cluster_fcm	total_transaksi	total_nominal	rata_rata_transaksi	total_admin	jumlah_pelanggan	jumlah_produk	cluster_kmeans
0	74.059542	9.700915e+06	331176.714441	2.902144e+05	73.162595	3.015267	1.000763
1	282.791277	3.255522e+07	153561.220521	1.075190e+06	271.783489	8.442368	0.828660
2	1053.985646	1.123801e+08	176190.863451	3.789616e+06	1043.038278	8.162679	0.000000

**Gambar 13.** Statistik Deskriptif Tiap Cluster (FCM)

## 3.5 Perbandingan K-Means dan Fuzzy C-Means

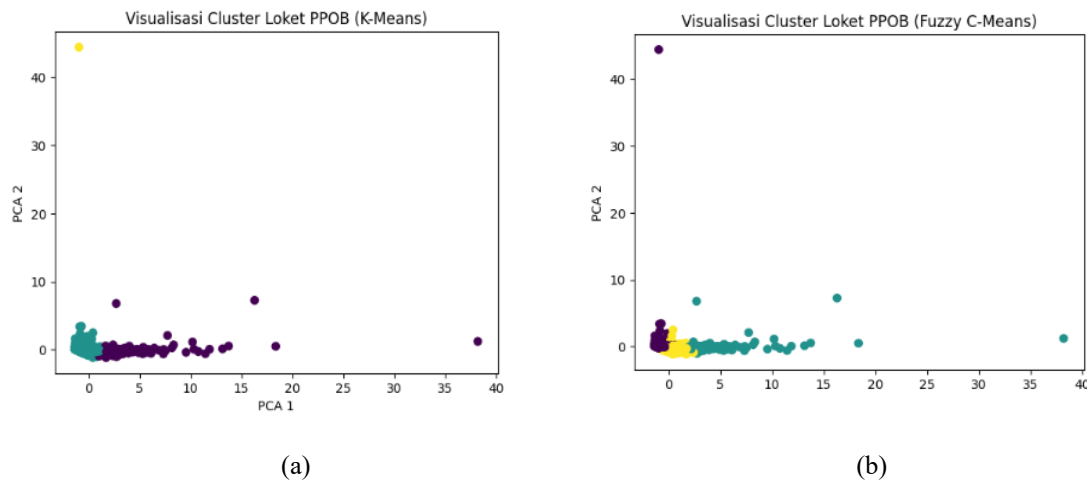
### 3.5.1 Perbandingan Karakteristik Metode

Tabel berikut menyajikan perbandingan karakteristik antara metode K-Means dan *Fuzzy C-Means* dalam pengelompokan loket PPOB.

**Tabel 4.** Perbandingan K-Means dan Fuzzy C-Means

Aspek Perbandingan	K-Means	Fuzzy C-Means
Tipe Keanggotaan	Tunggal (crisp)	Ganda ( <i>Fuzzy</i> )
Penugasan Cluster	Loket hanya masuk satu <i>cluster</i>	Loket memiliki derajat keanggotaan
Zona Transisi	Tidak terdeteksi	Terdeteksi
Fleksibilitas	Rendah	Tinggi
Interpretasi Bisnis	Sederhana	Lebih realistis
Kesesuaian Data PPOB	Cukup	Sangat sesuai
Nilai DBI	0.6010510082164627	1.0450578626244225

### 3.5.2 Perbandingan Hasil *Clustering* Berdasarkan Visualisasi *Cluster* Loker PPOB



**Gambar 14.** (a) Visualisasi Hasil *Clustering* Loker PPOB Menggunakan Metode K-Means (b) Visualisasi Hasil *Clustering* Loker PPOB Menggunakan Metode Fuzzy C-Means

Gambar 14. menyajikan perbandingan hasil *clustering* loket PPOB menggunakan metode K-Means dan Fuzzy C-Means (FCM) yang divisualisasikan dalam ruang dua dimensi hasil reduksi fitur menggunakan Principal Component Analysis (PCA). Hasil *clustering* dengan metode K-Means menunjukkan pembentukan *cluster* yang bersifat tegas (*hard clustering*), di mana setiap data loket hanya menjadi anggota satu *cluster*. Pola ini mengindikasikan adanya pemisahan *cluster* yang cukup jelas, namun beberapa data berada jauh dari pusat *cluster* utama, yang mencerminkan sensitivitas K-Means terhadap data ekstrem.

