

## PENINGKATAN KINERJA ALGORITMA K MEANS DENGAN MENGGUNAKAN PARTICLE SWARM OPTIMIZATION DALAM PENGELOMPOKAN DATA PENYEDIAAN AKSES SANITASI DAN AIR BERSIH

Ari Yunus Hendrawan M. Kom  
[aritubil@gmail.com](mailto:aritubil@gmail.com)

### Abstrak

Air merupakan salah satu hal yang memegang peranan yang sangat penting dalam kelangsungan hidup manusia, karena pemerintah Indonesia mempunyai program penyediaan air dan sanitasi berbasis masyarakat (PAMSIMAS), sehingga semua program tersebut dapat berjalan dengan baik memerlukan suatu teknik pengelompokan status wilayah dalam penelitian ini. Dengan algoritma K-means. K-means merupakan algoritma partisi yang bertujuan untuk membagi data menjadi jumlah cluster yang ditentukan, hasil dari algoritma K means tergantung pada pemilihan pusat klater awal namun permasalahan yang sering terjadi saat pemilihan centroid awal diambil secara acak dari solusinya. dari pengelompokan kurang tepat. Untuk mengatasi masalah tersebut penulis ingin menggunakan algoritma PSO in initial centroid selector untuk algoritma K-means, dalam penelitian ini juga membandingkan pemilihan 3 centroid pertama secara random, kedua menurut standar pemerintah nilainya tinggi, kualitas air minum sedang dan rendah maka metode ketiga yang diusulkan dengan algoritma PSO kemudian diuji dengan Davies Bouldin Index. Dari hasil pengujian, metode K-means dengan pemilihan centroid awal secara acak dengan nilai 0,208856082, metode K-means pemilihan centroid sesuai dengan standar pemerintah tentang kondisi SAM 0,280077 dan metode pemilihan terbaik adalah K -berarti PSO 0, 08383. Jadi pengujian data PAMSIMAS menggunakan K-means PSO menemukan bahwa metode tersebut lebih optimal.

**Kata kunci:** Algoritma K – means, Algoritma PSO, Data Pamsimas

### Abstract

Water is one of the things that plays a very important role in human survival, because the Indonesian government has a community-based water supply and sanitation (PAMSIMAS) program, so that all the programs run well need a regional status grouping technique in this research. with the K-means algorithm. K-means is a partition algorithm that aims to divide the data into the specified number of clusters, the results of the K means algorithm depend on the selection of the initial klater center but problems that often occur when selecting the initial centroid are randomly drawn from the solution. from the grouping is not quite right. To overcome this problem the author wants to use the PSO algorithm in the initial centroid selector for the K-means algorithm, in this study also compared the selection of the first 3 centroids according to random, second according to government standards the value of high, medium and low drinking water quality then the third method proposed by the PSO algorithm was then tested with Davies Bouldin Index. From the test results, the K-means method with the selection of random initial centroid with a value of 0.208856082, the K-means method with the selection of centroids in accordance with government standards about SAM conditions of 0.280077 and the best selection method is K-means PSO 0, 08383. So testing the PAMSIMAS data using K-means PSO found that the method was more optimal.

**Keywords:** Algoritma K – means, Algoritma PSO, Data Pamsimas

## 1. PENDAHULUAN

### 1.1 Latar Belakang

Air menjadi peran utama bagi kelancaran hidup manusia, dimana air dapat digunakan untuk kelancaran hidup masyarakat yang kegunaannya meliputi air minum, kesehatan sanitasi serta masih banyak kegunaan yang lain, sehingga air menjadi salah satu komponen utama yang paling dasar dalam kehidupan masyarakat. Oleh karena prasarana sanitasi dan air bersih menjadi bagian utama untuk melaksanakan tujuan dari Millennium Development Goals yang disingkat dengan MDGs pada tahun 2015 dengan tujuan berkurangnya jumlah masyarakat yang belum mendapatkan akses untuk sanitasi air bersih[1]. Perencanaan pembangunan sanitasi dirasa penting dalam pembangunan berkelanjutan dan telah menjadi tantangan berat, untuk mengatasi tantangan tersebut memerlukan solusi yang tepat dalam pembangunan sarana dan prasarana sanitasi, dengan adanya agenda MDGs diharapkan dapat mewujudkan tujuan dari pembangunan yaitu untuk kesejahteraan masyarakat. Dalam MDGs perencanaan penyediaan akses sarana dan prasarana sanitasi memiliki 17 tujuan, tujuan tersebut salah satunya mewujudkan dan memastikan tersedianya dalam pengolahan air bersih serta sanitasi yang mencapai 100%, yang terdapat pada tujuan no. 6 MDGs, dalam upaya mencapai tujuan tersebut Indonesia termasuk kelompok negara yang melakukan pembangunan sesuai capaian MDGs, upaya tersebut dapat dilihat pada tahun 2014 yang mencapai 59,71% Indonesia telah mencapai capaian dalam pembangunan sanitasi dan air bersih, dan melakukan peningkatan dalam pembangunan pada tahun 2015 dengan nilai capaian target 62,41% [2]. Dari tujuan SDGs tersebut pemerintah telah membuat suatu kebijakan nasional tentang Penyediaan air minum dan sanitasi berbasis masyarakat yang disingkat dengan PAMSIMAS, salah satu programnya meliputi pelarangan membuang air besar di tempat terbuka, menggunakan sabun untuk mencuci tangan, pemanfaatan sampah dan limbah, sehingga setiap daerah provinsi maupun kabupaten wajib memperbaiki dan melakukan pembangunan di bidang sarana prasarana sanitasi dengan acuan pedoman STBM.[3] Untuk mendukung program nasional PAMSIMAS telah dilaksanakan dalam kurun waktu 8 tahun, pada tahap ke-2 tahun 2013 sampai tahun 2016, mendapatkan hasil yang cukup baik dimana awal mula yang menjalankan program PAMSIMAS sejumlah 218 kabupaten, 5.000 jumlah desa, menjadi 12.000 desa, dengan pelaksanaan sekitar 12.000.000 jiwa[4]. Dari hasil evaluasi pembangunan tahap ke-2 PAMSIMAS dapat di rasakan dan berjalan dengan baik di lingkungan masyarakat, oleh karena itu agar dapat memanfaatkan sarana dan prasarana air dan sanitasi yang layak sepanjang masa dalam pengelolaannya yang berkelanjutan dan agar Semua instalasi sesuai yang diharapkan yaitu berjalan dengan baik dan bersifat kekal, perlu adanya teknik pengelompokan status daerah yang akan di bangun sarana prasarana sanitasi dan air bersih guna untuk pembangunan yang merata. Untuk dapat mengelompokkan data bidang sistem keberlanjutan PAMSIMAS perlu di cari suatu metode yang akurat, dalam thesis ini akan digunakan beberapa metode untuk keperluan tersebut di atas dengan menggunakan data pada modul 7.3 kinerja badan pengelola status PAMSIMAS dari kurun waktu 2008-2017 periode laporan juni tahun 2018, dengan menggunakan teknik data mining penulis ingin mengusulkan cara pengelompokan data yang dapat mudah diimplementasikan dengan cepat, tepat dan efisien, data PAMSIMAS yang di gunakan untuk pengelompokan kriteria pembangunan sarana dan prasarana sanitasi sehingga tepat guna untuk dapat melakukan pembangunan berkelanjutan sesuai aturan, teknik data mining yang ingin diimplementasikan adalah metode klastering yang nantinya diharapkan akan mendapatkan sebuah klaster yang optimal pada data PAMSIMAS. Klastering merupakan metode yang dapat digunakan untuk mengatasi sebuah permasalahan dengan cara pengelompokan data, sehingga akan mendapatkan sebuah partisi dalam subset yang akan menghasilkan sebuah informasi sesuai kebutuhan dan tujuan.

