

## **CLUSTERING CREDIT CARD HOLDER BERDASARKAN PEMBAYARAN TAGIHAN MENGGUNAKAN IMPROVED K-MEANS DENGAN PARTICLE SWARM OPTIMIZATION**

Farhanna Mar'i<sup>1</sup>, Ahmad Afif Supianto<sup>2</sup>

<sup>1,2</sup>Program Studi Magister Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Brawijaya  
Email: <sup>1</sup>farhannamar@student.ub.ac.id, <sup>2</sup>afif.supianto@ub.ac.id

(Naskah masuk: 10 Juni 2018, diterima untuk diterbitkan: 07 November 2018)

### **Abstrak**

Kartu kredit merupakan salah satu bentuk media bagi nasabah untuk melakukan kredit dalam sebuah proses transaksi yang telah disetujui oleh bank bersangkutan. Bank harus selektif dalam menganalisis nasabah yang ingin mengajukan penerbitan kartu kredit untuk menghindari adanya kredit macet yang dapat menimbulkan kerugian pada bank, sehingga sangat penting untuk mengetahui karakteristik nasabah dengan melakukan *clustering*. Bank akan dapat mengambil keputusan untuk pertimbangan penerbitan kartu kredit dengan mencocokkan nasabah baru kedalam *cluster-cluster* yang telah dibentuk dan mengetahui kelayakan nasabah untuk diberikan akses kartu kredit dalam melakukan transaksi. *K-Means* adalah metode populer yang diterapkan dalam proses *clustering*. Tetapi, metode *K-Means* tidak dapat memberikan solusi optimum karena keterbatasannya dalam penentuan titik *centroid* yang optimal, sehingga untuk memperbaiki metode *K-Means* dalam penelitian ini digunakan salah satu algoritma evolusi yaitu *Particle Swarm Optimization* (PSO) untuk *generate* titik *centroid* optimum yang digunakan dalam proses perhitungan *K-Means*. Hasil pengujian dilakukan dengan membandingkan nilai *Silhouette Coefficient* dari *cluster* yang dibentuk menggunakan *K-Means* murni dan *Improved K-Means* dengan PSO yang menghasilkan nilai masing-masing yaitu 0,31614 dan 0,39484.

**Kata kunci:** *credit card holders, clustering, improved k-means, particle swarm optimization*

## **CLUSTERING CREDIT CARD HOLDER BASED ON BILLING PAYMENT USING IMPROVED K-MEANS WITH PARTICLE SWARM OPTIMIZATION**

### **Abstract**

*Credit card is one form of media for customers to credit in a transaction process that has been approved by the bank concerned. Banks should be selective in analyzing customers who want to apply for credit card issuance to avoid bad debts that can cause losses to banks, so it is very important to know the characteristics of customers by clustering. The Bank will be able to take decisions for credit card issuance by matching new customers into the established clusters and knowing the eligibility of customers to be granted credit card access in making transactions. K-Means is a popular method that is applied in the clustering process. However, the K-Means method can not provide the optimum solution because of its limitation in determining the optimal centroid point, so to improve the K-Means method in this research is used one of the evolution algorithm namely Particle Swarm Optimization (PSO) to generate optimum centroid point used in k-means calculation process. The test results were performed by comparing the coefficient silhouette values of the clusters formed using pure K-Means and Improved K-Means with PSO which yielded respective values of 0,31614 and 0,39484, respectively.*

**Keywords:** *credit card holders, clustering, improved k-means, particle swarm optimization*

### **1. PENDAHULUAN**

Kartu Kredit merupakan salah satu media pembayaran yang dimiliki oleh nasabah bank atau (*card holder*) dalam melakukan transaksi. Pemegang kartu kredit dapat melakukan pembayaran dengan

kartu kredit untuk membeli barang maupun jasa di toko atau (*merchant*) yang menyediakan alternatif pembayaran melalui kartu kredit (Lestari dkk., 2017). Kartu kredit diterbitkan oleh bank selaku (*issuer*) yang bersangkutan disertai dengan

persyaratan dan ketentuan yang berlaku. Penentuan penerbitan kartu kredit bagi nasabah merupakan masalah penting yang dihadapi oleh bank, karena selain memberikan keuntungan terhadap bank dengan adanya nilai bunga yang harus dibayarkan oleh nasabah sesuai dengan persyaratan yang telah disepakati. Penerbitan kartu kredit untuk nasabah juga dapat memberikan kerugian untuk bank apabila nasabah tersebut tidak membayar tagihan pada kartu kredit sesuai dengan kesepakatan yang telah ditentukan.

Bagi bank selaku (*issuer*) catatan rekam jejak seorang nasabah dalam berhutang dinilai penting sebagai salah satu penilaian untuk dapat menerbitkan kartu kredit (Lestari dkk., 2017). Namun, untuk nasabah baru yang tidak memiliki rekam jejak dalam berhutang atau kredit di bank yang bersangkutan akan sulit bagi bank untuk mengetahui *track record* nasabah tersebut, sehingga diperlukan sebuah model *cluster* untuk menganalisis karakteristik nasabah tersebut dan memasukkan data nasabah sesuai dengan *cluster* yang cocok.

