

# Pengelompokan UMKM Batik Madura Menggunakan Metode K-Means Dan Sillhoutte Coefficient

Yeni Kustiyahningsih<sup>1\*</sup>, Achmad Khozaimi<sup>2</sup>, Jaka Purnama<sup>3</sup>

<sup>1,2</sup> Jurusan Teknik Informatika, Fakultas Teknik, Universitas Trunojoyo Madura, Jawa Timur

<sup>3</sup> Program Studi Teknik Industri, Fakultas Teknik, Universitas 17 Agustus 1945 Surabaya, Jawa Timur

Email: <sup>1\*</sup>ykustiyahningsih@trunojoyo.ac.id, <sup>2</sup>khozaimi@trunojoyo.ac.id, <sup>3</sup>jakapurnama@untag-sby.ac.id

(Naskah masuk: 16 Mar 2024, direvisi: 2 Jun 2024, diterima: 3 Jun 2024)

## Abstrak

UMKM merupakan salah satu sektor pendukung perekonomian di Indonesia. UMKM Batik Madura memberikan kontribusi yang cukup tinggi terhadap penyerapan tenaga kerja dan peningkatan ekonomi masyarakat daerah. Hal ini terbukti pada penyerapan tenaga kerja UMKM di Kabupaten Bangkalan Madura sebesar 210.003 dan Sampang sebesar 264.569. Permasalahan penelitian ini berkaitan dengan banyaknya UMKM Madura, sehingga menyulitkan Dinas Koperasi dalam menjalankan kebijakan pemerintah dalam memberikan pelatihan, bantuan pengembangan UMKM dan pendampingan. Tujuan penelitian adalah mengelompokan UMKM Batik Madura menjadi beberapa kluster menggunakan metode K-Means dan *Sillhoutte Coefficient*. Metode *K-Means* dapat melakukan pengelompokan berdasarkan data yang sama atau mempunyai *similarity* yang tinggi. Data UMKM akan dilakukan preprosesing terlebih dahulu untuk mengatasi data yang kosong dan normalisasi. Metode *Sillhoutte Coefficient* (SC) digunakan untuk menentukan jumlah kluster yang paling optimal. Pengelompokan UMKM Batik ini berdasarkan perpektif *balance scorecard* yaitu bisnis internal, keuangan, *learning and growth* dan pelanggan. Hasil *cluster* yang paling optimal adalah  $K=3$ . Nilai SC adalah sebesar 0,275, dengan 9 fitur dan  $SC = 0,403$  dengan 5 fitur, artinya dengan metode seleksi fitur *information gain* terjadi peningkatan nilai SC sebesar 0,128. Prosentase hasil pengelompokan adalah *cluster* 1 sebesar 15 %, *cluster* 2: 25 % dan *cluster* 3 : 60 %. Kategori pemetaan *cluster* 1 adalah sangat baik, *cluster* 2 baik dan *cluster* 3 sedang.

**Kata Kunci** : *K-Means Clustering*, *Sillhoutte Coefficient*, *Similarity*, UMKM Batik

## Grouping of Madurese Batik MSMEs Using the K-Means and Silhouette Coefficient Method

### Abstract

MSMEs are one of the supporting sectors for the economy in Indonesia. Madura Batik MSMEs make a fairly high contribution to the absorption of labor and improving the economy of regional communities. This is proven by the absorption of MSME labor in Bangkalan Madura Regency at 210,003, and Sampang at 264,569. The problem of this research is related to the large number of Madurese MSMEs, making it difficult for the Cooperative department to implement government policies in providing training, assistance for developing MSMEs and mentoring. The aim of the research is to group Madura Batik MSMEs into several clusters using the K-Means and Silhouette Coefficient methods. The K-Means method can group based on data that is the same or has high similarity. MSME data will be preprocessed first to overcome empty data and normalize it. To determine the most optimal number of clusters, the Silhouette Coefficient (SC) comparison method is used. The grouping of Batik MSMEs is based on a balanced scorecard perspective, namely internal business, finance, learning and growth and customers. The most optimal cluster result is  $K=3$ . The SC value is 0.275, with 9 features and  $SC = 0.403$  with 5 features, meaning that with the information gain feature selection method there is an increase in the SC value of 0.128. The percentage of grouping results is cluster 1 at 15%, Cluster 2: 25% and Cluster 3: 60%. Cluster 1 mapping category is very good, cluster 2 is good and cluster 3 is moderate.