Sebaliknya, metode Fuzzy C-Means menghasilkan pembentukan *cluster* yang lebih fleksibel melalui penerapan derajat keanggotaan, sehingga satu data dapat memiliki keterkaitan dengan lebih dari satu *cluster*. Pendekatan ini memungkinkan representasi struktur data yang lebih halus, khususnya pada data loket PPOB yang memiliki karakteristik saling tumpang tindih. Berdasarkan hasil visualisasi tersebut, dapat disimpulkan bahwa K-Means sesuai untuk data dengan batas *cluster* yang jelas, sedangkan Fuzzy C-Means lebih adaptif dalam menangani data dengan pola distribusi yang kompleks.

### 3.6 Pembahasan Hasil

Hasil penelitian menunjukkan bahwa karakteristik loket *Payment Point Online Bank* (PPOB) bersifat heterogen, yang tercermin dari perbedaan signifikan pada volume transaksi, nilai nominal, jumlah pelanggan, dan variasi produk yang dilayani. Kondisi ini menyebabkan tidak semua loket dapat diklasifikasikan secara tegas ke dalam satu kelompok tertentu, sebagaimana terlihat pada hasil *clustering* K-Means yang masih menunjukkan tumpang tindih antar *cluster*.

Penerapan metode Fuzzy C-Means memberikan hasil yang lebih sesuai karena mampu merepresentasikan derajat keanggotaan ganda pada loket-loket yang berada di zona transisi antar *cluster*. Keberadaan zona transisi tersebut membuktikan bahwa pendekatan Fuzzy lebih tepat digunakan dalam konteks bisnis PPOB yang dinamis, di mana performa loket dapat berubah seiring waktu dan tidak selalu berada pada kategori yang bersifat mutlak.

### 3.7 Implikasi Hasil Penelitian

Hasil *clustering* loket PPOB menggunakan metode K-Means dan Fuzzy C-Means memberikan implikasi strategis yang signifikan bagi penyedia layanan *Payment Point Online Bank* (PPOB) dalam mendukung pengelolaan dan pengembangan jaringan loket. Segmentasi loket yang dihasilkan memungkinkan penyedia layanan untuk menentukan strategi insentif secara lebih tepat sasaran, di mana loket dengan performa tinggi dapat diberikan penghargaan atau program loyalitas untuk mempertahankan kinerjanya, sementara loket pada kategori menengah dapat diberikan insentif bertahap guna mendorong peningkatan aktivitas transaksi.

Selain itu, hasil *clustering* juga mempermudah identifikasi loket potensial, khususnya loket yang berada pada zona transisi dengan nilai keanggotaan Fuzzy yang tinggi pada lebih dari satu *cluster*, sehingga dapat diprioritaskan dalam program pengembangan atau ekspansi layanan. Di sisi lain, loket dengan tingkat aktivitas rendah dapat teridentifikasi secara objektif sehingga pembinaan dan pendampingan loket kurang aktif dapat dilakukan secara lebih terarah, misalnya melalui pelatihan operasional, perluasan jenis produk, atau dukungan promosi. Secara keseluruhan, pemanfaatan hasil *clustering* ini mendukung pengambilan keputusan berbasis data (*data-driven decision making*), yang memungkinkan penyedia layanan PPOB untuk merumuskan kebijakan dan strategi bisnis secara lebih efektif, adaptif, dan berkelanjutan sesuai dengan dinamika performa loket di lapangan.

#### 4. KESIMPULAN

Pada penelitian ini berhasil membandingkan kinerja *K-Means* dan *Fuzzy C-Means* untuk *Clustering* Transaksi PPOB Berdasarkan Validitas *Cluster*. Hasil analisis menunjukkan bahwa metode *Fuzzy C-Means* memberikan hasil yang lebih sesuai karena mampu merepresentasikan derajat keanggotaan ganda pada loket-loket yang berada di zona transisi antar *cluster*. Keberadaan zona transisi tersebut membuktikan bahwa pendekatan *Fuzzy* lebih tepat digunakan dalam konteks bisnis PPOB yang dinamis, di mana performa loket dapat berubah seiring waktu dan tidak selalu berada pada kategori yang bersifat mutlak. Untuk penelitian lanjutan disarankan agar metode *Fuzzy C-Means* dikembangkan agar transaksi PPOB lebih efisien.

