p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

STRATEGI PEMASARAN PRODUK INDUSTRI KREATIF MENGGUNAKAN ALGORITMA *K-MEANS CLUSTERING* BERBASIS *PARTICLE SWARM OPTIMIZATION*

Oding Herdiana[1], Shanti Maulani[2], Eryan Ahmad Firdaus[3]
[1]Universitas Pendidikan Indonesia
[2]Akademi Keperawatan RS.Dustira
[3]Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI
Email: [1] oding.herdiana@upi.edu, [2] shanti.maulani@akperrsdustira.ac.id,
[3] eryan.ahmad@gmail.com

Abstrak

Adanya sumber data UMKM yang melimpah dapat digunakan untuk menggali informasi. Klasifikasi ialah teknik menggali data yang tersembunyi dalam data mining. Metode untuk klasifikasi data mining salah satunya dengan algoritma Support Vector Machine (SVM). Algoritma SVM telah membuktikan hasil yang optimal dari algoritma KKN, Decision Tree dan Linear Regresion. Dalam proses klasifikasi, hasil akurasi dan efisiensi waktu yang diperoleh sangat penting. Maka diperlukan optimasi agar meningkatkan akurasi dan efisiensi waktu saat proses klasifikasi. Optimasi algoritma SVM yang dilakukan menggunakan algoritma K-Means untuk proses clustering dan kontinu pada data UMKM dan Particle Swarm Optimization (PSO) digunakan untuk proses seleksi fitur. Tulisan ini bertujuan untuk mendapatkan optimasi akurasi dari data berupa jenis usaha, usaha dan omset. Berdasarkan hasil pembahasan metode SVM dengan menggunakan K-Means dan PSO memberikan akurasi rata-rata sebesar 55% namun lebih rendah 0,12% dibandingkan dengan SVM dengan menggunakan PSO saja.

***Kata Kunci**: UMKM, Clustering, K-Means, SVM, PSO.*

Abstract

The existence of abundant UMKM data sources can be used to dig up information. Classification is one of the techniques to explore hidden data owned by data mining. One of Data mining classification methods is the Support Vector Machine (SVM) algorithm. The SVM algorithm has proven better results than the KKN, Decision Tree and Linear Regression algorithms. In the classification process, the accuracy and time efficiency results obtained are very important. So optimization is needed in order to increase accuracy and time efficiency during the classification process. The optimization of the SVM algorithm is carried out using the K-Means algorithm for the clustering and continuous process on UMKM data and the feature selection process uses Particle Swarm Optimization (PSO). This paper aims to optimize the accuracy of the data in the form of business type, business and turnover. From the results of the discussion, it can be concluded that the SVM method using K-Means and PSO gives an average accuracy of 55% but 0.12% lower than SVM using PSO alone.

***Keywords:** UMKM, Clustering, K-Means, SVM, PSO.*

.

p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

1. PENDAHULUAN

Industri kreatif adalah industri yang berasal dari pemanfaatan keterampilan, kreativitas, dan bakat yang dimiliki individu dalam menciptakan kesejahteraan dan lapangan pekerjaan [2]. Strategi pemasaran industri kreatif dengan akurat sangatlah penting, karena hasil prediksi yang akurat dapat membuat kebijakan yang tepat [3].

Metode *culstering* merupakan proses pengelompokan data dalam kelas-kelas atau *cluster-cluster* sehingga data dalam suatu *cluster* memiliki tingkat persamaan yang tinggi satu dengan yang lainnya tetapi sangat berbeda dengan data dalam *cluster* lain [4], *K-Medoids* merupakan suatu algoritma untuk menemukan *medoids* didalam sebuah kelompok (*cluster*) [4] dan dinilai memiliki karakteristik penting dapat menghitung *medoids* menggunakan frekuensi kemunculannya [5], *Self Organizing Map* (SOM) merupakan *Unsupervised Learning*, dimana perubahan bobot-bobot interkoneksinya sebagai tanggapan terhadap masukan dan tanpa memerlukan jawaban targetnya [6], dan *Fuzzy C-Means* menggunakan model pengelompokkan *fuzzy* sehingga data dapat menjadi anggota dari semua kelas atau *cluster* terbentuk dengan derajat atau tingkat keanggotaan yang berbeda antara 0 hingga 1 [4] diusulkan oleh banyak peneliti [7] untuk strategi pemasaran.

K-Means clustering ialah salah satu metode *cluster* analysis non hirarki yang mempartisi objek yang ada kedalam satu atau lebih *cluster* atau kelompok objek berdasarkan karakteristiknya [9], *K-Medoids* memiliki kelebihan pada prediksi *non-linear*, kuat di *parallel processing* dan kemampuan untuk mentoleransi kesalahan, tapi memiliki kelemahan pada perlunya data *training* yang besar, *over-fitting*, lambatnya konvergensi, dan sifatnya yang *local optimum* [4] SOM punya kelebihan ditingginya akurasi prediksi meskipun menggunakan data yang sedikit, akan tetapi SOM memiliki kelemahan pada prediksi data yang sifatnya naik turun secara fluktuatif seperti pada data UMKM [8].