Tujuan dari klastering adalah untuk dapat mengelompokkan data kedalam beberapa kelompok data atau disebut juga klaster dan diharapkan memiliki tingkat kemiripan yang baik/maksimal dalam satu kelompok dan memiliki tingkat kemiripan yang minimum antar kelompok[5]. Metode klastering yang sering digunakan oleh para peneliti adalah metode partisi dan metode hirarki, metode klastering dalam penelitian ini ingin menggunakan metode klastering berbasis partisi. Dalam metode klastering berbasis partisi Algoritma yang sering digunakan adalah Algoritma K-means dan pada algoritma K-means sangat bergantung pada data dan teknik clustering yang digunakan [6]. K-means merupakan algoritma dari teknik klastering berbasis partisi yang bertujuan membagi data kedalam K buah klaster yang sudah ditentukan sebelumnya. Permasalahan yang sering terjadi pada klastering berbasis partisi terletak pada proses pemilihan centroid awal yang masih dilakukan secara acak [7]. Hasil dari pengelompokan Algoritma K-means tergantung pada pemilihan pusat awal klaster yang ditentukan secara random jika penentuan pusat awal klaster jauh dari solusi akhir maka kemungkinan hasil dari pengelompokan tersebut menjadi kurang tepat, hal ini menyebabkan kualitas hasil klastering pada teknik partisi tidak selalu menghasilkan klaster dengan kualitas terbaik, karena hasil dari teknik partisi sangat bergantung pada ketepatan terhadap pemilihan centroid awal, untuk mengatasi masalah tersebut penulis ingin menggunakan algoritma optimasi karena algoritma optimasi pada umumnya memiliki manfaat untuk meningkatkan atau mengoptimalkan hasil dari metode, algoritma PSO yang akan digunakan untuk dapat memperbaiki kelemahan algoritma klastering berbasis partisi dalam pemilihan centroid awal [8].

Dalam penelitian ini juga dibandingkan pemilihan centroid terbaik dengan tiga cara yaitu yang pertama adalah secara acak sesuai dengan K- mean yang asli, yang ketiga pemilihan centroidnya berdasarkan nilai kondisi Sanitasi air minum (SAM) daerah yang memiliki nilai SAM tertinggi, medium dan selanjutnya daerah yang memiliki nilai terendah, dan yang ketiga adalah pemilihan centroid nya menggunakan algoritma PSO, dari hasil pengujian didapat penggabungan K- means dan algoritma PSO lah yang terbaik. Oleh karena itu untuk menyelesaikan persoalan tersebut peneliti ingin menerapkan penggabungan algoritma K means dan PSO untuk mencari centroid awal terbaik sehingga menghasilkan cluster yang optimal pada algoritma K means. Dari algoritma K-means PSO tersebut. Kemudian ketiga cara pemilihan tersebut akan diuji performanya menggunakan metode pengujian evaluasi untuk mencari nilai centroid yang paling minimum menggunakan Davies Bouldin Indeks (DBI)[9] pengujian dengan DBI menggambarkan bahwa semakin kecil nilai dari DBI maka semakin dekat antara klaster satu dengan yang lain, dalam penelitian ini algoritma K- means akan ditingkatkan kinerjanya dengan PSO dan peneliti ingin memberikan judul yaitu “Peningkatan Kinerja Algoritma K Means Dengan Menggunakan Particle Swarm Optimization Dalam Pengelompokan Data Penyediaan Akses Sanitasi Dan Air Bersih Dan Air Bersih”.

## 1.2 Identifikasi Masalah

Berdasarkan latar belakang tersebut K-Means merupakan algoritma yang sering digunakan, namun memiliki kelemahan pada penentuan pusat awal klaster dimana pada algoritma murni K-Means ditentukan secara acak. Sehingga mungkin menyebabkan kinerja algoritma K-Means menjadi menurun dan mudah terjebak dalam lokal optimal.

## 1.3 Perumusan Masalah

Bagaimana cara melakukan peningkatan kualitas, kinerja K-Means algoritma K-Means dengan menggunakan algoritma PSO (Particle Swarm Optimization) untuk mengatasi masalah penentuan pusat klaster awal pada algoritma K-Means?

#### 1.4 Tujuan Penelitian

Dari rumusan masalah penelitian yang ada, penulis memiliki beberapa tujuan yang ingin dicapai, adapun tujuan dari penelitian meliputi: Meningkatkan kinerja algoritma K-Means dengan menggunakan algoritma PSO untuk dapat mengatasi masalah pemilihan centroid awal pada algoritma K-Means.

#### 1.5 Manfaat Penelitian

Dari tujuan penelitian yang ingin dicapai, terdapat beberapa manfaat penelitian yang diperoleh. Adapun manfaat sebagai berikut:

1. Dari hasil pengujian data PAMSIMAS penulis akan mendapatkan 3 kluster daerah PAMSIMAS berdasarkan perhitungan jarak optimal untuk memberi masukan kepada Pemerintah sebagai pengambil kebijakan nasional
2. Sebagai bahan referensi bagi akademik dan adik kelas.