*Clustering* adalah sebuah metode dalam data mining untuk pemrosesan data yang digunakan untuk menemukan pola terhadap data, sehingga data akan dapat dikelompokkan berdasarkan pola yang sesuai. Apabila data memiliki pola yang sama akan dijadikan satu di sebuah *cluster*, dan sebaliknya untuk data yang tidak berpola sama akan berada di *cluster* yang berbeda (Adolfsson dkk., 2018).

Metode populer yang diterapkan untuk proses *clustering* data adalah metode *K-Means*. Namun, *K-Means* memiliki kelemahan dalam penentuan titik *centroid* awal yang dilakukan secara random, sehingga tidak bisa didapatkan titik *centroid* optimum. Titik *centroid* tersebut akan menentukan seberapa baik sebuah *cluster* yang dapat diukur menggunakan perhitungan rumus *Silhouette Coefficient* (Paramartha dkk., 2017).

Untuk mendapatkan nilai *centroid* yang optimum, diperlukan sebuah algoritma heuristik yang telah terbukti dapat memberikan optimasi terhadap suatu nilai. *Particle Swarm Optimization* (PSO) adalah kelompok dari algoritma *swarm intelligents* yang bersifat *meta-heuristik* dan dapat digunakan untuk optimasi titik *centroid* pada metode *K-Means*, sehingga akan didapatkan *cluster-cluster* yang tepat (Tan, L., 2015).

Dalam metode *K-Means*, juga perlu ditentukan terlebih dahulu jumlah *cluster* atau nilai *K* yang akan dibuat. Dalam penelitian ini, untuk menemukan nilai *K* yang tepat digunakan metode *elbow* dengan menghitung nilai dari SSE (*Sum of Square Error*) dari masing-masing *cluster* sehingga dapat memberikan hasil berupa nilai *K* yang optimal untuk sebuah data (Irwanto dkk., 2012). Dalam penelitian ini, *clustering* dilakukan menggunakan metode *K-Means* dengan jumlah *cluster* yang penentuannya didapatkan dari perhitungan nilai SSE yaitu sejumlah 2 *cluster*. Sedangkan, untuk mendapatkan

titik *centroid* yang optimum dari kedua *cluster* tersebut, metode PSO juga digunakan dalam penelitian ini sehingga dataset default pembayaran nasabah pemegang kredit dapat diklasterisasi secara optimum. Data yang digunakan dalam penelitian ini merupakan dataset *default* pembayaran dari nasabah pemegang kartu kredit sejumlah 30.000 nasabah yang memiliki 24 features yang didapatkan dari sumber terbuka.

Belakangan ini, analisis data pelanggan dan pengembangannya dilakukan pada beberapa industri perbankan yaitu klasifikasi maupun *clustering*, baik dalam bidang analisis Bank *Direct Marketing* pada nasabah yang menerima kredit (Oktanisa dan Supianto, 2018) maupun analisis kartu kredit (Viji dan Banu, 2018). Pada penelitian sebelumnya dalam analisis kartu kredit ada beberapa topik penelitian salah satunya yaitu deteksi penipuan pada transaksi menggunakan kartu kredit atau *fraud detection* yang menggunakan pendekatan *clustering* dengan metode *K-Means* yang berhasil melakukan *clustering* pada transaksi kartu kredit menjadi 4 cluster yaitu *low, high, risky, and high risky* (Vaishali, 2014). Selain itu, juga telah dilakukan klasifikasi untuk *fraud detection* pada *e-tail merchant* menggunakan perbandingan beberapa metode *machine learning* seperti *random forest, logistic regression, dan support vector machine* yang menghasilkan hasil akurasi yang baik (Carneiro, N. dkk. 2017).

Topik penelitian lain yang membahas tentang analisis kartu kredit adalah *behavioral analysis dan default payment analysis*. Pada penelitian sebelumnya dalam *behavioral analysis* telah dilakukan prediksi penggunaan kartu kredit oleh pemegang kartu dengan beberapa faktor seperti pekerjaan dan kebiasaan pemegang kartu menggunakan kartu kredit. Dalam penelitian tersebut terfokus pada keterkaitan antara pendapatan *customer* dan penggunaan penuh *credit limit* yang dimiliki oleh *customer* (Dewri, L. V. dkk. 2016).

Sedangkan pada penelitian ini akan fokus pada topik *default payment analysis* yang sebelumnya juga pernah dilakukan penelitian menggunakan data yang sama yaitu klasifikasi *customer default payments* menggunakan metode data mining sebagai perbandingan untuk mengetahui metode yang memiliki nilai akurasi terbaik (Yeh dan Lien, 2009). Pada penelitian ini akan dilakukan *clustering* untuk mengelompokkan *customer* dengan kemiripan berdasarkan *default payment* dari *customer* tersebut.