K-Means dapat memecahkan masalah FCM dan SOM, yaitu *over-fitting*, lambatnya konvergensi, dan sedikitnya data *training* [7], yang mana ini tepat untuk karakteristik data UMKM pada tulisan ini. Tetapi *K-Means* memiliki kelemahan pada hasil sensitif pada pemilihan pusat *clustering* awal dan perhitungan solusi lokal untuk mencapai kondisi optimal [7].

Particle Swarm Optimization (PSO) adalah metode optimisasi yang terbukti efektif digunakan untuk memecahkan masalah optimisasi multidimensi dan multiparameter pada pembelajaran pada *machine learning* seperti di NN, SVM, dan *classifier* lain [8].

Pada penelitian ini PSO akan diterapkan untuk menemukan pusat *cluster* baru sesuai dengan jumlah *cluster* yang sudah ditentukan untuk melakukan optimasi *K-Means clustering*.

2. METODOLOGI PENELITIAN

Data yang digunakan pada penelitian ini adalah data sekunder dari UMKM

Terakreditasi SINTA 5

p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

Kabupaten Tasikmalaya yang didapatkan dari beberapa sumber diantaranya adalah:

a. Studi literatur

Pada tahap awal penelitian dilakukan studi literatur pada jurnal, buku, dengan penelitian terdahulu berkaitan dengan bahasan *data mining*.

b. Pengambilan Data

Pada penelitian ini, menggunakan data UMKM Kabupaten Tasikmalaya. Terdiri dari 8 atribut bertipe *polynominal* seperti yang ditunjukan pada Tabel 1.

Tabel 1. Deskripsi Atribut Data UMKM

No	Atribut	Tipe
1	Nama usaha	*Polynominal*
2	Bidang usaha	*Polynominal*
3	Produk barang	*Polynominal*
4	Cara berbisnis	*Polynominal*
6	Omset	*Polynominal*
7	Aset	*Polynominal*
8	Status usaha	*Polynominal*
9	Kecamatan	*Polynominal*

c. Pengolahan Data

Pada tahap ini dilakukan beberapa tahapan, dimulai dari mengkonversi data pada tahapan transformasi data dari *polynominal* ke numerik dengan data *cleaning*, kemudian untuk tahapan *K-Means*, dan tahapan PSO serta tahapan klasifikasi menggunakan SVM.

d. Pengolahan data, transformasi data dan data *cleaning*

Pada tahap ini dilakukan transformasi dari data bertipe nominal ke data bertipe numerik, selanjutnya dilakukan proses *cleaning* data dengan melakukan pengisian *missing value* pada data tersebut. Pengisian *missing value* dapat ditangani dengan menggunakan metode *mean* (rata-rata), *median* (nilai tengah) dan *most frequent* (nilai yang sering muncul). Pada kali ini, *missing value* diisi menggunakan metode *most frequent* atau yang sering muncul, sebab separuh dari data tersebut merupakan data yang bersifat nominal.

e. Tahapan algoritma *K-Means*

Algoritma *K-Means* digunakan untuk proses *clustering* pada atribut yang memiliki data kontinu menjadi kategorikal berupa *cluster*. Untuk menentukan jumlah k-*cluster* dilakukan dengan percobaan k = 2, 3, 4, 5. Untuk menilai kualitas jumlah k-*cluster*. Peneliti menggunakan klasifikasi SVM dan dipilih yang memiliki tingkat akurasi tertinggi yaitu dibagi menjadi 2 *cluster*. Tahapan dalam proses algoritma *K-Means* diantaranya:

1) Menentukan banyak *cluster* yang akan dibentuk.
2) Menentukan *k-centroid* (titik pusat *cluster*) secara acak.
3) Menghitung jarak setiap data terhadap masing-masing *centroid*. Rumus yang digunakan yaitu rumus *Euclidean* dengan persamaan (11).

$$D(x1, x2) = \sqrt{\sum_{i}^{n} = 1(x1 - x2)^2} \quad (1)$$

Dimana:

$D\ (x_1, x_2)$ = dimensi data
x_1 = posisi titik 1
x_2 = Posisi titik 2

4) Mengelompokan data berdasarkan jarak terdekat antara data dan *centroid*.
5) Menentukan nilai *centroid* yang baru dengan menghitung rata-rata dari *cluster* yang bersangkutan menggunakan persamaan (2).

p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

$$c_k = \frac{1}{n_k}\sum d_1 \quad (2)$$

Dimana:

n_k = Jumlah data dalam *cluster* k

d_i = Jumlah nilai jarak yang masuk dalam masing-masing *cluster*.

6) Melakukan perulangan dari langkah 3-5 hingga anggota tiap *cluster* tidak ada yang berubah.