## 2. Tinjauan Pustaka

### 2.1 Penelitian Tentang K-Means

Pada tahun 2016 Fina Nasari dkk, melakukan penelitian menggunakan Algoritma K-Means tentang pengelompokan daerah yang terkena penyakit diare dan dari hasil penelitian tersebut dapat diketahui daerah-daerah mana saja yang bisa dikelompokkan terkena penyakit diare [11]. Pada tahun 2016 Harianto Kristanto dkk, melakukan penelitian menerapkan metode k-means clustering pada Sistem akan pengelompokkan rasio profitabilitas semua perusahaan pada sektor pertambangan dan sektor industri barang konsumsi dalam 1 tahun hasil dari pengelompokan tersebut Setelah dilakukan uji coba terhadap jumlah centroidnya, dapat ditarik kesimpulan bahwa semakin banyak jumlah centroid dalam setiap proses clustering, maka makin spesifik kelompok cluster yang dihasilkan. Dengan demikian pengambilan kesimpulan kesamaan dalam kelompok cluster makin mudah [12]. Pada tahun 2016 Yoga Religia melakukan penelitian tentang “Pengelompokan status desa menggunakan Algoritma K-means” Hasil pengujian menunjukkan bahwa penentuan jarak yang paling optimum dan baik menggunakan Chebyshev dan yang efisien pada waktu menggunakan Euclidean [13].

Pada tahun 2015 Windha Mega Pradnya Duhita melakukan penelitian pengklasteran status gizi balita menggunakan metode K-Means dan dihasilkan bahwa algoritma K-Means hanya memiliki akurasi 34% tepat [14]. Dwi Anggraeni Kuntjoro, Budi Darma Setiawan, Rizal Setya Perdana melakukan penelitian tentang “Algoritme Genetika Untuk Optimasi K-Means Clustering Dalam Pengelompokkan Data Tsunami” dengan tujuan Mengimplementasikan optimasi Algoritme Genetika pada metode K-Means untuk mengelompokkan data tsunami. Dan dari penelitian tersebut dapat disimpulkan bahwa pada cluster 1 tsunami biasanya terjadi pada kedalaman 17,28 kilometer dengan ketinggian air 66,6 meter diatas permukaan laut dan terjadi pada kisaran wilayah 48,20 lintang utara 148,37 bujur timur. Begitu juga pada cluster-cluster selanjutnya[15]. Persamaan matematika dinomori dengan angka Arab di dalam tanda kurung buka-tutup pada posisi rata kanan kolom. Persamaan ditulis menjorok ke dalam sejauh satu 7,5 mm. Untuk persamaan yang tidak cukup ditulis dalam lebar 1 kolom, penulisannya dapat melintasi 2 kolom, ditulis di bagian bawah halaman dan diberi nomor urut yang sesuai.

## 2.2 Penelitian tentang PSO

Particle Swarm Optimization salah satu algoritma optimasi yang sering digunakan. Adapun cara kerja PSO yaitu mengoptimasi permasalahan dengan cara membangkitkan partikel didalam salah satu ruang menggunakan fungsi tertentu, salah satunya fungsi fitness. particle swarm optimization (PSO) sering digunakan untuk memberikan solusi dari masalah optimasi. Cara kerja PSO yang gampang sehingga dapat mengefesienkan perhitungan baik dari segi memori yang di butuhkan dan juga dari segi kecepatan. PSO, juga dapat digunakan untuk menentukan parameter optimum.[16] Pada tahun 2018 Yuni Eka Achyani, dalam penelitian digunakan seleksi atribut pada SVM dengan PSO dan menghasilkan nilai akurasi sebesar 88,71 %, nilai precision 89,47% dan nilai AUC sebesar 0,896. Kemudian menggunakan PSO Hasil menunjukkan nilai akurasi yang lebih tinggi yaitu sebesar 89,38%, nilai precision 89,89% dan nilai AUC sebesar 0,909 dengan menunjukkan nilai akurasi klasifikasi excellent dan terjadi peningkatan akurasi sebesar 0,67 %, dan peningkatan AUC sebesar 0,013 [17]. Pada tahun 2009 Evi Ria Zerda dan Suyanto, melakukan penelitian tentang penggunaan PSO, pada penjadwalan sumber daya proyek untuk melakukan optimasi, hasil penelitian menunjukkan bahwa algoritma PSO dapat digunakan pada masalah optimasi perataan penggunaan sumber daya.. dari uji coba yang dilakukan memberika hasil pemetaan yang bagus dan fluktuasi yang minimal[18]. Achmad Yasid melakukan penelitian tentang Implementasi Automatic Clustering Menggunakan PSO Dan GA Pada Data Kemahasiswaan dengan masalah Menganalisis data calon mahasiswa perguruan tinggi untuk mendapatkan input mahasiswa yang berkualitas sesuai dengan kebijakan yang ada. Tujuan penelitian itu sendiri adalah untuk memperoleh jumlah cluster akhir dan membagi data sesuai karakteristik setiap objek. Dan hasil dari penelitian tersebut adalah , jumlah cluster akhir yang diperoleh sebanyak 7 cluster dengan nilai VI indek terkecil 0.1788 [19].

Achmad saiful dan Joko lianto buliali melakukan penelitian dengan menerapkan PSO Pada K-Means Untuk Clustering Data Automatic Dependent Surveillance Broadcastde dengan tujuan penelitian mengelompokkan rute penerbangan pada data Atomatic Dependent Surveillance-Broadcast menggunakan metode clustering untuk mendapatkan similaritas rute penerbangan. Dan berdasarkan uji coba dari analisis Berdasarkan pada uji coba dari analisis yang dilakukan, optimasi pada k-means menggunakan particle swarm optimization dengan fitness function silhouette coefficient berhasil dilakukan. Hal tersebut dapat dilihat pada penegompokkan pada automatic dependent surveillance-broadcast (ADS-B), metode tersebut menjadi kelompok metode terbaik pada lima dari enam segmen yang ada serta menghasilkan nilai davies-boulding index lebih baik pada satu segmen dibandingkan dengan metode k-means, k-medoids dan fuzzy c-means sebesar 0,779[20].

Max Teja Ajie Cipta Widiyanto melakukan penelitian dengan judul optimasi pso untuk metode clustering fuzzy c-means dalam pengelompokkan kelas, dengan masalah penelitian untuk meningkatkan pengelompokkan FCM dengan Particle Swarm Optimization dan tujuan penelitian nya adalah Dapat merealisasikan suatu metode yang dapat menggabungkan keunggulan dari metode pengklusteran Fuzzy C-Means (FCM) selama ini yaitu terjebak dalam titik local optimal, dengan metode yang diusulkan Penggabungan Fuzzy C-Means dengan revolutionary algorithm semacam Algoritma Genetika, Simulated Annealing (SA), Ant Colony Optimization ACO) dan PSO untuk mendapatkan hasil terbaik, hasil dari penelitian tersebut adalah Penelitian ini menggunakan variable nilai akademik dengan metode clustering FCM yang dioptimasi dengan PSO menunjukkan peningkatan nilai yang lebih bagus dalam semua index validitas cluster serta meminimalkan iterasi, sehingga optimasi ini telah terbukti berhasil membentuk kelas cluster yang lebih kompak dari metode FCM. Lama prosesnya masih cepat pada metode FCM tanpa optimasi PSO seperti penelitian yang terdahulu [21].