**Keywords**: *K-Means Clustering*, *Sillhoutte Coefficient*, *Similarity*, SMEs Batik.

## I. PENDAHULUAN

Dalam kebijakan penanggulangan kemiskinan dan pengangguran, UMKM memberikan kontribusi yang paling besar. Hal ini karena UMKM dapat memajukan perekonomian Indonesia. Terdapat sekitar 90% usaha di Indonesia berasal dari UMKM dan memberikan kontribusi sebesar 11% terhadap total produksi di Jawa Timur. Saat ini terdapat sekitar 64,2 juta unit usaha yang tergolong dalam UMKM. Namun, sejalan dengan perkembangan teknologi dan persaingan usaha, pengembangan UMKM memiliki tantangan lebih besar [1]. Permasalahannya adalah berkaitan dengan kebijakan pemerintah mengenai bantuan dan pelatihan untuk mendukung pengembangan, peningkatan daya saing dan kapasitas UMKM. Karakteristik UMKM berdasarkan beberapa indikator yaitu jumlah tenaga kerja, jenis produksi, kapasitas produksi, satuan produksi, harga satuan, nilai produksi, nilai investasi, umur, pendidikan, surat izin usaha, motif, dan *marketplace* [2]. Data karakteristik terdapat data kategorikal dan numerik, sehingga perlu normalisasi sebelum data di kluster. Tujuan penelitian ini adalah mengelompokkan data UMKM berdasarkan karakteristik yang sama, atau tingkat kemiripan data yang sama sehingga memudahkan dinas dalam pengambilan keputusan untuk kebijakan strategis.

Sebelum data diklusterisasi atau dikelompokkan, data tersebut akan dilakukan *pre-processing* untuk menyeragamkan data dengan menggunakan *z-score* [3]. Setelah data ternormalisasi maka akan dilakukan pengelompokan dengan metode *K-Means Clustering*. Untuk menentukan jumlah kluster paling optimal di gunakan metode *Sillhoutte Coefficient* [4]. Kelebihan *K-Means* adalah dapat melakukan pembagian atau partisi data UMKM ke dalam beberapa kluster sesuai dengan karakteristik masing-masing data [4]. *K-Means* digunakan untuk pengelompokan UMKM pada dinas koperasi dan usaha mikro Kabupaten Pamekasan dan menghasilkan akurasi yang tinggi yaitu 89,47% [5]. Integrasi metode *balanced scoredcard* dan *Fuzzy Analytic Network Process* (FAHP) untuk mengukur kinerja UMKM. FAHP sangat sesuai untuk menentukan bobot indikator [6][7]. Penelitian lain mengenai Pengelompokan UMKM menggunakan metode *K-Means Clustering* dengan optimasi *Davies Bouldin Indexes* (DBI) dan *Cosine Similarity* sudah dilakukan, tetapi belum terdapat *pre-processing* data sehingga hasil pengelompokan mempunyai akurasi yang rendah [8]. Selanjutnya mengenai pemodelan klusterisasi UMKM menggunakan *K-Means* dengan mengklasifikasikan usaha menjadi mikro, kecil, dan menengah menjadi 3 kluster. penelitian ini belum terdapat kluster yang optimal, sehingga dikelompokkan hanya sesuai dengan [9].

Dengan penelitan di atas belum menerapkan metode optimasi kluster menggunakan *Silhouette Coefficient* dan pembobotan fitur dengan *information gain*. Kelebihan *Information Gain* adalah penerapannya sederhana dibandingkan metode seleksi fitur yang lain, sehingga banyak digunakan dalam pembobotan fitur hasil dari klusterisasi [10][11]. *Information Gain* merupakan metode yang dapat memberikan bobot sebagai perangkingan pada fitur.

*Information Gain* dapat menentukan fitur-fitur yang memiliki informasi terbanyak, Sehingga ditemukan fitur-fitur yang paling dominan [12][13][14]. Fitur yang akan terpilih dilihat dari nilai *Information Gain* terbesar [15]. Oleh karena itu penelitian ini akan mengembangkan model pengelompokan UMKM berbasis kluster menggunakan metode *K-Means clustering* dan *information gain* dengan metode optimasi *cluster* menggunakan *Sillhoutte Coefficient* dan *preprocessing* data menggunakan *zscore*

## II. METODE PENELITIAN

### A. Tahapan Penelitian

Tahapan penelitian dapat dilihat pada Gambar 1 yang terdiri dari:

1. Studi Literatur, Wawancara dan Kuisisioner  
Pada tahap ini dilakukan studi literatur dan jurnal dengan mempelajari teori-teori mengenai SPK, data mining, klustering, *Artificial Intelligence* (AI), *data pre-processing*, metode *K-Means*, metode *z-Score*, metode *Sillhoutte Coefficient* (SC) untuk menghitung jarak *similarity* dengan *Euclidean Distance*, *Minkowski*, dan *information gain*. Wawancara dan kuisisioner dilakukan kepada koperasi dan usaha mikro Kabupaten Bangkalan dan Pamekasan serta pengrajin UMKM Batik.
2. Pengumpulan Data  
Pada tahap pengumpulan data terdiri dari:
  - Data UMKM Batik yang digunakan adalah sebanyak 150 data.
  - Data yang digunakan sebagai objek penelitian ini adalah data UMKM Batik di Bangkalan dan Pamekasan Madura.
  - Data diperoleh dari koperasi, Disperindag dan kuisisioner kepada pemilik UMKM Batik.
  - Data indikator dari dinas dan pelaku UMKM sesuai dengan data pada Tabel 1.
3. Tahap pemodelan aplikasi pemetaan UMKM berbasis kluster. Kriteria yang digunakan dalam penelitian ini adalah sesuai dengan kebutuhan yang terdapat pada UMKM Madura. Tahap berikutnya adalah *pre-processing*. Sebelum data di klusterisasi atau dikelompokkan, data tersebut akan dilakukan *pre-processing* untuk menyeragamkan data menggunakan metode *Z-Score*. Setelah data ternormalisasi maka akan dilakukan pengelompokkan dengan metode *K-Means Clustering*. Algoritma *K-Means Clustering* dan *information gain* dapat dilihat pada Gambar 2. Tahap awal dilakukan normalisasi data UMKM Batik dan menentukan jumlah *cluster* yang akan digunakan sebagai titik pusat *centroid*. Menghitung jarak *Euclidean Distance*, menentukan hasil *cluster*, jika terjadi perubahan pada salah satu *cluster*, maka perhitungan akan dilanjutkan pada iterasi berikutnya dan seterusnya sampai tidak terdapat perubahan pada hasil *cluster*. Selanjutnya jika *dataset* sudah terbentuk *cluster*, maka akan di lakukan proses seleksi fitur menggunakan metode *information gain*. Hasil dari tahap pemodelan *information gain* adalah

pembobotan fitur semua kriteria klusterisasi UMKM, dan meranking kriteria dari terbesar ke kriteria terkecil.

4. Analisis Kebutuhan dan Perancangan Sistem  
Tahap ini terdiri dari spesifikasi kebutuhan *software* dan *hardware*, serta perancangan sistem untuk klusterisasi UMKM Batik. Spesifikasi perangkat keras terdiri dari:

- Prosesor *Genuine Intel Core i5*
- Memori RAM 4 GB
- *Hardisk* 1 TB

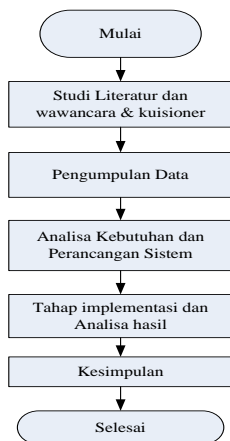
Spesifikasi Perangkat Lunak (*Software*)

- Sistem Operasi *Windows 10*
- *My SQL database XAMPP Versi 5.6*
- *My SQL Workbench*
- *Power Designer 12.5*
- *Microsoft Visio 2021*
- *Star UML*

Dalam pembuatan perangkat lunak ini akan di dapatkan keluaran sebagai berikut :

- Informasi *database* UMKM dari dinas koperasi dan Disperindag Bangkalan Madura.
- Rekomendasi dan klusterisasi UMKM berdasarkan metode *K-Means clustering*
- Bobot yang paling berpengaruh terhadap klusterisasi
- Implementasi, uji coba, dan analisis system