#### REFERENCE

- [1] J. R. J. Thompson, L. Feng, R. M. Reesor, and C. Grace, *Know Your Clients' behaviours: a cluster analysis of financial transactions*. Department of Statistical and Actuarial Sciences: University of Western Ontario, 2020. Accessed: Jan. 28, 2026. [Online]. Available: <https://arxiv.org/abs/2005.03625>
- [2] S. Nofal, "Identifying highly-valued bank customers with current accounts based on the frequency and amount of transactions," *Heliyon*, vol. 10, no. 13, pp. 1–23, Jul. 2025, doi: 10.1016/j.heliyon.2024.e33490.
- [3] D. F. Suntoro, N. Fitriani, and A. Wibowo, "SEGMENTASI PENDAPATAN DARI PAYMENT AGGREGATOR MENGGUNAKAN METODE KLASTERISASI K-MEANS," *jicon*, vol. 13, no. 1, pp. 18–25, Mar. 2025, doi: 10.35508/jicon.v13i1.12691.
- [4] M. R. Dzirkillah, D. H. Gutama, D. P. Wijaya, and D. Danianti, "Implementasi Algoritma *K-Means* dalam Pengelompokan Tingkat Transaksi Produk PPOB Nusantara," *JIP*, vol. 10, no. 3, pp. 341–350, May 2024, doi: 10.33795/jip.v10i3.4917.
- [5] M. H. Zikri, S. Monalisa, and F. Muttakin, "Customer Segmentation Using the RFMD Model and *Fuzzy C-Means* Algorithm," *JSC*, vol. 7, no. 3, pp. 386–395, Dec. 2024, doi: 10.37396/jsc.v7i3.481.
- [6] S. Ghosh and S. Kumar, "Comparative Analysis of *K-Means* and *Fuzzy C-Means* Algorithms," *IJACSA*, vol. 4, no. 4, 2013, doi: 10.14569/IJACSA.2013.040406.
- [7] A. Abdulhafedh, "Applying *K-Means* Clustering and *Fuzzy C-Means* Clustering in Vehicle Crashes," *OALib*, vol. 12, no. 04, pp. 1–11, 2025, doi: 10.4236/oalib.1112856.
- [8] D. L. Aditya and D. Fitriannah, "COMPARATIVE STUDY OF FUZZY C-MEANS AND K-MEANS ALGORITHM FOR GROUPING CUSTOMER POTENTIAL IN BRAND LIMBACK," *JRisInf*, vol. 3, no. 4, pp. 327–334, Sep. 2021, doi: 10.34288/jri.v3i4.241.
- [9] L. Firlan, R. Fadila, M. K. Ridho, E. Sunandi, and U. Aflah, "Pendekatan Data Mining dalam Optimalisasi Margin Penjualan Adidas: Studi Klasterisasi dengan *K-Means* dan *Fuzzy C-Means*," *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 13, no. 2, pp. 229–237, Jul. 2025, doi: 10.37905/euler.v13i2.32417.
- [10] G. Daruhadi and P. Sopiati, "Pengumpulan Data Penelitian," *J-CEKI: Jurnal Cendekia Ilmiah*, vol. 3, no. 5, pp. 5423–5423, 2024, doi: <https://doi.org/10.56799/jceki.v3i5.5181>.
- [11] H. Taherdoost, "Data Collection Methods and Tools for Research; A Step-by-Step Guide to Choose Data Collection Technique for Academic and Business Research Projects," *International Journal of Academic Research in Management (IJARM)*, vol. 10, no. 1, pp. 10–38, 2021.
- [12] V. O. Ajayi, "A Review on Primary Sources of Data and Secondary Sources of Data," *SSRN Journal*, 2025, doi: 10.2139/ssrn.5378785.
- [13] P. Koukaras and C. Tjortjis, "Data Preprocessing and Feature Engineering for Data Mining: Techniques, Tools, and Best Practices," *AI*, vol. 6, no. 10, p. 257, Oct. 2025, doi: 10.3390/ai6100257.
- [14] A. P. Joshi and B. V. Patel, "Data Preprocessing: The Techniques for Preparing Clean and Quality Data for Data Analytics Process," *Orient. J. Comp. Sci. and Technol*, vol. 13, no. 0203, pp. 78–81, Jan. 2021, doi: 10.13005/ojcs13.0203.03.
- [15] Moh. Baha'Uddin and Z. Fatah, "Penerapan Data Mining Clustering *K-Means* Dalam Mengelompokkan Data Penduduk Penyandang Disabilitas," *Gudang Jurnal Multidisiplin Ilmu*, vol. 2, no. 10, pp. 86–94, 2024, doi: <https://doi.org/10.59435/gjmi.v2i11.1040>.
- [16] O. Gurler, "COMPREHENSIVE ANALYSIS OF FUZZY C - MEANS CLUSTERING AND ITS VARIANTS," *JMTE*, vol. 9, no. 1, pp. 69–93, Aug. 2024, doi: 10.62476/jmte9269.