### 2.3 Penelitian yang Relevan

Hasan pada tahun 2009 [22] mengusulkan metode Robust Initialization menggunakan menggunakan pendekatan local outlier factor (LOF) yang diusulkan oleh Breuning et al untuk menghindari titik outlier untuk terpilih menjadi pusat klaster. Pertama pusat klaster dipilih secara acak, selanjutnya hitung jarak tiap titik data ke pusat klaster, urutkan jarak dari terkecil ke terbesar. Kemudian titik data yang mempunyai nilai local outlier factor (LOF) mendekati satu dipilih sebagai pusat klaster selanjutnya [23]. Khan dan Ahmad pada tahun 2004 mengusulkan Cluster Center Initialization Algorithm (CCIA) untuk memecahkan masalah inisialisasi pusat awal klaster pada K-means. CCIA didasarkan pada dua pengamatan, dimana beberapa pola mempunyai kemiripan yang hampir sama satu dengan yang lain. Metode ini dimulai dengan menghitung nilai rata-rata dan standar deviasi untuk semua atribut data. Kemudian memisahkan data dengan kurva normal menjadi beberapa partisi. CCIA menggunakan K-means dan Density-based Multi Scale Data Condensation (DBMSDC)[24].

Kumar mengusulkan metode untuk penentuan pusat awal klaster. Metode ini menghitung jarak tiap data dengan Origin  $O(0,0)$ . Metode yang diusulkan menghasilkan hasil klaster yang sama meskipun dijalankan berulang kali. Langkah- langkah metode Kumar:

1. Tentukan banyaknya jumlah klaster  $k$ .
2. Untuk setiap objek hitung jarak tiap objek dengan Origin  $O(0,0)$ .
3. Jarak yang didapatkan, diurutkan dari terbesar ke terkecil atau sebaliknya.
4. Partisi data yang sudah diurutkan menjadi  $k$  partisi yang sama.
5. Nilai rata-rata setiap partisi digunakan untuk inisialisasi pusat awal klaster [25].

### 2.4 Landasan Teori

#### Data Mining

Data mining merupakan suatu metode yang bertujuan untuk menemukan pola di dalam sebuah data yang tidak terlihat sebelumnya. Dan diharapkan dari pola tersebut dapat menolong untuk mengambil sebuah keputusan diberbagai bidang ilmu. Data mining sering digunakan untuk pengolahan data yang besar, dan oleh sebab itu data mining sangat digunakan untuk pengolahan data dalam bidang industri, keuangan, cuaca, ilmu dan teknologi. Fungsi data mining di antaranya class, asosiasi, analisa korelasi, klasifikasi, prediksi, analisa klaster dan lainnya [26].

#### Algoritma K-mean

Algoritma K-Means pertama kali diperkenalkan oleh MacQueen JB pada tahun 1976, dan merupakan bagian dari algoritma clustering yang cukup sederhana yang memiliki tujuan untuk mengelompokkan suatu data besar kepada kelompok-kelompok tertentu yang disebut  $k$ . Dengan tujuan mempartisi data yang ada ke dalam bentuk satu atau lebih klaster[28]. Karena Algoritma K-Means memiliki kelebihan, salah satunya mampu mengelompokkan objek data besar, tetapi juga memiliki kelemahan diantaranya kesensitivitasnya tinggi pada pembangkitan titik pusat awal secara random, kadang-kadang hasil pengelompokan berubah-ubah sehingga keakuratannya belum tentu yang baik [10].

Cara kerja algoritma K-Means :

1. Menetapkan  $k$  total klaster yang diinginkan
2. Memunculkan centroid (titik pusat *cluster*) secara random
3. Menghitung jarak antar klaster  
 Dalam pengukuran jarak antar klaster algoritma yang sering digunakan adalah Euclidian Distance , dengan rumus sebagai berikut:

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

Dimana,

x : Titik data pertama,

y : Titik data kedua,

n : Jumlah karakteristik (attribut) dalam terminologi data mining,

d(x,y) : *Euclidian distance* yaitu jarak antara data pada titik x dan titik y menggunakan kalkulasi matematika

4. Menentukan centroid terdekat,  
Dalam hal ini dilihat jarak mana yang terdekat, jarak yang paling dekatlha yang akan masuk dalam kluster tersebut.
5. Menghitung nilai rata rata untuk menentukan cendroid baru,dari data-data yang ada pada *centroid* yang sama, kemudian melakukan iterasi selanjutnya dengan pusat kluster yang baru terbentuk apabila mendapatkan hasil belum konvergen.
6. Jika posisi cendroid yang varu dan yang lama tidak sama makan kembali kepada langkah yang ketiga

### Particle Swarm Optimization

*Particle Swarm Optimizaton* (PSO) merupakan algoritma pencarian berbasis populitas yang diinisialisasi dengan populasi solusi acak,dan di gunakan untuk memecahkan masalah optimasi[29]. PSO diperkenalkan oleh kennedy dan Eberhart pada tahun 1995 berdasarkan penelitian terhadap perilaku kawanan burung dan ikan. Setiap partikel dalam PSO juga di kaitkan dengan kecepatan partikel terbang melalui ruang pencarian dengan kecepatan yang dinamis disesuaikan untuk perilaku historis mereka. Oleh karena itu, partikel memiliki kecenderungan terbang menuju daerah yang lebih baik selama proses pencarian [8].

### Algoritma Davies Bouldin Index (DBI)

Dalam pengukuran metode clustering yang digunakan adalah Algoritma Davues Bouldin Index. Pengukuran ini bertujuan memaksimalkan jarak interkluster antara satu kluster dengan kluster yang lain. Dalam penelitian ini DBI akan digunakan untuk mendeteksi outlier pada masing-masing kluster yang terbentuk. DBI diperkenalkan oleh David L. Davies dan Donald W. Bouldin (1979) yang digunakan untuk mengevaluasi kluster. Validasi internal yang dilakukan adalah seberapa baikkah klastering yang sudah dilakukan, yaitu dengan menghitung kuantitas dan fitur turunan dari set data [30].