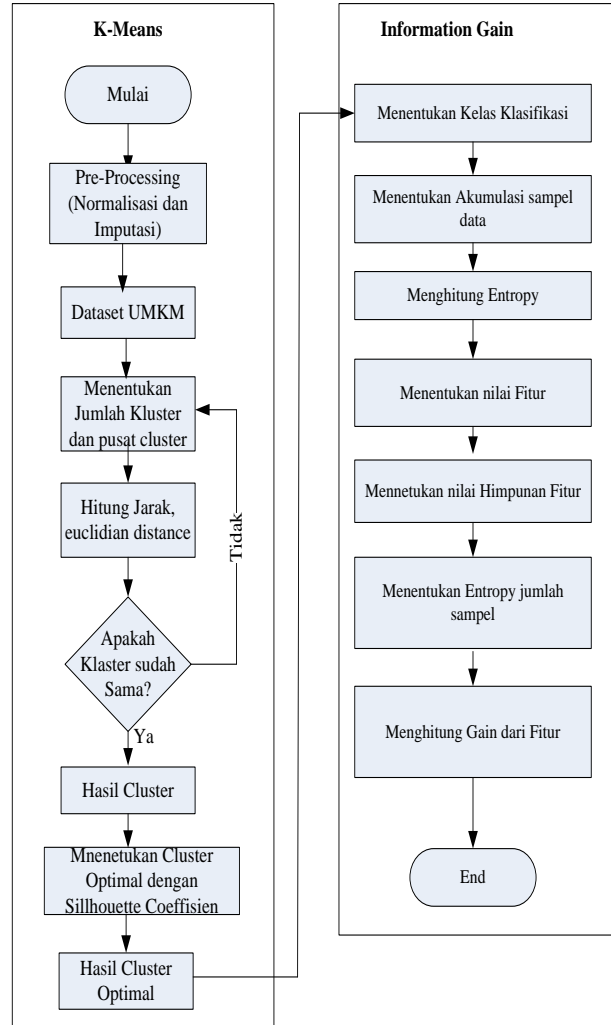
5. Tahap implementasi dan analisis hasil, akan dilakukan pada implementasi model sistem yang sesuai dengan rancangan dan analisa kebutuhan sistem yang telah dibuat sebelumnya. Hasil implementasi klusterisasi juga mempermudah pihak UMKM dan dinas dalam menentukan rekomendasi perbaikan masing-masing indikator, dan diarahkan pada peningkatan produktivitas UMKM Batik. Tahapan uji coba terdiri dari, *pre-processing* dengan membandingkan antara *z-score* dan *min-max*, selanjutnya menentukan jumlah *cluster* dan pusat *centroid*, untuk perhitungan jarak terdapat 2 metode yaitu *Euclidian Distance*, *manhattan*, selanjutnya adalah hasil *cluster*, dan dioptimasi menggunakan SC. Tahap akhir adalah hasil *cluster* akan dilakukan perankingan dengan *information gain*.



Gambar 1. Bagan Alir Penelitian

**B. Integrasi K-Means dan Information Gain**

Integrasi kedua metode dapat dilihat pada Gambar 2. menunjukkan bahwa data sebelum di *cluster*, dilakukan di *pre-processing*, setelah itu di *cluster*. *Cluster* yang optimal ditentukan menggunakan *Sillhouette Coefficient* dan hasilnya *cluster* digunakan untuk mennetukan pembobotan menggunakan *information gain*.



Gambar 2. Flowchart Algoritma K-Means dan Information Gain

**C. K-Means Clustering**

Algoritma *K-Means* dapat dibagi menjadi beberapa tahapan sebagai berikut [14].

1. Tentukan jumlah *cluster* yang akan dibuat dari kumpulan informasi.
2. Secara acak menentukan pusat *cluster*, titik pusat secara acak (*randomly*).
3. Hitung jarak setiap objek ke setiap pusat cluster menggunakan jarak *Euclidean* untuk menghitung jarak antara objek dan pusat massa.

$$D(x, y) = ||x - y||_2$$

$$= \sqrt{\sum_{i=1}^n (x_i - y_i)^2 ; i = 1,2,3, \dots, n} \tag{1}$$

Dimana:

$x_i$ =objek x ke-i

$y_i$ =centroid y ke-i

$n$ = banyaknya objek

4. Tetapkan setiap objek ke *centroid* terdekat.
5. Ulangi dan gunakan Persamaan 1 untuk menentukan posisi *centroid* baru.
6. Ulangi langkah 3 sampai 5 sampai Anda menemukan bahwa *cluster* konvergen adalah akhir dari *clustering*.

**D. Silhoutte Coefficient**

*Silhouette Coefficient* merupakan gabungan dari dua metode yaitu metode *cohesion* yang berfungsi untuk mengukur seberapa dekat relasi antara objek dalam sebuah *cluster*, dan metode *separation* yang berfungsi untuk mengukur seberapa jauh sebuah *cluster* terpisah dengan *cluster* lain [15]. Langkah metode ini adalah:

1. Hitung rata-rata jarak objek dengan semua objek lain yang berada di dalam satu *cluster*.

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j) \tag{2}$$

A= banyaknya data *cluster* A

2. Hitung rata-rata jarak objek dengan semua objek lain yang berada pada *cluster* lain, kemudian ambil nilai paling minimum dengan persamaan:

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} d(i, j) \tag{3}$$

C = banyaknya data *cluster* C

3. Hitung nilai *silhouette coefficient* dengan persamaan

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \tag{4}$$

Nilai s(i) berada antara -1 dan 1, di mana setiap nilai diinterpretasi sebagai berikut:

- s(i) ≈ 1 => data ke-i digolongkan dengan baik (dalam A)
- s(i) ≈ 0 => data ke-i berada di tengah antara dua klaster (A dan B)
- s(i) ≈ -1 => data ke-i digolongkan dengan lemah (dekat ke klaster B daripada A)

**E. Information Gain**

Tahapan *information gain* adalah sebagai berikut:

1. Menghitung nilai *information gain* pada setiap atribut.
2. Menentukan *threshold*, dengan tujuan untuk mempertahankan bobot yang sama dan diatas *threshold* serta membuat batas yang di bawahnya.
3. *Dataset* diperbaiki untuk mengurangi atribut.

**III. HASIL DAN PEMBAHASAN**

Hasil dan pembahasan penelitian ini adalah dimulai dari pengolahan *dataset*. *Dataset* yang digunakan adalah 10 kriteria yaitu new normal, jumlah konsumen, jumlah klaim konsumen, kerjasama dengan mitra, pelatihan karyawan pertahun, terdapat branding produk, pelatihan pemilik pertahun, memiliki surat ijin usaha, jumlah variasi motif batik, mempunyai marketplace, dengan 150 data UMKM Batik di Bangkalan, dapat dilihat pada Tabel 1. *Dataset* UMKM Batik.

Tabel 1. Transformasi Kriteria Pengelompokan UMKM

| Kriteria          | Nilai                    | Konversi Data |
|-------------------|--------------------------|---------------|
| Jumlah konsumen   | >=200                    | 4             |
|                   | 150-199                  | 3             |
|                   | 50-149                   | 2             |
|                   | <=50                     | 1             |
| Jumlah klaim      | 0                        | 4             |
|                   | 1-5                      | 3             |
|                   | 6-10                     | 2             |
| Kerjasama mitra   | >=5                      | 4             |
|                   | 2-4                      | 3             |
|                   | 1-2                      | 2             |
|                   | 0                        | 1             |
| New normal        | Masker;FS                | 4             |
|                   | Masker;Hand sanitizer;FS | 3             |
|                   | Masker                   | 2             |
|                   | Tidak Menggunakan        | 1             |
| Pel. karyawan     | 0                        | 1             |
|                   | 1 sampai 2               | 2             |
|                   | 3 sampai 6               | 3             |
|                   | 7 sampai 10              | 4             |
| Branding produk   | Terdapat merek           | 4             |
|                   | Terdaftar                | 3             |
|                   | Proses                   | 2             |
|                   | Tidak Bermerek           | 1             |
| Pelatihan pemilik | 0                        | 1             |
|                   | 1 - 2                    | 2             |
|                   | 3 - 6                    | 3             |
|                   | 7 - 10                   | 4             |
| Surat ijin usaha  | Tdk Ada                  | 1             |
|                   | Ditolak                  | 2             |
|                   | Proses Urus              | 3             |
|                   | SIUP                     | 4             |
| Jml motif batik   | <=20                     | 1             |
|                   | 30 Sampai 50             | 2             |
|                   | 60 sampai 90             | 3             |
|                   | >=100                    | 4             |
| Marketplace       | Tidak ada                | 1             |
|                   | 1                        | 2             |
|                   | 2                        | 3             |
|                   | 3                        | 4             |

Tahap selanjutnya adalah konversi data batik yang terdapat pada Tabel 2. Sebelum klasterisasi, data terlebih dahulu dilakukan *pre-processing* menggunakan normalisasi dan imputasi data menggunakan metode *average* yang dapat dilihat pada Tabel 3 *pre-processing dataset* UMKM.

Tabel 2. *Dataset* UMKM Batik

| Nama Batik    | Jumlah konsumen | Kerjasama mitra | ... | Pelatihan pemilik |
|---------------|-----------------|-----------------|-----|-------------------|
| Sumber Arafat | 4               | 3               | ... | 4                 |
| Tia Batik     | 4               | 3               | ... | 4                 |
| .....         | ...             | ...             | ..  | ...               |
| Zulpah Batih  | 3               | 2               | ... | 4                 |
| Rasiman Batik | 1               | 1               | ... | 1                 |
| Rusna Batik   | 1               | 1               | ... | 1                 |