Metode yang digunakan dalam penelitian ini adalah Metode *K-Means* yang pada penelitian sebelumnya juga pernah digunakan untuk *clustering personal credit analysis* untuk mendapatkan analisis trend dari kebiasaan *customer* yang digunakan oleh bank untuk menganalisis kelompok dari *customer* (Han dan Chai, 2012). Untuk meningkatkan performa dari metode *K-Means*, pada penelitian sebelumnya telah dilakukan *Improve* menggunakan

beberapa metode seperti Hidden Markov Model (Bhati dan Sharma, 2015) (Viji dan Bhanu, 2018)(Kumari dkk., 2017). Improve pada K-Means juga dapat dilakukan menggunakan *Evolutionary Computing* seperti algoritma genetika yang dihybrid menjadi *Genetic K-Means* (Chougule dkk., 2015). Sehingga pada penelitian ini dilakukan *improve* pada metode *K-Means* menggunakan metode yang termasuk dalam *Evolutionary Computing* dan juga *swarm intelligents* yaitu *Particle Swarm Optimization* (PSO). Pada penelitian sebelumnya juga telah dilakukan *Improve K-Means* menggunakan PSO yang terbukti dapat meningkatkan performa dari metode *K-Means* (Tan, 2015)(Rani dan Parthiban, 2015)(Karimov dan Ozbayoglu, 2015).

### K-MEANS CLUSTERING

*K-Means* adalah algoritma yang cukup sederhana dan efektif dalam proses pengelompokan data berdasarkan kemiripannya atau disebut *clustering*. Penentuan kemiripan data dihitung menggunakan rumus *Euclidean Distance* (Anggodo dkk., 2017). *K-Means* termasuk dalam salah satu algoritma yang bersifat *unsupervised* yaitu sebuah algoritma yang tidak membutuhkan proses *training* atau pembelajaran terlebih dahulu. Setiap *cluster* dibentuk berdasarkan titik *centroid* atau titik pusat yang ditentukan. Jumlah titik *centroid* pada *cluster* menggambarkan jumlah *K* atau *cluster* itu sendiri.

Pada algoritma *K-Means* penentuan awal titik *centroid* didapatkan secara random, dan pada iterasi berikutnya titik *centroid* didapatkan dari perhitungan jarak rata-rata dari anggota *cluster* terhadap titik *centroid* awal. Adapun algoritma dari *K-Means* adalah sebagai berikut:

1. Tentukan Nilai *K* terlebih dahulu.  
Nilai *K* adalah jumlah *cluster* yang ingin dibuat.

Dalam penelitian ini digunakan perhitungan metode *Elbow* atau juga disebut dengan *SSE* (*Sum of Square Error*) untuk mengetahui jumlah *cluster* yang optimal, adapun perhitungannya disajikan pada persamaan (1) berikut (Irwanto dkk., 2012).

$$SSE = \sum_{K=1}^K \sum_{x_i \in S_k} \|x_i - C_j\|_2^2 \quad (1)$$

Keterangan =  
 $X_i$  = Data ke-*i*  
 $C_j$  = Titik pusat *cluster* atau (*centroid*)  
*K* = Jumlah *cluster*

Nilai *K* yang paling sesuai adalah nilai *K* yang mengalami penurunan nilai *SSE* secara drastis, dan apabila digambarkan akan membentuk siku atau (*elbow*).

2. *Generate* titik *centroid* awal secara *random*, apabila *K=2* maka diperlukan 2 titik *centroid* awal yang didapatkan secara acak dari dataset.
3. Perhitungan jarak antara data  $x_i$  dengan titik *centroid* dengan rumus *Euclidean Distance* (*D*) pada persamaan (2) berikut.

$$D(X_i, C_j) = \sqrt{\sum_{i=1}^p (X_i - C_j)^2} \quad (2)$$

Keterangan =  
 $X_i$  = data ke-*i*  
 $C_j$  = titik pusat *cluster* atau (*centroid*)  
*P* = dimensi data

4. Menentukan setiap objek data masuk kedalam *cluster* dengan jarak *euclidean* terdekat atau terkecil.
5. Melakukan perhitungan ulang titik *centroid* pada iterasi berikutnya dengan menghitung rata-rata jarak antara semua objek dalam *cluster*.
6. Ulangi langkah ke-3 hingga nilai *centroid* tidak berubah atau telah dalam kondisi berhenti (*stopping condition*).

Untuk mengukur seberapa baik sebuah *cluster* digunakan rumus perhitungan *Silhouette Coefficient* yang merupakan perpaduan antara rumus *separation* dan *cohesion* sebagai berikut (Wahyuni dkk., 2016).