### Metode yang diusulkan

Untuk dapat mengatasi masalah dalam penentuan centroid awal pada algoritma K- means Metode K means – PSO. Penelitian ini mengusulkan Algoritma K-means Particle Swarm Optimization (K means – PSO) yaitu gabungan algoritma yang di usulkan untuk menemukan pusat centroid awal terbaik agar menghasilkan cluster terbaik dalam algoritma K-means, adapun langkah langkah yang diusulkan adalah :

1. Penentuan jumlah cluster  
Dalam penelitian ini mengikuti tahap dari algoritma K-means yaitu untuk menentukan jumlah kluster, secara acak atau random, hanya saja dalam data PAMSIMAS pemerintah ada menetapkan indicator penting dalam pelaksanaan PAMSIMAS yaitu kondisi Sanitasi Air Minum (SAM) dengan indicator, berfungsi dengan baik, berfungsi sebagian, dan tidak berfungsi, oleh sebab itu peneliti menentukan tiga kluster dalam penelitian ini.

## 2. Penentuan titik awal centroid

Langkah selanjutnya proses pencarian centroid awal terbaik, untuk dapat menghasilkan cluster yang lebih baik, peneliti mengusulkan untuk dapat melakukan proses pencarian centroid terbaik menggunakan metode *Particle Swarm Optimization* (disingkat PSO).

## 3. Melakukan inialisasi nilai ke dalam partikel

Untuk menentukan nilai centroid dilakukan pembangkitan partikel dan akan dipakai menjadi solusi terbaik, dan setiap partikel dibangkitkan dari dataset yang ada. Pada penelitian ini jumlah centroid yang akan dibangkitkan diperoleh dari partikel Inialisasi nilai partikel pada algoritma K means – PSO ditentukan secara acak dari data didalam dataset, sebagai contoh kita memiliki delapan atribut dan tiga puluh partikel yang dibangkitkan dan total akan dibentuk adalah 3 klaster, dari data asli yang dapat dilihat di table 3.1 pada data PAMSIMAS 2008 – 2017, centroid yang dibangkitkan memiliki jumlah atribut yang sama dengan data asli yang ada, dan penentuan nya dipilih secara random, Total partikel yang dibangkitkan menyesuaikan dengan jumlah swarm size yang telah ditentukan sejak awal.

## 4. Inialisasi Kecepatan

Tujuan dari fungsi kecepatan pada penelitian yang diusulkan adalah untuk merubah posisi suatu partikel menuju partikel dengan nilai centroid yang lebih optimal. Nilai suatu kecepatan merupakan range antara 0 sampai 1, karna apabila nilai suatu kecepatan semakin besar maka akan semakin lebar jumlah iterasi algoritma PSO untuk menentukan suatu solusi pada penelitian nilai kecepatan yang diberikaan adalah 0,1.

## 5. Menghitung nilai Fitness

Langkah yang akan dilakukan adalah perhitungan nilai fitness setelah memiliki nilai awal dan kecepatan pada partikel dan dalam penelitian ini fungsi nilai fitness yang digunakan adalah nilai minimum SSE, karna nilai minimum SSE akan digunakan untuk menentukan Partikel best dan Global best dalam sebuah iterasi PSO, dan tahap tahap yang akan dilakukan adalah sebagai berikut:

- a. Langkah yang dimulai dengan Perhitungan jarak yang digunakan adalah perhitungan euclidean distance untuk Setiap data yang telah dihitung nilai jaraknya dengan setiap centroid dimasukan ke dalam cluster yang memiliki nilai jarak terkecil. Proses ini terus dilakukan sampai seluruh data yang ada pada dataset telah memiliki kelompok kalster ,dengan rumus sebagai berikut:

$$D(x_1 - cd_1) = \sqrt{\sum_i (x_i - cd_i)^2} \quad (2)$$

Dimana,

x : Titik data pertama,

cd : titik centroid ,

D : Jumlah karakteristik (atribut) dalam terminologi data mining,

d(x,y) : *Euclidian distance* yaitu jarak antara data pada titik x dan titik y menggunakan kalkulasi matematika

- b. Menghitung SSE (*Sum Square Error*)

Pada penetian ini tahap perhitungan yang digunakan dalam perhitungan SSE adalah memilih satu titik pusat pusat cluster yang memiliki nilai yang paling minimum dari seluruh titik pusat



kluster dalam partikel tersebut, SSE diterjemahkan sebagai penjumlahan nilai minimum jarak data dengan pusat *cluster*. SSE dinyatakan dengan rumus berikut :

$$SSE = \sum_{i=1}^n (d)^2 \quad (3)$$

dimana,  $d$  adalah jarak antara data dengan pusat kluster

Dalam penelitian ini, SSE merupakan *fitness function* yang akan dicari nilainya dalam algoritma *clustering*. SSE inilah yang akan dicari nilai optimalnya (minimum) dengan menggunakan algoritma PSO.

Nilai minimum dengan persamaan sebagai berikut :

### Centroid 1

$$\begin{aligned} & \min \{(P_1cd_1), (P_1cd_2), (P_1cd_3)\} \\ & \min \{(P_1cd_1), (P_1cd_2), (P_1cd_3)\} = \min\{(0), (12331), (7653,12)\} = 0 \end{aligned}$$

### Centroid 2

$$\min\{(P_1cd_1), (P_1cd_2), (P_1cd_3)\} = \min\{(12331), (0), (4678,37)\} = 0$$

### Centroid 3

$$\begin{aligned} & \min\{(P_1cd_1), (P_1cd_2), (P_1cd_3)\} = \min\{(15173, 1), (2842), (7520, 42)\} \\ & = 2842 \end{aligned}$$

Hasil perhitungan jarak antara data dan partikel 1 yang dibangkitkan dapat dilihat pada table hasil nilai SSE pada partikel 1. Berdasarkan dari perhitungan tersebut, nilai Sum of Squared Error pada partikel satu yang dibangkitkan adalah 9571691, langkah berikutnya adalah menghitung kembali nilai Sum of Squared Error pada partikel selanjutnya sampai semua swarm size yaitu tiga puluh partikel.