Tabel 3. *Pre-processing Dataset* UMKM

| Nama Batik    | Jumlah Konsumen | Kerjasama Mitra | ... | Pelatihan Karyawan |
|---------------|-----------------|-----------------|-----|--------------------|
| Sumber Arafat | 1,367           | 1,374           | ... | 0,948              |
| Tia Batik     | 1,367           | 1,374           | ... | 0,948              |
| ...           | ...             | ...             | ... | ...                |
| Zulpah Batih  | 1,027           | 0,321           | ... | 0,143              |
| Rasiman Batik | -0,126          | -1,213          | ... | -1,467             |
| Rusna Batik   | -0,126          | -1,213          | ... | -1,467             |

Setelah tahap *pre-processing dataset* UMKM, maka selanjutnya ditentukan titik pusat *cluster* secara *random*, kemudian mencari jarak *cluster* terdakat menggunakan *Euclidean Distance*. Terdapat beberapa percobaan untuk pengujian klasterisasi yaitu K=2, K=3, K=4 sampai K=9. Dari beberapa uji coba dihasilkan *cluster* yang optimal menggunakan metode *Sillhoutte Coefficient*. Hasil dari nilai perhitungan *Sillhoutte Coefficient* masing-masing Kluster terdapat pada Tabel 4.

Tabel 4. Hasil Nilai *Sillhoutte Coefisien* Tanpa Seleksi Fitur

| Nilai K | Nilai <i>Sillhoutte Coefficient</i> |
|---------|-------------------------------------|
| 2       | 0,236                               |
| 3       | 0,275                               |
| 4       | 0,212                               |
| 5       | 0,203                               |
| 6       | 0,192                               |
| 7       | 0,195                               |
| 8       | 0,196                               |
| 9       | 0,195                               |

Tabel 3 menunjukkan bahwa *cluster* terbaik atau paling optimal adalah *cluster* 3, karena mempunyai nilai SC mendekati 1 yaitu 0,275. Grafik dari nilai SC masing-masing *cluster* dapat dilihat juga pada Gambar 3. Tabel 5 merupakan data hasil *cluster* yang paling optimal, yaitu K=3, dengan C1 terdapat 23 UMKM, C2 terdapat 32 UMKM dan C3 terdapat 90 UMKM, jika dipersentase 15 % UMKM pada *cluster* 1, 25 % UMKM di *cluster* 2 dan 60 % UMKM di *cluster* 3. karyawan pertahun, terdapat *branding* produk, jumlah variasi motif batik, menerapkan *new normal*, terdapat *marketplace*, kecuali indikator memiliki surat ijin usaha / SIUP lebih rendah dari *cluster* 2 dan 3. Berdasarkan hasil uji coba tersebut, maka *cluster* 1 merupakan pemetaan untuk UMKM yang baik, *cluster* 2: UMKM sedang, dan *cluster* 3: UMKM rendah. Selanjutnya dilakukan proses pembobotan menggunakan metode *information gain*. Hasil pembobotan masing-masing kriteria ini akan dijadikan sebagai rekomendasi perbaikan dari masing-masing UMKM, dengan melakukan perangkingan pada setiap *cluster*. Hasil pembobotan indikator dapat dilihat pada Tabel 5.

Tabel 5. Hasil Klasterisasi K=3 UMKM Batik

| Nama Batik    | Jumlah konsumen | Kerjasama mitra | .. | Cluster |
|---------------|-----------------|-----------------|----|---------|
| Sumber Arafat | 3,126           | 0,081           | .. | C1      |
| Tia Batik     | 3,367           | 0,081           | .. | C1      |
| ...           | ...             | ...             | .. | ...     |
| Zulpah Batih  | 1,563           | 0,081           | .. | C2      |
| Rasiman Batik | 0,021           | 0,034           |    | C3      |
| Rusna Batik   | 0,234           | 0,033           | .. | C3      |

Tabel 6. Pembobotan Fitur Menggunakan *Information Gain*

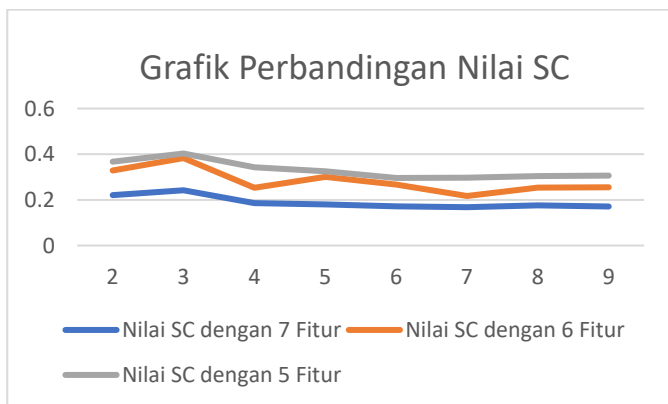
| Indikator                   | <i>Information Gain</i> |
|-----------------------------|-------------------------|
| <i>Branding</i> Produk      | 0,543                   |
| SIUP                        | 0,444                   |
| Pelatihan Karyawan          | 0,425                   |
| Kerjasama Mitra             | 0,396                   |
| <i>New Normal</i>           | 0,309                   |
| Jumlah Konsumen             | 0,282                   |
| Terdapat <i>Marketplace</i> | 0,253                   |
| Variasi Motif Batik         | 0,152                   |
| <i>Reward</i> Pelanggan     | 0,011                   |
| Jumlah Komplain             | 0,001                   |