1. Perhitungan jarak rata - rata  $C_j$  yang merupakan titik *centroid* terhadap semua objek  $X_i$  dalam satu *cluster* yang dihitung menggunakan persamaan (3).

$$\alpha(i) = \frac{1}{|A|} \sum_{j \in A, j \neq i} C d(i, j) \quad (3)$$

Keterangan :  
*j* = objek dalam *cluster* *A*  
 $d(i, j)$  = jarak objek *i* ke titik *cluster* *j*

2. Perhitungan jarak rata - rata  $C_j$  terhadap objek pada *cluster* lain menggunakan persamaan (4), lalu diambil nilai terkecil sesuai dengan persamaan (5).

$$d(i, C) = \frac{1}{|A|} \sum_{j \in A, j \neq i} C d(i, j) \quad (4)$$

$$b(i) = \min_{C \neq A} d(i, j) \quad (5)$$

3. Hitung nilai *Silhouette Coefficient* untuk tiap *centroid* menggunakan persamaan (6).

$$s(i) = \frac{b(i) - \alpha(i)}{\max(\alpha(i), b(i))} \quad (6)$$

Maka nilai *Sillhouette Coeficient* merupakan nilai rata-rata  $s(i)$  semua *cluster* yang menggambarkan kemiripan data dalam *cluster* dan menunjukkan seberapa tepat data diletakkan di dalam sebuah *cluster*.

## PARTICLE SWARM OPTIMIZATION

*Particle Swarm Optimization* (PSO) adalah sebuah algoritma yang termasuk dalam kelompok *swarm intelligents* yang merepresentasikan solusi sebagai sebuah partikel atau *swarm*. Partikel secara iteratif bergerak di dalam *search space* yang dihitung berdasarkan perpindahan posisi partikel dan rumus perhitungan *update* kecepatan (*velocity*) untuk menemukan posisi *global best*. Posisi dan kecepatan sebuah partikel  $i$  disetiap iterasi  $t$  dalam *search space* dilambangkan dengan vektor seperti pada persamaan (7) dan (8) berikut (Pornsing, 2014).

$$X_{i(t)} = (X_{i1(t)}, X_{i2(t)}, \dots, X_{in(t)}) \quad (7)$$

$$V_{i(t)} = (V_{i1(t)}, V_{i2(t)}, \dots, V_{in(t)}) \quad (8)$$

Keterangan :

$X_{i(t)}$  = Posisi Partikel  $i$  dalam iterasi  $t$ .

$V_{i(t)}$  = Kecepatan partikel  $i$  dalam iterasi  $t$ .

Setiap partikel akan bergerak menjelajahi *search space*. Pergerakan partikel dipengaruhi oleh kecepatan untuk mendapatkan posisi terbaik. Adapun rumus perhitungan kecepatan (*velocity*) seperti pada persamaan (9) berikut (Pornsing, 2014).

$$V_i^{t+1} = \omega \cdot v_i^t + c_1 \cdot r_1 \cdot (p_i - x_i) + c_2 \cdot r_2 \cdot (pg_i - x_i) \quad (9)$$

Keterangan :

$v_i$  = Nilai kecepatan partikel  $i$  hingga  $n$

$t$  = Waktu iterasi

$\omega$  = Nilai vektor inerti

$p_i$  = Posisi terbaik untuk tiap partikel

$pg_i$  = Posisi terbaik untuk semua partikel

$c_1, c_2$  = Konstanta *cognitive* dan *sosial*

$r_1, r_2$  = Bilangan acak antara  $[0, 1]$

Setelah menghitung kecepatan partikel, maka selanjutnya dihitung perpindahan posisi partikel menggunakan rumus *update* posisi menggunakan persamaan (10) sebagai berikut.

$$x_i^{t+1} = x_i + V_i^{t+1} \quad (10)$$

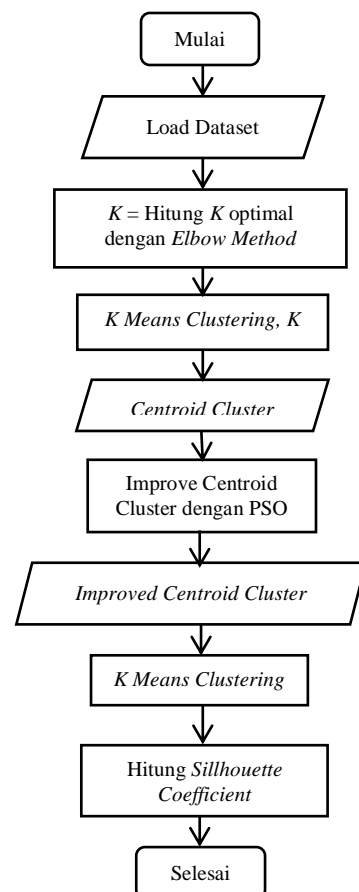
Dalam penelitian ini partikel dalam PSO merepresentasikan titik *centroid* untuk perhitungan *K-Means*. Adapun algoritma *Improved K-Means* dengan PSO adalah sebagai berikut (Wahyuni dkk., 2016).

- i. Inisialisasi partikel yang berisi *centroid* setiap *cluster* dan kecepatan ( $v$ ) partikel.
- ii. Menentukan partikel best dan *global best* awal.