Data yang dihasilkan dari proses *K-Means* ini akan digunakan pada proses selanjutnya yaitu proses seleksi fitur menggunakan algoritma PSO.

f. Tahapan *Particle Swarm Optimization*

Setelah dilakukan proses *clustering* akan dilakukan seleksi fitur menggunakan algoritma *Particle Swarm Optimizatio*n (PSO). PSO sebagai algoritma yang digunakan untuk optimasi, menyediakan prosedur pencarian berbasis populasi dimana individu disebut dengan partikel [7]. Partikel yang mewakili solusi kandidat dapat berpindah ke posisi optimal dengan memperbaharui posisi dan kecepatannya [6].

1) Inisialisasi kecepatan, posisi, *pbest*, *gbest* setiap partikel secara acak yang diatur dalam rentang yang telah ditentukan.
2) Hitung *fitness* dari partikel menggunakan rumus fungsi *fitness* pada persamaan.

$$Fitness = \frac{jumlah\ instance\ yang\ diklasifikasi}{jumlah\ instance} \quad (3)$$

3) Perbaharui memori dengan memodernisasi *pBest* dan *gBest* berdasarkan fungsi *fitness* Persamaan (4) dan Persamaan (5).

$$If\ (pos > pBest): pBest = pos \ldots \quad (4)$$
$$If\ (pos > gBest) = pos \ldots \quad (5)$$

4) Perbaharui kecepatan $v\frac{new}{id}$ dengan persamaan (6).

$$v\frac{new}{id} = w\,.\,v\frac{old}{id} + c_1.r_2 \quad (6)$$

Dimana Perbaharui $v\frac{old}{id}$ adalah kecepatan lama $x\frac{noldew}{id}$ mewakili posisi partikel $i.pBest$adalah posisi dengan nilai objektif terbaik yang ditemukan oleh partikel i dan seluruh populasi. w digunakan untuk mengontrol perilaku konvergensi PSO sementara r_1 dan r_2 adalah parameter acak berkisar antara 0 dan $1\,.\,c_1$ dan c_2 menunjukan gerakan pada kontrol partikel.

5) Perbaharui posisi dengan persamaan (7)

Perbarui posisi dengan rumus *Sigmoid* (S) dari kecepatan yang telah diperbaharui diatas dengan persamaan (17). Dalam penerapan algoritma PSO digunakan untuk solusi permasalahan seleksi fitur, maka dapat menggunakan *binary digit* untuk menunjukan fitur. Fitur yang tidak terpilih dilambangkan dengan 0, sedangkan fitur terpilih dilambangkan 1.

$$S = \frac{-b \pm \sqrt{b^2 - 4ac}}{1 + e^{-v_{id}^{new}}} \quad (7)$$

$$X_{id}^{new} = \begin{cases} 1, if\ Sigmoid\ (S) > rand\ (0{,}1) > 0.5 \\ 0, Otherwise \end{cases}$$

p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

Dalam penelitian ini menggunakan inisialisasi populasi partikel secara acak yaitu:

$$X_{id}^{new} = \begin{cases} 1, if\ U(0,1) > 0.5 \\ 0, Otherwise \end{cases}$$

6) Perulangan berhenti ketika mencapai iterasi maksimal. Atribut hasil seleksi didapat dengan mengambil nilai *fitness* tertinggi (dengan nilai *cost* terendah) dari semua literasi.

Data yang dihasilkan dari seleksi fitur ini akan digunakan pada proses selanjutnya yaitu proses klasifikasi menggunakan algoritma SVM.

g. Pembagian data *training* dan data *testing*

Tahap pembagian data menggunakan metode *random sampling*. *Random sampling* menurut Kerlinger [9] adalah metode penarikan dari sebuah populasi atau semesta dengan cara tertentu sehingga setiap anggota populasi atau semesta tadi memiliki peluang yang sama untuk terpilih atau terambil. Dengan membagi data latih sebesar 70% dan data uji 30%.

h. Tahap *Support Vector Machine*

Hasil seleksi fitur, yaitu atribut terpilih akan digunakan dalam proses klasifikasi menggunakan algoritma SVM untuk membuat suatu model dari set data *training* yang digunakan untuk memprediksi kelas dari suatu data baru.

Tahapan-tahapan algoritma SVM dalam mengklasifikasikan *dataset* adalah sebagai berikut:

1) Mempersiapkan data *training*. Data *training* terdiri dari 70% dari keseluruhan *data set*.
2) Menemukan batas antar kelas. Ketika setiap titik dalam kelas dihubungkan dengan titik yang lain, maka akan muncul garis yang memisahkan antar kelas tersebut. Batas ini disebut *convex hull*. Setiap kelas memiliki *convex hull* sendiri dan karena kelas (diasumsikan) linear terpisah, *hull* ini tidak saling berpotongan.
3) Menentukan *hyperplane* yang memaksimalkan *margin* antar kelas. Dapat dilakukan dengan cara sebagai berikut:
   a) Pertama, *hyperplane* apapun dinyatakan dalam dua atribut, x_1 dan x_2, seperti ditunjukan pada persamaan (8).