Berdasarkan pembobotan dengan menggunakan *information gain*, terdapat 5 indikator yang paling berpengaruh terhadap pengelompokan UMKM Batik yaitu *Branding* Produk, SIUP, Pelatihan Karyawan, Kerjasama Mitra, dan *New Normal*. Terdapat perbandingan nilai SC antara sebelum dan sesudah seleksi fitur menggunakan *information gain*. Hasil perbandingan nilai SC dapat dilihat pada Tabel 7. Berdasarkan tabel tersebut dapat dilihat bahwa nilai SC optimal adalah pada K=3, dengan 10 fitur dan nilai SC = 0,275. Metode *information gain* digunakan untuk

menentukan seleksi fitur berdasarkan fitur yang mempunyai korelasi yang tinggi dengan pembobotan. Hasil SC seleksi 7 fitur adalah 0,242, 6 fitur = 0,383 dan 5 fitur = 0,403. Nilai SC yang paling baik adalah 5 fitur dan terjadi peningkatan nilai SC sebesar 0,128. Jadi nilai SC lebih tinggi setelah proses seleksi *information gain*. Persentase hasil pengelompokan adalah *cluster* 1 sebesar 15 %, *cluster* 2: 25 % dan *cluster* 3 : 60 %. Kategori pemetaan *cluster* 1 adalah sangat baik, *cluster* 2 baik dan *cluster* 3 sedang.

Tabel 7. Hasil Perbandingan SC *Information Gain*

| Nilai K | Nilai SC dengan 7 Fitur | Nilai SC dengan 6 Fitur | Nilai SC dengan 5 Fitur |
|---------|-------------------------|-------------------------|-------------------------|
| 2       | 0,221                   | 0,329                   | 0,367                   |
| 3       | 0,242                   | 0,383                   | 0,403                   |
| 4       | 0,186                   | 0,253                   | 0,343                   |
| 5       | 0,180                   | 0,301                   | 0,325                   |
| 6       | 0,172                   | 0,267                   | 0,296                   |
| 7       | 0,168                   | 0,217                   | 0,297                   |
| 8       | 0,176                   | 0,254                   | 0,304                   |



Gambar 3. Grafik Perbandingan Nilai SC Dengan *Information Gain*

Berdasarkan Gambar 3 maka dapat dilihat bahwa nilai SC pada K=3 dengan jumlah fitur 7, yaitu 0,242, jika jumlah fitur 6 = 0,383 dan jumlah fitur 5 = 0,403, sehingga dapat dilihat bahwa jumlah fitur yang paling optimal adalah 5. Artinya bahwa 5 fitur yang paling baik untuk klasterisasi UMKM Batik Bangkalan dan Pamekasan adalah *Branding* Produk, SIUP, Pelatihan Karyawan, Kerjasama Mitra, *New Normal*. 5 fitur tersebut direkomendasikan kepada dinas dalam menentukan langkah strategis dalam menerapkan kerja dinas, sehingga terdapat 5 indikator penting dalam menentukan bantuan, prioritas kerja, pelatihan maupun program lain yang mendukung peningkatan kinerja dan produktivitas UMKM.

#### IV. KESIMPULAN

Dari hasil penelitian dapat diambil kesimpulan bahwa metode *K-Means*, *Sillhoutte Coefficient* dan *Information Gain* dapat digunakan untuk menentukan pengelompokan data

UMKM Batik. Metode *Information Gain* digunakan untuk mengetahui pengaruh atribut dataset terhadap klasifikasi. Atribut yang memiliki pengaruh tinggi adalah atribut yang mempunyai nilai *information* yang besar. Berdasarkan percobaan menggunakan seleksi fitur dan tanpa seleksi fitur terjadi kenaikan nilai SC sebesar 0,128. Nilai SC yang paling baik adalah pada K=3 dengan jumlah fitur = 5.