- iii. Mulai iterasi dan lakukan pada setiap vektor.
  - (i) Perhitungan *K-Means clustering* menggunakan titik *centroid* yang telah di *generate*.
  - (ii) Menentukan *cluster* setiap objek dengan *K-Means*.
  - (iii) Hitung nilai *Sillhouette Coefficient*.
  - (iv) Update nilai titik *centroid* menggunakan persamaan *update* kecepatan pada persamaan (2) dan update partikel dengan persamaan (3).
  - (v) Update *particle best* dan *global best*.
- iv. Ulangi langkah ke-3 hingga iterasi berhenti setelah mencapai *stopping condition* atau iterasi maksimum.

## 2. METODOLOGI PENELITIAN

Dalam penelitian ini diusulkan metode *improved K-Means* menggunakan *Particle Swarm Optimization* (PSO). PSO digunakan untuk *generate* titik *centroid* perhitungan *K-Means* sehingga didapatkan nilai titik *centroid* optimum. *Objective function* yang digunakan sebagai *fitness* dari PSO merupakan nilai *Silhouette Coefficient* yang digunakan untuk mengukur seberapa baik sebuah *cluster*. Adapun metodologi yang digunakan dalam penelitian ini disajikan pada Gambar 1.



Gambar 1. Metodologi Penelitian

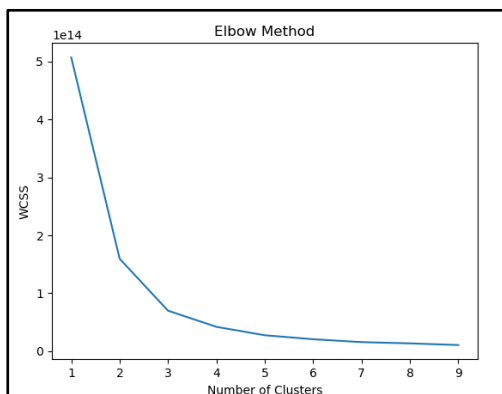
## 2.1 Dataset

*Dataset* yang digunakan adalah data pembayaran tagihan kartu kredit di sebuah bank yang didapatkan dari <https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients> yang memiliki 30000 *instances* dan 24 *features*. Adapun 24 *features* tersebut adalah sebagai berikut :

1. X1 : Jumlah *limit* kartu kredit yang diberikan kepada nasabah
2. X2 : *Gender* (1=Laki-Laki, 2=Perempuan)
3. X3 : Pendidikan (1= *graduate school*, 2 = *university*, 3= *high school*, 4= *others*)
4. X4 : Status (1=Menikah, 2= Lajang, 3=Others)
5. X5 : Usia (tahun)
6. X6-X11: Catatan pembayaran nasabah mulai bulan April hingga September.  
X6 = pembayaran bulan September, X7 = Pembayaran bulan Agustus, dan seterusnya hingga X11 = pembayaran bulan April. Ukuran untuk status pembayaran adalah :  
-1 = tidak ada  
1 = penundaan pembayaran dalam 1 bulan  
2 = penundaan pembayaran dalam 2 bulan  
8 = penundaan pembayaran dalam 8 bulan
7. X12-X17 : Jumlah tagihan yang harus dibayar  
X12 = jumlah tagihan bulan september, hingga X17 = jumlah tagihan bulan April.
8. X18 – X23 : Jumlah tagihan yang telah dibayar  
X18 = jumlah tagihan yang telah dibayar pada bulan september.  
X23 = jumlah tagihan yang telah dibayar pada bulan april.
9. X24 = *default* pembayaran bulan depan (1 = Ya, 2 = Tidak).

## 2.2 Elbow Method

Untuk mencari Nilai  $K$  yaitu jumlah *cluster* yang optimal, pada penelitian ini digunakan perhitungan dengan metode *elbow*, dan menghasilkan nilai *SSE* yang disajikan pada Gambar 2.



Gambar 2. Hasil perhitungan metode *elbow*

Nilai  $K$  optimal adalah nilai  $K$  yang mengalami penurunan nilai *SSE* secara drastis, dapat dilihat

pada gambar 2 nilai  $K$  optimal adalah 2. Maka dapat disimpulkan bahwa *clustering* akan dilakukan dengan jumlah 2 *cluster*.

## 2.3 K-Means Clustering

Pada tahap ini dilakukan proses *clustering* menggunakan metode *K-Means* murni dengan jumlah  $K=2$ , pada penelitian ini dilakukan *clustering* sehingga didapatkan data yang telah berkelompok menjadi 2 *cluster* berdasarkan titik *centroid* yang didapatkan secara random. Pada perhitungan *K-Means* murni, juga akan dihitung nilai *sillhouette coefficient* untuk digunakan sebagai perbandingan untuk pengujian.