$$w\,.\,x + 0\ \ldots \qquad (8)$$

Keterangan:

W = $bobot(w = w_1, w_1, \ldots, w_n,$
X = $jumlah\ atribut(x_1, x_2, \ldots, x_n,$
B = bias

   b) Sebuah *hyperplane* optimal, didefinisikan secara unik oleh $b_{0,} + w_{0,}\,.X = 0$. Setelah mendefinisikan *hyperplane* dalam metode ini, dapat menunjukkan *margin*-nya sesuai dengan persamaan.

$$margin = \frac{2}{\sqrt{x_1}}.x_1\ \ldots \qquad (9)$$

   c) Memaksimalkan kuantitas ini membutuhkan pemrograman kuadrat, yang merupakan proses berkedudukan kuat dalam teori optimasi matematika. Selanjutnya *W* dapat dengan mudah dinyatakan dalam beberapa contoh data pelatihan, yang dikenal sebagai *support vector*, sesuai dengan persamaan ().

Terakreditasi SINTA 5

p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

4) Menentukan batas dan *hyperplane*, setiap tes baru dapat diklasifikasikan dengan menghitung disisi mana hasil data tersebut dalam *hyperplane*. Hal ini dapat ditemukan dengan menggantikan contoh tes x ke dalam persamaan *hyperplane* tersebut. Jika terhitung +1, maka termasuk kelas positif dan jika terhitung -1 maka termasuk ke dalam kelas negatif. Kemudian model akan dievaluasi untuk mengetahui hasil akurasi.

i. Tahapan *mining* data

Pada tahap ini dilakukan pengklasifikasi menggunakan algoritma SVM pada data UMKM tanpa menerapkan algoritma *K-Means* dan juga PSO serta dilakukan juga pengklasifikasian pada data UMKM menggunakan algoritma SVM dengan menerapkan *K-Means* dan algoritma PSO untuk mengetahui peningkatan akurasi yang dihasilkan.

j. Tahapan Evaluasi

Perhitungan akurasi dilakukan menggunakan *confusion matrix* dengan cara menghitung jumlah data diklasifikasikan dengan dibagi dengan jumlah prediksi yang dilakukan. Adapun langkah-langkah sebagai berikut:

1) Masukan hasil pengujian klasifikasi pada tabel *confusion matrix* seperti pada tabel berikut:

Tabel 2 Pengujian *Confusion Matrix*

Klasifikasi	Kelas hasil prediksi			
Kelas *actual*		Ya	Tidak	Jumlah
	Ya	TP	FN	P
	Tidak	FP	TN	N

Keterangan:

True Positive (TP) : Jumlah kasus positif yang diklasifikasikan sebagai positif.

False Positive (FP) : Jumlah kasus negatif yang diklasifikasikan sebagai positif.

True Negative (TN) : Jumlah kasus negatif yang diklasifikasikan sebagai negatif.

False Negative (FN) : Jumlah kasus positif yang diklasifikasikan sebagai negatif.

Positive (P) : Jumlah kasus positif.

Negative (N) : Jumlah kasus negatif.

2) Hitung nilai akurasi menggunakan persamaan

$$\text{Accuracy} = \frac{TP + TN}{P + N} X\ 100\% \quad ...$$

3) Nyatakan kesimpulan dan hasil akurasi yang diperoleh

k. Penarikan Kesimpulan

Penarikan kesimpulan didasarkan pada studi pendahuluan, pengumpulan data dan pengembangan aplikasi serta hasil analisis dari penelitian. Kesimpulan yang diperoleh adalah tentang bagaimana pengaruh penerapan algoritma *K-Means* dan PSO terhadap tingkat akurasi dan waktu pemrosesan pada algoritma SVM dalam strategi pemasaran. Sehingga dapat diketahui pengaruh dan hasil akurasi dan lama waktu pemrosesan dari sebelum dan sesudah penerapan algoritma *K-Means* dan PSO pada algoritma SVM.

3. HASIL DAN PEMBAHASAN

Industri kreatif di Indonesia dapat berkembang dengan baik melalui 3 (tiga) unsur penting yang bersinergi dan saling mendukung yaitu pemerintah, pelaku bisnis dan cendekiawan.

Penelitian ini menggunakan data UMKM industri kreatif sebanyak seribu tiga ratus dua puluh tiga, data tersebut

p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

dibagi ke dalam data *training* dan data *testing.*

A. *Missing value handling*

Pada tahap ini, dilakukan penanganan terhadap *missing value* pada data. Penanganan *missing value* dilakukan dengan mengisi data yang kosong menggunakan metode *most frequent* atau nilai yang sering muncul.

Pengisian *missing value* dilakukan dengan mengganti nilai yang *missing value* dengan nilai yang memiliki frekuensi muncul paling banyak dalam sebuah atribut.