## 6. Menentukan nilai Global best dan partikel best

Berikutnya adalah melakukan inisialisasi dan pembaruan terhadap nilai *particle best* ( $Pbest$ ) dan *global best* ( $GBest$ ). Nilai  $Pbest$  dan  $GBest$  digunakan untuk mencari perubahan kecepatan maupun posisi partikel. Solusi pencarian *centroid* terbaik, didapatkan dengan melakukan *update* terhadap nilai  $Pbest$  dan  $Gbest$  selama proses iterasi PSO berlangsung, sehingga menghasilkan partikel baru dengan nilai *centroid* yang lebih baik dari partikel sebelumnya. Nilai  $Pbest$  didapatkan dengan membandingkan nilai *fitness* partikel pada iterasi saat ini dengan nilai partikel pada iterasi sebelumnya.  $Pbest$  pada iterasi pertama akan bernilai sama dengan posisi partikel saat pertama diinisialisasikan. Sedangkan untuk iterasi lebih dari satu, maka penentuan nilai  $Pbest$  dilakukan dengan membandingkan nilai *fitness* antar partikel. Partikel yang mempunyai nilai *fitness* terkecil yang akan terpilih sebagai nilai  $Pbest$  di iterasi saat ini.

Global best didapatkan dari membandingkan semua partikel di seluruh iterasi PSO berdasarkan pada nilai *fitness*. Pada iterasi pertama,  $Gbest$  diperoleh dari partikel dalam satu populasi yang memiliki nilai *fitness* terkecil. Dimana partikel dengan nilai terkecil dan paling baik yang dipilih

sebagai *Gbest*. Sedangkan untuk iterasi lebih dari satu, *Gbest* akan melakukan perbandingan dengan nilai fitness partikel yang didapat dari seluruh proses iterasi. Selama proses iterasi PSO berlangsung apabila ada partikel yang menghasilkan nilai fitness paling kecil, maka partikel tersebut akan menjadi *Gbest* algoritma dan nilai dari *Gbest* tidak akan berubah apabila selama iterasi tidak ditemukan lagi partikel dengan nilai *Gbest* yang lebih kecil, untuk dapat lebih memahami proses penentuan partikel dapat dilihat hasil perhitungan Sum of Squared Error pada table 3.7. pada perhitungan tersebut ada kolom yang berwarna gelap itulah nilai partikel yang paling minimum diantara tiga puluh partikel yang lain. Pada perhitungan Sum Squared of Error dari tiga puluh partikel yang dibangkitkan dapat dilihat pada partikel ke 15 dengan nilai 515410 adalah nilai paling minimum

#### 7. Menghitung nilai velocity dan posisi

Langkah selanjutnya adalah melakukan *update* kecepatan pada setiap partikel didalam populasi. Untuk mencari posisi partikel terbaik yang dapat menjadi kandidat solusi *centroid* dalam ruang pencarian, maka secara iteratif nilai kecepatan harus mengalami suatu perubahan. Nilai kecepatan pada partikel dapat dihitung dengan menggunakan persamaan rumus velocity terlebih dahulu menentukan nilai pada Parameter swarm size, jumlah iterasi, nilai inersia ,  $C_1, C_2$  ,  $R_1$ , dan  $R_2$  diinisialisasi terlebih dahulu pada awal proses penentuan cluster.

Tabel 3.8 Parameter

No	Nama parameter	Nilai parameter
1	Swarm size	30 partikel
2	Iterasi	7 iterasi
3	Inersia	0,5
4	$C_1$	1
5	$C_2$	1
6	$R_1$	0,4
7	$R_2$	0,5

Sedangkan nilai *Pbest* dan *Gbest* didapatkan dari inisialisasi perhitungan nilai fitness. Perubahan kecepatan akan dilakukan secara berulang selama iterasi PSO berlangsung, sehingga nilai kecepatan dari seluruh partikel di iterasi sebelumnya akan berubah menjadi nilai kecepatan yang baru. Setelah melalui proses untuk mengupdate nilai kecepatan di setiap partikel, selanjutnya akan dilakukan proses perubahan posisi untuk mencari posisi terbaik partikel. Untuk mengupdate posisi partikel dapat dilihat berdasarkan nilai posisi partikel pada iterasi sebelumnya dan nilai kecepatan pada iterasi saat ini. Kemudian dengan menggunakan persamaan rumus, maka posisi partikel untuk iterasi saat ini dapat dihitung. Seperti halnya dengan perubahan nilai kecepatan, proses *update* posisi partikel juga dilakukan pada keseluruhan partikel didalam populasi. Perhitungan velocity dengan persamaan sebagai berikut :

$$V_{j(i)} = \omega * (i - 1) + c_1.r_1 [(P\_pbest_j - x_j(i - 1))] + c_2.r_2 [(G\_best_j - x_j(i - 1))]$$

$$V_{j(i)} = 0,5 * 0,1 + 1 * 0,4 * (103 - 20) + 1 * 0,5 * (103 - 20) = 74,5$$

$$V_{j(i)} = 0,5 * 0,1 + 1 * 0,4 * (21 - 13) + 1 * 0,5 * (21 - 13) = 7,25$$

$$V_{j(i)} = 0,5 * 0,1 + 1 * 0,4 * (115 - 20) + 1 * 0,5 * (115 - 20) = 85,55$$

$$V_{j(i)} = 0,5 * 0,1 + 1 * 0,4 * (109 - 20) + 1 * 0,5 * (109 - 20) = 80,15$$

Perhitungan posisi

$$X_{(i)}(t) = X_{(i)}(t - 1) + Vi(t)$$

$$X_{(i)}(t) = X_{(i)}(t - 1) + Vi(t)$$

$$X_{(i)}(t) = 20 + 75 = 95$$

$$X_{(i)}(t) = 13 + 7 = 20$$

$$X_{(i)}(t) = 20 + 86 = 106$$

$$X_{(i)}(t) = 20 + 80 = 10$$

$$X_{(i)}(t) = 15173 + 63302 = 78475$$

$$X_{(i)}(t) = 20 + 59 = 79$$

#### 4. Ulangi Langkah no 1 Selama *stopping criteria* Belum Terpenuhi

Stopping *criteria* yang digunakan pada proses pencarian *centroid* terbaik adalah nilai maksimum dari iterasi PSO. Nilai maksimum iterasi yang digunakan pada penelitian ini adalah tujuh iterasi. Saat proses evaluasi partikel PSO sudah mencapai tujuh kali iterasi, maka proses pengulangan algoritma akan dihentikan. Jika *stopping criteria* ini sudah tercapai maka proses akan dilanjutkan ke langkah ke-7. Sebaliknya jika belum terpenuhi maka proses akan diulangi dari langkah no 1 dengan menggunakan posisi dan kecepatan partikel baru yang dihasilkan oleh iterasi saat ini

#### 5. Gunakan pusat kluster dari hasil PSO sebagai pusat *Cluster* awal

Pada langkah terakhir pencarian *centroid* awal terbaik adalah menentukan partikel mana yang menghasilkan solusi paling baik dari seluruh partikel didalam populasi. Partikel terbaik diambil dari nilai *Gbest* yang didapatkan setelah batas iterasi maksimum PSO terpenuhi. Setelah ditemukan nilai *Gbest* terbaik, maka solusi *centroid* terbaik dari keseluruhan populasi partikel yang didapatkan yang menjadi pusat *cluster* awal terbaik yang digunakan pada *cluster* awal K- Means.