## 2.4 Particle Swarm Optimization

PSO akan melakukan *improve* pada titik *centroid* hingga didapatkan nilai titik *centroid* yang optimal dengan parameter sebagai berikut :

$$C1 = 1.2$$

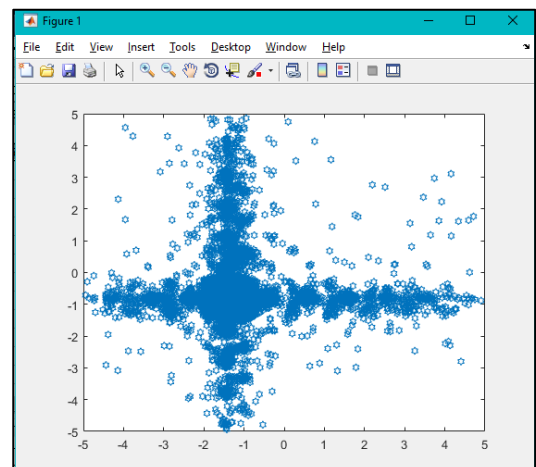
$$C2 = 1.2$$

Nilai  $\omega$  akan diuji coba mulai rentang 0,1 hingga 1 untuk menemukan parameter  $\omega$  yang optimal. Untuk jumlah partikel juga akan diuji coba dengan jumlah partikel 50, 100, dan 200. Untuk percobaan pertama akan dilakukan PSO untuk optimasi titik *centroid* dengan jumlah partikel yaitu 50 dan nilai  $\omega$  adalah 0.9 sehingga didapatkan nilai *global best position* untuk titik *centroid* disajikan pada Tabel 1.

Tabel 1. Representasi Partikel

Cluster	Global best position
1	-1.425128
2	-0.800321

*Global best position* merupakan representasi dari nilai titik *centroid* yang optimal yang diukur dengan nilai *sillhouette coefficient*. Dengan menggunakan software Matlab 2018, visualisasi data pencarian titik *centroid* optimal dapat digambarkan pada Gambar 3.



Gambar 3. Visualisasi data

Setelah mendapatkan titik centroid optimal maka akan kembali dilakukan proses perhitungan *K-Means*, sehingga akan didapatkan data yang telah berkelompok secara optimal.

### 3. PENGUJIAN DAN HASIL

Pada bagian ini akan dilakukan 2 skenario pengujian. Adapun di antaranya adalah pengujian parameter PSO yaitu  $\omega$  dan jumlah partikel, dan pengujian pengaruh jumlah data terhadap nilai *sillhouette coefficient*.

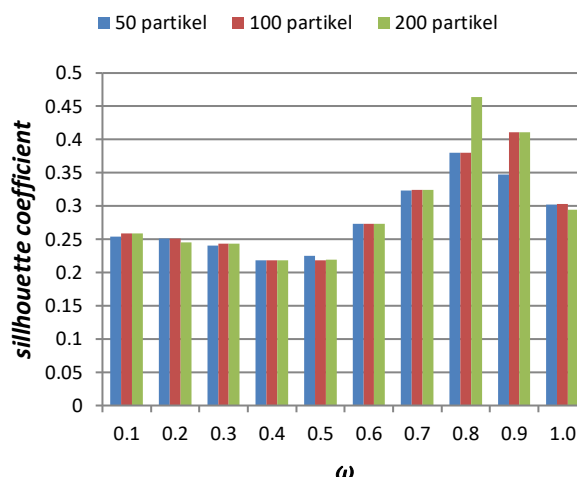
#### 3.1 Pengujian Parameter PSO

Pada pengujian parameter PSO akan dilakukan uji coba menggunakan parameter  $\omega$  dan jumlah partikel yang digunakan dalam PSO. Nilai  $\omega$  akan diuji coba mulai rentang 0,1 hingga 1 untuk menemukan parameter  $\omega$  yang optimal. Untuk jumlah partikel juga akan diuji coba dengan jumlah partikel 50, 100, dan 200. Jumlah data yang akan digunakan adalah 10 ribu data. *Objective function* untuk parameter optimal adalah nilai *Sillhouette Coefficient*.

Tabel 2. Pengujian parameter PSO

$\omega$	50 partikel	100 partikel	200 partikel
0.1	0,254321	0,258763	0,258816
0.2	0,251498	0,251242	0,244896
0.3	0,240041	0,243586	0,243411
0.4	0,218405	0,218310	0,218265
0.5	0,225221	0,218805	0,218924
0.6	0,273313	0,273613	0,273630
0.7	0,323769	0,323825	0,323946
0.8	0,379909	0,380440	0,463787
0.9	0,347294	0,410731	0,411159
1	0,3016896	0,303527	0,293984

Pada Tabel 2 disajikan perbandingan nilai *sillhouette coefficient* menggunakan parameter  $\omega$  dan jumlah partikel yang berbeda. Pada Tabel 3. dapat disimpulkan bahwa nilai *sillhouette coefficient* terbaik adalah sebesar 0.463787 menggunakan 10 ribu data dengan parameter PSO yaitu nilai inerti atau  $\omega$  yaitu 0.8 dan dengan jumlah partikel yaitu 200. Plot dalam grafik disajikan pada Gambar 4.