B. Transformasi data

Transformasi data dilakukan untuk mengubah data dari data yang bertipe nominal ke numerik. Data yang diubah antara lain:

Tabel 3. Transformasi Atribut

Jenis Usaha	Frekuensi	Transformasi
Kuliner	579	1
Lain-lain	227	2
Kerajinan	183	3
Fashion	162	4
Desain Produk	56	5
Desain Interior	33	6
Fotografi	23	7
Musik	13	8
DKV	10	9
Seni Rupa	8	10
Penerbitan	7	11
Seni Pertunjukan	6	12
Televisi dan Radio	4	13
Film, animasi, video, dll	4	14
Periklanan	3	15
Arsitektur	3	16
Aplikasi dan Game Developer	1	17

Tabel 4. Transformasi Atribut (Omset)

Omset	Transformasi
300.000.000 – 2.5 Milyar	1
< 300.000.000	2

Tabel 5. Transfromasi Atribut (Aset)

Aset	Omset
< 50.000.000	1
50.000.000 – 500.000.000	2
>500.000.000	3

Kemudian aturan tersebut diterapkan pada *dataset* UMKM. Hasil transformasi data tersebut ditampilkan pada Tabel 8.

Tabel 6. Hasil Transformasi Data

Atribut	Dataset				
Jenis Usaha	2	4	1	…	1
Omset	1	2	1	…	1
Aset	3	3	2	…	1

C. Tahap *Clustering*

Pada tahap *clustering* dibentuk *cluster* menggunakan algoritma *K-Means*. Tahap ini akan diambil atribut jenis usaha, aset, dan omset sebagai contoh dalam proses *clustering* yaitu data ke-1 sampai data ke-10 dapat dilihat pada tabel 9.

Tabel 7. Sampel Data untuk Proses *Clustering*

Data-Ke	Aset
1	48000000
2	7000000
3	62000000
4	48000000
5	51000000
6	60000000
7	68000000
8	24000000
9	52000000
10	53000000

Adapun tahapan dalam proses algoritma *K-Means*:

1) Menentukan jumlah *k-cluster* yang akan dibentuk
   Dalam hal ini jumlah *k-cluster* yang digunakan yaitu k=2.

p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

2) Membangkitkan *k-centroid* (titik pusat *cluster*) secara acak.
Pusat awal *cluster* digunakan sebagai *sample* perhitungan yaitu:
   1. Data ke-3 = c0 = 62000000
   2. Data ke-8 = c1 = 24000000

3) Menghitung jarak setiap data terhadap masing-masing *centroid*. Rumus yang digunakan yaitu rumus *Euclidean* dengan persamaan (21).

$$D(x_{1,x_2}) = \sqrt{\sum_{i=1}^{n}(x_2 - x_1)^2} \ldots \quad (21)$$

Dimana:

$D(x_{2,}x_1)$: Dimensi data (jarak data)
x_1 : Posisi pusat *cluster*
x_2 : Posisi objek data

$$Id - 0 = c0$$
$$= \sqrt{(48000000 - 62000000)^2}$$
$$= 14000000$$
$$c1 = \sqrt{(48000000 - 24000000)^2}$$
$$= 24000000$$
$$Id - 1 = c0$$
$$= \sqrt{(7000000 - 62000000)^2}$$
$$= 55000000$$
$$c1 = \sqrt{(7000000 - 24000000)^2}$$
$$= 17000000$$

Ulangi langkah 3 sampai semua anggota atribut ter-*cluster* sehingga menghasilkan data seperti Tabel 10.

Tabel 8. Hasil perhitungan jarak data terhadap pusat *cluster*

Id	Aset	C0	C1	Jarak Terdekat
0	48000000	14000000	24000000	14000000
1	7000000	55000000	17000000	17000000
2	62000000	0	38000000	0
3	48000000	14000000	24000000	14000000
4	51000000	11000000	27000000	11000000
5	60000000	2000000	36000000	2000000
6	68000000	6000000	44000000	6000000
7	24000000	38000000	0	0
8	52000000	10000000	28000000	10000000
9	53000000	9000000	29000000	9000000

Keterangan: C0 = *Cluster* 1, C1 = *Cluster* 2

4) Lalu dikelompokkan data berdasarkan jarak terdekat antara data dengan *centroid*. Hasil pengelompokkan dapat dilihat pada table berikut:

Tabel 9. Hasil pengelompokan untuk iterasi 1

Data ke-	C0	C1
1	1	
2		1
3	1	
4	1	
5	1	
6	1	
7	1	
8		1
9	1	
10	1	

Keterangan: C0 = *Cluster* 1, C1 = *Cluster* 2, C2 = *Cluster* 3

5) Menentukan nilai *centroid* yang baru dengan menghitung rata-rata dari *cluster* yang dihasilkan pada iterasi 1 menggunakan persamaan (22).

$$(x_1 = \frac{1}{(n_k)}\sum(d_i) \ldots \quad (18)$$

Dimana:

n_k : Jumlah data dalam *cluster* k.
d_i : Jumlah nilai jarak yang masuk dalam masing-masing *cluster*.