#### 6. Proses Pembentukan *Cluster*

Langkah ketiga dari algoritma K means -PSO adalah melakukan proses pembentukan *cluster* dengan cara melakukan pengelompokan setiap data ke dalam *cluster* menggunakan *centroid* yang dihasilkan pada proses sebelumnya sebagai *centroid* awal. Proses ini berbeda dengan algoritma K-Means yang membangkitkan *centroid* secara acak dari *dataset*. Pengelompokan data ke dalam *cluster* pada algoritma K means – PSO menggunakan fungsi *Euclidean distance* sebagai parameter pembandingan. Setiap data akan dihitung nilai jaraknya dengan setiap *centroid*. Kemudian data akan dikelompokkan ke dalam *cluster* yang *centroids* nya memiliki jarak terdekat dengan data. Proses pengelompokan data ke dalam *cluster* akan terus dilakukan sampai kondisi data yang ada didalam

*cluster* tidak mengalami perubahan dari hasil *cluster* sebelumnya. Hasil akhir proses pembentukan *cluster* pada algoritma K means - PSO adalah kumpulan *cluster* yang telah berisi data yang sesuai dengan karakteristik dari *cluster* tersebut.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Evaluasi dengan DBI

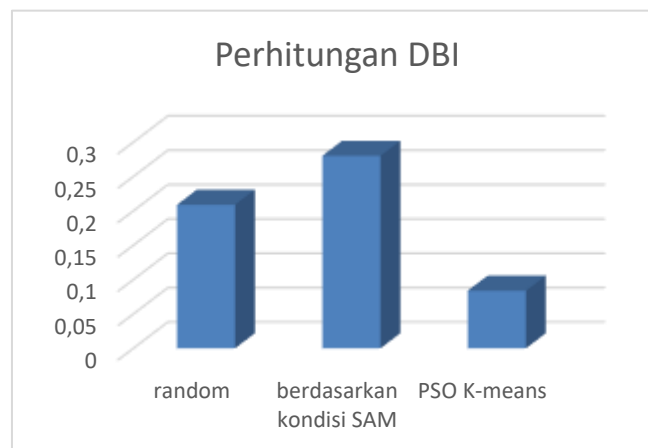
Mengevaluasi hasil kluster yang terbentuk menggunakan Davies Bouldin Index (DBI). Untuk evaluasi dengan nilai DBI yaitu semakin kecil nilai DBI maka hasil kluster semakin baik. Dengan semakin kecilnya nilai DBI maka dengan demikian performa metode yang digunakan dapat diketahui metode yang paling baik dalam klustering data. Validitas internal yang dilakukan adalah seberapa baik klustering sudah dilakukan dengan menghitung kuantitas dan fitur turunan dari *dataset*. *Sum of square within cluster* (SSW) sebagai metrik kohesi dalam sebuah kluster ke-*i*.

#### 3.2 Hasil dan Pembahasan

Eksperimen untuk dapat menentukan daerah yang menjadi acuan dalam menentukan kelompok daerah pada data PAMSIMAS dilakukan dengan menggunakan algoritma K- means sementara itu algoritma K-means memiliki masalah pada penentuan pusat awal kluster, oleh sebab itu peneliti mengusulkan untuk mengatasi masalah tersebut menggunakan algoritma PSO dalam pemilihan pusat kluster awal. Pusat awal kluster yang didapatkan digunakan untuk penentuan pusat awal kluster pada algoritma K-means. selanjutnya dihitung evaluasi *Davies-Bouldin Index*.

#### 3.3 Hasil Perbandingan Kinerja Metode

Pengukuran metode dilakukan dengan dataset PAMSIMAS tahun 2008- 2017 periode pelaporan bulan juni tahun 2018, dan metode yang digunakan adalah metode algoritma K- means menggunakan centroid secara acak dan penentuan centroid sesuai dengan ketentuan pemerintah tentang kondisi SAM dan metode K- menas PSO dimana algoritma PSO digunakan untuk menentukan centroid awal terbaik dari partikel yang dibangkitkan, hasil dari metode tersebut dapat dilihat di table 4.26



Gambar 1. Perbandingan hasil perhitungan DBI

## 4. KESIMPULAN DAN SARAN

### 4.1 Kesimpulan

Dalam penelitian ini dilakukan pengujian model dengan menggunakan *K-means* dan *K-means PSO* dalam pengelompokan kedalam 3 *cluster*. Model yang dihasilkan diuji untuk mendapatkan nilai Bouldin Index dari setiap algoritma sehingga didapat pengujian dengan menggunakan data dan setelah dilakukan pengujian dengan tools excel 2016 didapat nilai Bouldin Index adalah metode *K-means dengan pemilihan centroid awal acak dengan nilai 0,208856082*, metode *K-means dengan pemilihan centroid sesuai dengan standar pemerintah tentang kondisi SAM sebesar 0,280077 dan metode pemilihan yang terbaik adalah metode K-means PSO 0,08383*. Dengan menggunakan *Davies Bouldin Index* suatu *cluster* akan dianggap memiliki skema *clustering* yang optimal adalah yang memiliki *Index Davies Bouldin* minimal maka dapat disimpulkan pengujian data PAMSIMAS menggunakan *K-means PSO* didapat bahwa metode tersebut lebih optimal.

### 4.2 Saran

Dari hasil pengujian yang telah dilakukan dan hasil kesimpulan yang diberikan maka ada saran atau usul yang di berikan antara lain:

1. Dari penelitian ini di ketahui bahwa setiap cluster yang dihasilkan memiliki jarak dari cluster yang lain, dengan hasil cluster yang diperoleh diharapkan dapat diukur jarak setiap atribut antar cluster agar dapat ditentukan atribut mana yang perlu diperhatikan sebagai prioritas perkembangan pembangunan sarana dan prasaran air dan sanitasi.
2. Mencoba menerapkan metode optimasi sebagai bahan perbandingan seperti EM (*Expectation Maximisation*).