Gambar 4. Visualisasi pengujian parameter pso

Pada Gambar 4. dapat dilihat hasil nilai *sillhouette coefficient* dengan parameter  $\omega$  dan jumlah partikel tertentu.

#### 3.2 Pengujian terhadap jumlah data

Pengujian dilakukan dengan membandingkan hasil *sillhouette coefficient* untuk *clustering* menggunakan *K-Means* murni dan *Improved K-Means* dengan PSO dengan data yang diambil secara acak dengan jumlah data 10 ribu, 15 ribu, 20 ribu, 25 ribu, dan 30 ribu untuk membandingkan pengaruh jumlah data terhadap nilai *sillhouette coefficient* yang disajikan pada Tabel 3.

Tabel 3. Perbandingan dari nilai *sillhouette coefficient* terhadap jumlah data

Jumlah data	<i>K-Means</i>	<i>Improved K-Means dengan PSO</i>
10 ribu	0,3230	0,4637
15 ribu	0,3106	0,3700
20 ribu	0,3250	0,3176
25 ribu	0,3097	0,4461
30 ribu	0,3124	0,3768
<b>Average</b>	<b>0,31614</b>	<b>0,39484</b>

Dari hasil pengujian menunjukkan bahwa *improved K-Means* dengan PSO memiliki nilai *sillhouette coefficient* yang lebih baik dibandingkan dengan *K-Means* murni. Parameter PSO yang digunakan yaitu nilai  $\omega$  adalah 0.8 dan jumlah partikel adalah 200. Dari Tabel 3. tersebut dapat disimpulkan bahwa jumlah data berpengaruh pada nilai *sillhouette coefficient*, semakin besar jumlah data maka akan semakin rendah nilai *sillhouette*

*coefficient*. Jumlah data yang memiliki nilai *sillhouette coefficient* terbaik adalah 10 ribu data.

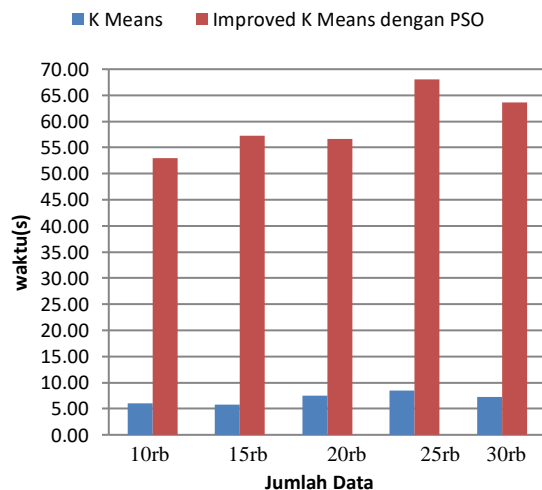
Berdasarkan pengujian parameter PSO yang telah dilakukan untuk mendapatkan parameter optimum, maka juga didapatkan hasil dari pengujian terhadap waktu komputasi sebagai berikut.

**Tabel 4.** Hasil perbandingan waktu komputasi

Metode	Rata-rata waktu komputasi (s)
K-Means	7,009
<i>Improved K-Means</i> dengan PSO	59,690

Pada Tabel 4 dapat diambil kesimpulan bahwa metode *Improved K-Means* menggunakan PSO membutuhkan waktu komputasi yang lebih lama dibandingkan dengan *K-Means* murni. Hal tersebut terjadi karena adanya proses pencarian titik *centroid* optimal oleh PSO yang sudah pasti akan membutuhkan waktu dibandingkan dengan proses random pencarian titik *centroid* dari *K-Means* murni.

Adapun visualisasi dari perbandingan waktu komputasi antara *K-Means* murni dan *Improved K-Means* dengan PSO disajikan pada Gambar 5.



**Gambar 5.** Visualisasi perbandingan waktu komputasi

#### 4. KESIMPULAN

Berdasarkan dari hasil penelitian ini dapat disimpulkan yaitu metode *Improved K-Means* menggunakan algoritma *swarm intelligents* yaitu *Particle Swarm Optimization* telah berhasil melakukan *clustering* nasabah pemegang kartu kredit berdasarkan pembayaran *credit card bill*. Pengujian dilakukan dengan membandingkan nilai *sillhouette coeffisien* terhadap metode *K-Means* murni dan *Improved K-Means* dengan PSO yang

membuktikan bahwa *Improved K-Means* dengan PSO dapat memberikan model *cluster* yang lebih baik yaitu sebesar 0,39484 dibandingkan *K-Means* murni yang sebesar 0,31614. Pada penelitian selanjutnya untuk meningkatkan hasil yang lebih baik dapat dilakukan *Improved* pada parameter PSO dengan membuat nilai  $\omega$  atau *inertia* menjadi dinamis pada setiap iterasinya, sehingga diharapkan akan menghasilkan *objective function* yang lebih baik.