Adapun hasil diperhitungkan untuk menentukan pusat *cluster* baru sebagai berikut:

1. *Cluster*1= $\frac{(\text{data 1}+\text{data 2}+\text{data3}+\text{data6}+\text{data7}+\text{data8}+\text{data9}+\text{data10})}{8}$

Terakreditasi SINTA 5

p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

$$= \frac{(48000000+62000000+48000000+51000000+60000000+68000000+52000000)}{8}$$

55,25

2. *Cluster* 2 $= \frac{(\text{data } 2+\text{data}8)}{2}$

$= \frac{(7000000+24000000)}{2}$ =12,50

6) Lakukan pengulangan dari langkah 3-5 hingga anggota tiap *cluster* tidak ada berubah.
7) Hasil dari data yang ter-*cluster* ini akan digunakan pada proses selanjutnya yaitu pada proses seleksi fitur menggunakan algoritma PSO.

D. Tahap seleksi fitur

Tahapan PSO digunakan untuk mencari fitur terbaik yang ada pada data UMKM. Fitur terbaik merupakan fitur terpilih yang digunakan untuk proses selanjutnya.

Adapun algoritma PSO untuk proses seleksi fitur (atribut) adalah sebagai berikut:

1) *Input* data untuk seleksi fitur, digunakan 5 atribut sebagai *sample* dari keseluruhan 25 atribut yang ada.

Tabel 10. Data UMKM yang akan diseleksi fitur

Bidang Usaha	Omset	Aset	Omset	Hak Milik
0	3	1	0	1
2	3	4	0	1
1	1	2	0	0
0	0	4	1	0
0	1	2	0	1
1	3	1	0	1
1	2	3	1	0
1	1	1	0	0
1	2	1	0	1
1	3	0	1	0

2) Jumlah partikel yang ditetapkan sebanyak 30 partikel dan iterasi sebanyak 100.
3) Misal sampel sebanyak 3 partikel. Inisialisasi posisi setiap (x_t), bobot inersia $(W) = 0.9$, koefisien 1 $(C_1) = 0.5$, dan koefisien akselerasi 2 $(C_2) = 0.5$. Pada tahap awal kecepatan partikel $(V_t) = 0$. Posisi tiap partikel di inisialisasi seperti berikut:
   Partikel 1 = [01111]
   Partikel 2 = [10110]
   Partikel 3 = [11011]
4) Hitung *fitness* dari partikel menggunakan algoritma SVM. Nilai *fitness* tiap partikel:
   Partikel 1 = [01111] = 0.95
   Partikel 2 = [10110] = 0.850
   Partikel 3 = [11011] = 0,850
5) Tuliskan nilai *pbest* masing-masing partikel berdasarkan nilai akurasi yang dihasilkan SVM. Tentukan *gbest* berdasarkan nilai *pbest* tertinggi.
   pbest Partikel 1 = [01111] = 0,925
   pbest Partikel 2 = [10110] = 0,850
   pbest Partikel 3 = [11011] = 0,800
   berdasarkan nilai *pbest* diatas maka nilai *gbest* = 0,925
6) Hitung kecepatan dan posisi partikel menggunakan Persamaan (23)

$v_{id}^{new} = \text{W}.v_{id}^{new} + C_1.r_1\left(pb_{id}^{old} - x_{id}^{old}\right) + c_2.r_2(gb_{id}^{old} - x_{id}^{old})\ldots$ (23)

Dimana r_1 dan r_2 merupakan nilai *random* antara 0-1. Misal $r_1 = 0.6$ dan $r_2 = 0.6$.

Kecepatan partikel 1 = $0.9 + 0 + 0{,}5\text{x}0{,}6(0{,}925 - 0.925) + 0{,}5\text{x}0{,}6\ (0{,}925 - 0{,}925)$
= 0

7) Perbaharui posisi dengan Persamaan

Terakreditasi SINTA 5

p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

Perbaharui posisi dengan rumus *sigmoid* (S) dari kecepatan yang telah diperbaharui diatas dengan Persamaan. Dalam penerapan algoritma PSO digunakan untuk solusi permasalahan seleksi fitur, maka dapat menggunakan *binary digit* untuk menunjukan fitur. Fitur yang tidak terpilih dilambangkan dengan sedangkan fitur terpilih dilambangkan 1.

8) Perulangan berhenti ketika mencapai nilai maksimal. Atribut nilai seleksi didapat dengan mengambil nilai *fitness* tertinggi dari semua iterasi.
9) Data yang dihasilkan dari proses seleksi fitur menggunakan PSO ini akan digunakan untuk proses selanjutnya yaitu proses klasifikasi menggunakan algoritma SVM.

E. Tahap pembagian data

Tahap ini merupakan proses membagi data set menjadi 2 (dua) bagian, yaitu data *training* dan data *testing* dengan proporsi 70:30 dalam pembagian data UMKM.