#### UCAPAN TERIMA KASIH

Terima kasih kepada LPPM UTM, Dinas Koperasi Mikro Kab. Bangkalan, UMKM Batik Bangkalan, atas kontribusinya dan dukungannya dalam pelaksanaan penelitian ini.

#### REFERENSI

- [1] A. Halim, "Pengaruh Pertumbuhan Usaha Mikro, Kecil Dan Menengah Terhadap Pertumbuhan Ekonomi Kabupaten Mamuju," *J. Ilm. Ekon. Pembang.*, vol. 1, no. 2, pp. 157–172, 2020, [Online]. Available: <https://stiemmamuju.e-journal.id/GJIEP/article/view/39>.
- [2] D. Sugiri, "Menyelamatkan Usaha Mikro, Kecil dan Menengah dari Dampak Pandemi Covid-19," *Fokus Bisnis Media Pengkaj. Manaj. dan Akunt.*, vol. 19, no. 1, pp. 76–86, 2020, doi: 10.32639/fokusbisnis.v19i1.575.
- [3] A. Supriyatna and D. Ekaputra, "Metode Fuzzy Simple Additive Weighting ( Saw ) Dalam Pemilihan Ketua Osis," *J. PETIR*, vol. 10, no. 1, pp. 71–76, 2017.
- [4] F. F. Dahayu Widyadhana, Rina Budi Hastuti , Iqbal Kharisudin, "Perbandingan Analisis Klaster K-Means dan Average Linkage untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah," *Prism. Pros. Semin. Nas. Mat.*, vol. 4, pp. 584–594, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>.
- [5] W. Maulidi Molyono, S. Achmadi, and Y. Agus Pranoto, "Pemetaan Tambak Garam Serta Produksi Garam Pada Kabupaten Pamekasan Menggunakan K-Means Clustering," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 5, no. 2, pp. 794–799, 2021, doi: 10.36040/jati.v5i2.3736.
- [6] Y. Kustiyahningsih and J. Purnama, "An integrated approach to determine mapping of SMEs during Covid-19 pandemic," 1945.
- [7] Y. Kustiyahningsih, E. Rahmanita, and J. Purnama, "Integration balanced scoredcard and fuzzy analytic network process (FANP) for measuring performance of small medium enterprise (SME)," *J. Theor. Appl. Inf. Technol.*, vol. 94, no. 2, pp. 343–352, 2016.
- [8] A. Viloría and O. B. P. Lezama, "Improvements for determining the number of clusters in k-means for innovation databases in SMEs," *Procedia Comput. Sci.*, vol. 151, no. 2018, pp. 1201–1206, 2019, doi: 10.1016/j.procs.2019.04.172.
- [9] P. Puntoriza and C. Fibriani, "Analisis Persebaran UMKM Kota Malang Menggunakan Cluster K-means,"

- JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 86–94, 2020, doi: 10.33633/joins.v5i1.3469.
- [10] Y. Kustiyahningsih, Mula'ab, and N. Hasanah, "Metode Fuzzy ID3 Untuk Klasifikasi Status Preeklamsi Ibu Hamil," *Teknika*, vol. 9, no. 1, pp. 74–80, 2020, doi: 10.34148/teknika.v9i1.270.
- [11] A. Harris and A. E. Mintaria, "Komparasi Information Gain , Gain Ratio , CFs-Bestfirst dan CFs-PSO Search Terhadap Performa Deteksi Anomali," vol. 5, pp. 332–343, 2021, doi: 10.30865/mib.v5i1.2258.
- [12] A. A. Syafitri Hidayatul AA, Yuita Arum S, "Seleksi Fitur Information Gain untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor dan Naïve Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 9, pp. 2546–2554, 2018.
- [13] M. R. Maulana and M. A. Al Karomi, "Information Gain Untuk Mengetahui Pengaruh Atribut," *J. Litbang Kota Pekalongan*, vol. 9, pp. 113–123, 2015.
- [14] R. Uli and B. Sitanggang, "Penerapan Algoritma K-Means Tingkat Kesehatan Bayi ( Studi Kasus : Dinas Kesehatan Kota Binjai ) The Application of K-means Algorithm for Baby Health Level ( Study Case : Binjai City Public Health Office )," pp. 1–7, 2022.
- [15] R. Hidayati, A. Zubair, A. Hidayat Pratama, L. Indana, P. Studi Sistem Informasi, and F. Teknologi Informasi, "Silhouette Coefficient Analysis in 6 Measuring Distances of K-Means Clustering," *Techno.Com*, vol. 20, no. 2, pp. 186–197, 2021.