#### DAFTAR PUSTAKA

- ADOLFSSON, A., ACKERMAN, M., BROWNSTEIN, N.C., 2018, *To cluster, or not to cluster : an analysis of clusterability methods*, *Pattern Recognition*, doi: <https://doi.org/10.1016/j.patcog.2018.10.026>
- ANGGODO, Y.P., CAHYANINGRUM, W., FAUZIYAH, A.N., CHOLISSODIN, I., 2017, *Hybrid K-Means dan Particle Swarm Optimization* untuk *clustering* nasabah kredit, *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)* Vol. 4, No. 2, Juni 2017, hlm. 104-110.
- BHATI, P. & SHARMA, M., 2015, *Credit Card Number Fraud Detection Using K-Means with Hidden Markov Method*, *SSRG International Journal of Mobile Computing & Application (SSRG-IJMCA) – volume 2 Issue 3 May to June 2015*.
- CARNEIRO, N., FIGUEIRA G., COSTA M., 2017, *A Data Mining based system for credit card fraud detection in e-tai*, *Decision Support System* doi:10.1016/j.dss.2017.01.00.
- CHOUGULE, THAKARE A.D., KALE P., GOLE M., NANEKAR P., 2015, *Genetic K-Means Algorithm for Credit Card Fraud Detection*, *International Journal of Computer Science and Information Technologies (IJCSIT)*, Vol. 6 (2), 2015, 1724-1727.
- DEWRI L. V., ISLAM M.R., SAHA N.K., 2016, *Behavioral analysis of credit card users in a developing country : A case of Bangladesh*, *International Journal of Business and Management*; Vol. 11, No. 4; 2016.
- HAN, P. Y & CHAI J., 2012, *The Application of K-Means in personal credit analysis*, *Trans Tech Publications, Switzerland*.
- IRWANTO, PURWANANTO, Y., SOELAIMAN R., 2012, *Optimasi Kinerja Algoritma Klasterisasi K-Means untuk Kuantisasi Warna Citra*, *Jurnal Teknik ITS* Vol. 1, No. 1 (Sept. 2012) ISSN: 2301-9271.
- KARIMOV J. & OZBAYOGLU M., 2015, *Clustering Quality Improvement of K-Means using a Hybrid Evolutionary Model*, *Procedia Computer Science* 61 (2015) 38 – 45 .



- KUMARI, S., CHOUBEY, ABHA., 2017, *Credit Card Fraud Detection Using HMM and K-Means Clustering Algorithm*, *International Journal of Scientific Research Engineering & Technology (IJSRET)*.
- LESTARI, B.A., SUHARJO, B., MUFLIKHATI, I., 2017, Minat Kepemilikan Kartu Kredit (Studi Kasus Kota Bogor), *Jurnal Aplikasi Bisnis dan Manajemen*, Vol. 3 No. 1, Januari 2017.
- OKTANISA, I., & SUPIANTO, A.A., 2018, Perbandingan teknik klasifikasi dalam Data Mining untuk Bank Direct Marketing, *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)* Vol. 5, No. 5, Oktober 2018, hlm. 567-576.
- PARAMARTHA, G. N. W., 2017, Analisis Perbandingan Metode K-Means dengan Improved Semi Supervised K-Means pada Data Indeks Pembangunan Manusia (IPM), *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)* Vol. 1, No. 9, Juni 2017, hlm. 813-824.
- PORNSING, 2014, *A Particle Swarm Optimization for the vehicle routing problem*. *Open Dissertation University of Rhode Island*.
- RANI, A. J. M. & PARTHIBAN L., 2015, *Improved Particle Swarm Optimization And K-means Clustering Algorithm For News Article*, *Chennai Fourth International Conference on Sustainable Energy and Intelligent Systems*.
- TAN L., 2015, *A Clustering K-Means Algorithm Based on Improved PSO Algorithm*, *2015 Fifth International Conference on Communication Systems and Network Technologies*.
- VAISHALI, 2014, *Fraud Detection in credit card clustering approach*, *International Journal of Computer Applications (0975 – 8887)* Volume 98– No.3, July 2014.
- WAHYUNI, I., AULIYA, Y.A., RAHMI, A. MAHMUDY, W. F., 2016. *Clustering nasabah bank berdasarkan tingkat likuiditas menggunakan hybrid particle swarm optimization*. *Jurnal Ilmiah Teknologi dan Informasi ASIA (JITIKA)*, vol. 10, no. 2, pp. 24-33.
- VIIJ, D. & BANU S. K.Z., 2018, *An Improved Credit Card Fraud Detection Using K-Means Clustering Algorithm*, *International Journal of Engineering Science Invention (IJESI)*.
- YEH I.C. & LIEN C.H., 2009, *The comparison of data mining techniques for the predictive accuracy of probability of default credit card clients*, *Expert System with Applications*.