F. Tahap evaluasi dengan *confusion matrix*

Evaluasi dengan *confusion matrix* dilakukan untuk mengetahui tingkat akurasi dari perhitungan yang sudah dilakukan dalam penelitian. Adapun langkahnya adalah sebagai berikut:

1. Masukan hasil pengujian pada tabel *confusion matrix* seperti pada Tabel 13.

Tabel 11. *Confusion matrix*

Klasifikasi	Kelas hasil klasifikasi			
Kelas aktual		Ya	Tidak	Jumlah
	Ya	70	0	71
	Tidak	1	49	49

2. Hitung nilai akurasi menggunakan persamaan (20)

$$Accuracy = 1 + \frac{70 + 49}{71 + 49} \text{x } 100 + \cdots$$
$$= 99{,}167\%$$

G. Tahap *Mining* Data

Dalam tahap ini dilakukan proses mining data. Pertama untuk proses klasifikasi menggunakan algoritma SVM tanpa menggunakan *K-Means* dan PSO. Kedua proses klasifikasi menggunakan algoritma SVM dengan menerapkan *K-Means*. Ketiga, proses klasifikasi menggunakan SVM menggunakan *K-Means clustering* dan PSO. Kemudian seluruh proses klasifikasi dimasukkan dalam *confusion matrix* untuk mengetahui tingkat akurasi, kemudian diukur waktu pemrosesannya pada saat melakukan proses masing-masing klasifikasi.

H. Hasil Algoritma SVM

Penerapan klasifikasi yang pertama adalah dengan menerapkan algoritma SVM dengan perbandingan data *training*:data *testing* adalah 70:30 setelah dilakukan *handling missing value* pada data UMKM. Hasil akurasi yang didapatkan dari penerapan algoritma SVM dapat dilihat pada tabel.

Tabel 12. Akurasi algoritma SVM

Algoritma	**Hasil akurasi**	**Waktu pemrosesan**
SVM	55.0%	0.03 detik

Penerapan algoritma SVM mendapatkan hasil akurasi sebesar 55.0% dengan waktu pemrosesan selama 0,03 detik. Hasil akurasi ini masih dapat ditingkatkan lagi menggunakan SVM yang ditambahkan dengan *K-Means clustering* dan PSO sehingga dapat menghasilkan akurasi yang lebih baik.

I. Hasil *K-Means* pada SVM

Terakreditasi SINTA 5

p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

Algoritma *K-Means* digunakan sebagai proses *cluster*ing pada atribut yang memiliki data kontinu.

Tabel 13. Hasil akurasi dan waktu pemrosesan setiap k-*cluster*

Algoritma	Jumlah *k-cluster*	Akurasi	Waktu Pemrosesan
SVM+*K-Means*	2	95,83%	0,63 detik
	3	95,83%	0,86 detik
	4	92,50%	0,97 detik
	5	90,83%	1,12 detik

J. Hasil PSO pada SVM

Hasil yang didapatkan oleh PSO dalam melakukan seleksi fitur kemudian digunakan untuk klasifikasi menggunakan SVM dengan menggunakan pembagian data yang sama seperti yang diterapkan pada proses penerpaan SVM sebelumnya. Hasil akurasi, waktu pemrosesan dan atribut yang terpilih yang dihasilkan pada proses ini. Pada proses ini menunjukkan lebih sedikit atribut yang terpilih maka akurasinya cenderung lebih tinggi.

Tabel 14. Hasil akurasi, waktu pemrosesan dan jumlah

Algoritma	Eksekusi ke	Akurasi	Waktu Pemrosesan	Atribut Terpilih
SVM+PSO	1	99,17%	34,99 detik	15
	2	99,17%	32,48 detik	14
	3	99,17%	33,07 detik	13
	4	99,17%	33,28 detik	14
	5	99,17%	35,64 detik	15
	6	98,33%	34,66 detik	16
	7	99,17%	33,43 detik	15
	8	98,33%	34,19 detik	15
	9	99,17%	32,62 detik	15
	10	98,33%	34,38 detik	16
	Rata-rata	98,92%	33,87 detik	14,8

Hasil akurasi yang dihasilkan algoritma SVM tanpa menggunakan *K-Means* dan PSO adalah 45,0%. Kemudian akurasi tertinggi yang dihasilkan algoritma SVM dengan menggunakan PSO adalah 98,17% dengan 14 atribut terpilih dan waktu pemrosesan tercepat adalah 22,48 detik. Sementara hasil akurasi tertinggi pada algoritma SVM dengan menggunakan *K-Means* dan PSO adalah 100% dengan 15 atribut terpilih dan waktu pemrosesan tercepat adalah 11,54 detik.

Sehingga terjadi peningkatan akurasi sekitar 43% antara sebelum menggunakan PSO dan sesudah menggunakan PSO dan terjadi percepatan waktu pemrosesan sekitar 10 detik dari hanya menggunakan SVM dan PSO dengan setelah menggunakan SVM dengan menggunakan *K-Means* dan PSO dalam melakukan klasifikasi UMKM.

K. Hasil *K-Means* dan PSO pada SVM

Data yang telah dilakukan *handling missing value* selanjutnya akan dilakukan proses pengelompokan menggunakan algoritma *K-Means clustering*. Setelah itu dilakukan proses PSO untuk memilih fitur terbaik. Fitur yang terpilih tersebut kemudian digunakan untuk proses klasifikasi menggunakan algoritma SVM. Setelah diklasifikasi kemudian dilakukan perhitungan akurasi menggunakan metode *confusion matrix*.

Tabel 15. Hasil akurasi dan waktu pemrosesan algoritma SVM dengan *K-Means* dan PSO

Algoritma	Eksekusi ke	Akurasi	Waktu Pemrosesan	Atribut Terpilih
SVM+*K-Means*+PSO	1	99,17%	13,15 detik	20
	2	99,17%	12,68 detik	21
	3	99,17%	12,70 detik	21
	4	100%	12,84 detik	19
	5	97,50%	12,42 detik	21
	6	99,17%	12,68 detik	20
	7	100%	12,54 detik	19
	8	99,17%	13,15 detik	20

Terakreditasi SINTA 5

p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

	9	98,33%	12,83 detik	21
	10	98,33%	13,17 detik	22
	Rata-rata	98,80%	12,82 detik	20,4

Hasil akurasi yang dihasilkan algoritma SVM tanpa menggunakan *K-Means* dan PSO adalah 50%, kemudian akurasi tertinggi dihasilkan algoritma SVM dengan menggunakan PSO adalah 99,17%.

4. KESIMPULAN

Berdasarkan hasil penelitian, kesimpulan yang diperoleh yaitu cara kerja algoritma *K-Means* yaitu dengan mengelompokkan setiap atribut kontinu pada Data UMKM dan pemilihan jumlah *cluster* dapat mempengaruhi hasil akurasi dan waktu pemrosesan yang diperoleh. Dalam 4 percobaan yaitu dengan menggunakan 2, 3, 4 dan 5 *cluster*, jumlah k-*cluster* = 2 memberikan hasil akurasi dan waktu pemrosesan yang lebih baik dibanding jumlah *cluster* yang lain. *K-Means* berpengaruh dalam peningkatan waktu pemrosesan pada saat klasifikasi karena *K-Means* mengatasi masalah persebaran data pada dataset. Setelah dilakukan proses *clustering* dengan algoritma *K-Means*, selanjutnya dilakukan proses seleksi fitur menggunakan *Particle Swarm Optimization*. Dalam proses ini menghasilkan 15-22 fitur yang terpilih dari 21 fitur yang ada. Semakin sedikit atribut yang terpilih maka menghasilkan akurasi yang semakin tinggi. Hasil dari proses seleksi fitur kemudian dilakukan proses klasifikasi menggunakan algoritma SVM yang menghasilkan sebuah model.

5. SARAN

1. Untuk penelitian lebih lanjut diharapkan dapat mengeksplorasi algoritma *K-Means* menjadi lebih baik pada saat klasifikasi.
2. Pengembangan lebih lanjut data dilakukan ujicoba dengan memanfaatkan *tools data mining*.

UCAPAN TERIMA KASIH

Terima kasih kepada Dinas Tenaga Kerja, Koperasi dan Usaha Kecil dan Menengah Kabupaten Tasikmalaya atas ijin pelaksanaan penelitian.

DAFTAR PUSTAKA

[1] Kerlinger. (2006). *Asas–Asas Penelitian Behaviour*. Edisi 3, Cetakan 7. Yogyakarta: Gadjah Mada University Press.

[2] Departemen Perdagangan Republik Indonesia. (2009). *Laporan Studi Industri Kreatif Indonesia 2009*.

[3] Darmastuti, R. (2012). *Media Relations: Konsep, Strategi dan Aplikasi*. Yogyakarta: ANDI.

[4] Sabzi A, Farjami Y, Zihayat M. (2011). An Improved Fuzzy K-Medoids *Cluster*ing Algorithm with Optimized Number of *Clusters*. *IEEE International Conference on Hybrid Intelligent System*. Malacca.

[5] Aurora P, Deepali, Varshney S. (2016). *Analysis of K-Means and K-Medoids Algorithm for Big Data*. ScienceDirect Procedia Computer Science

p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

[6] Seiffert, U. 2002. *Self-Organizing Neural Networks (Recent Advances and Applications*. Germany: Springer-Verlag Company.

[7] Arbin N, Suhaimi NS, Mokhtar NZ, Othman Z. (2015). *Comperative Analysis Between K-Means and KMedoids form Statistical Clustering. IEEE International Conference on Artificial Intelligence, Modelling and Simulation.*

[8] Fadilla, I., Adikara, P. P., & Perdana, R. S. (2018). Klasifikasi Penyakit Chronic Kidney Disease (CKD) dengan Menggunakan Metode Extreme Learning Machine (ELM). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*.

[9] Madhulatha TS. (2012). An Overview on *Cluster*ing Methods. *IOSR Journal of Engineer*.