## DAFTAR PUSTAKA

- [1] E. Keberhasilan *Et Al.*, "Evaluasi Keberhasilan Program Penyediaan Air Minum Dan Sanitasi Berbasis Masyarakat (Pamsimas) Di Kabupaten," No. 50, Pp. 1–15, 2011.
- [2] P. T. Penyusunan, R. Aksi, And T. Pembangunan, "Draft," 2017.
- [3] Unicef, "Air Bersih, Sanitasi & Kebersihan," *Ringkasan Kaji.*, Pp. 1–6, 2012.
- [4] D. Jenderal And C. Karya, *Pedoman Umum Pengelolaan Pamsimas.* .
- [5] W. Li Xiang, N. Zhu, S. Feng Ma, X. Lei Meng, And M. Qing An, "A Dynamic Shuffled Differential Evolution Algorithm For Data Clustering," *Neurocomputing*, Vol. 158, Pp. 144–154, 2015.
- [6] Z. Huang, "Extensions To The K-Means Algorithm For Clustering Large Data Sets With Categorical Values," *Data Min. Knowl. Discov.*, Vol. 2, No. 3, Pp. 283–304, 1998.
- [7] D. W. Van Der Merwe And A. P. Engelbrecht, "Data Clustering Using Particle Swarm Optimization," *Evol. Comput. 2003. Cec'03. 2003 Congr.*, Vol. 1, Pp. 215–220, 2003.
- [8] J. Kennedy And R. Eberhart, "Particle Swarm Optimization," *Neural Networks, 1995. Proceedings., Ieee Int. Conf.*, Vol. 4, Pp. 1942–1948 Vol.4, 1995.
- [9] S. Petrovic, "A Comparison Between The Silhouette Index And The Davies-Bouldin Index In Labelling Ids Clusters," *The 11th Nord. Work. Secur. It-Systems, Nord. 2006*, P. 53–64., 2006.
- [10] R. O. Ph, *Advances In K-Means Clustering A Data Mining Thinking Doctoral.* .
- [11] F. Nasari And C. J. M. Sianturi, "Penerapan Algoritma K-Means Clustering Untuk

- Pengelompokan Penyebaran Diare Di Kabupaten Langkat,” *Cogito Smart J.*, Vol. 2, No. 2, Pp. 108–119, 2016.
- [12] N. H. Kristanto, A. C. L. A, And H. B. S, “Implementasi K-Means Clustering Untuk Pengelompokan Analisis Rasio Profitabilitas Dalam Working Capital,” *Juisi*, Vol. 02, No. 01, Pp. 9–15, 2016.
- [13] I. Pendahuluan, K. Ilmiah, And D. Mining, “Metode Manhattan , Euclidean Dan Chebyshev Pada Algoritma K-Means Untuk Pengelompokan Status Desa,” 2016.
- [14] P. D. W. Mega, “Clustering Menggunakan Metode K-Means Untuk Menentukan Status Gizi Balita,” *J. Inform.*, Vol. 15, No. 2, Pp. 160--174, 2015.
- [15] B. Darma, D. Setiawan, And R. S. Perdana, “Algoritme Genetika Untuk Optimasi K-Means Clustering Dalam Pengelompokan Data Tsunami,” *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, Vol. 2, No. 10, Pp. 3865–3872, 2018.
- [16] N. Yusup, A. M. Zain, And S. Z. M. Hashim, “Overview Of Pso For Optimizing Process Parameters Of Machining,” *Procedia Eng.*, Vol. 29, Pp. 914–923, 2012.
- [17] Y. Eka, A. Stmik, N. M. Jakarta, And K. Kunci, “Penerapan Metode Particle Swarm Optimization Pada Optimasi Prediksi Pemasaran Langsung,” *Desember*, Vol. 5, No. 1, Pp. 1–11, 2018.
- [18] E. R. Zerda, “Analisis Dan Penerapan Algoritma Particle Swarm Optimization ( Pso ) Pada Optimasi Penjadwalan Sumber Daya Proyek,” *Tugas Akhir (Skripsi)*, P. Cover, I–Xi, 1-42, 2009.
- [19] A. Yasid, “Implementasi Automatic Clustering Menggunakan Particle Swarm Optimization Dan Genetic Algorithm Pada Data Mahasiswa,” Pp. 9–10, 2015.
- [20] A. Saidul And J. L. Buliali, “Implementasi Particle Swarm Optimization Pada K-Means Untuk Clustering Data Automatic Dependent Surveillance-Broadcast,” *Eksplora Inform.*, Vol. 8, No. 1, P. 30, 2018.
- [21] M. Teja, A. Cipta, And F. C-, “C-Means Dalam Pengelompokan Kelas,” Vol. 11, No. 1, Pp. 72–91, 2018.
- [22] M. Al Hasan, V. Chaoji, S. Salem, And M. J. Zaki, “Robust Partitional Clustering By Outlier And Density Insensitive Seeding,” *Pattern Recognit. Lett.*, Vol. 30, No. 11, Pp. 994–1002, 2009.
- [23] M. M. Breunig *Et Al.*, “Lof: Identifying Density-Based Local Outliers,” *Proc. 2000 Acm Sigmod Int. Conf. Manag. Data - Sigmod '00*, Vol. 29, No. 2, Pp. 93–104, 2000.
- [24] S. S. Khan And A. Ahmad, “Cluster Center Initialization Algorithm For K-Modes Clustering,” *Expert Syst. Appl.*, Vol. 40, No. 18, Pp. 7444–7456, 2013.
- [25] M. Goyal And S. Kumar, “Improving The Initial Centroids Of K-Means Clustering Algorithm To Generalize Its Applicability,” *J. Inst. Eng. Ser. B*, Vol. 95, No. 4, Pp. 345–350, 2014.
- [26] J. Han, M. Kamber, And J. Pei, *Data Mining: Concepts And Techniques*. 2012.
- [27] M. E. Celebi, H. A. Kingravi, And P. A. Vela, “A Comparative Study Of Efficient Initialization Methods For The K-Means Clustering Algorithm,” *Expert Syst. Appl.*, Vol. 40, No. 1, Pp. 200–210, 2013.
- [28] J. Wu, *Advanced In K-Means Clustering*. 2012.

- [29] C. Grosan, A. Abraham, And M. Chis, “Swarm Intelligence In Data Mining,” Vol. 34, No. 2006, Pp. 1–20, 2006.
- [30] D. L. Davies And D. W. Bouldin, “A Cluster Separation Measure,” *Ieee Trans. Pattern Anal. Mach. Intell.*, Vol. Pami-1, No. 2, Pp. 224–227, 1979.
- [31] F. Cao, J. Liang, And G. Jiang, “An Initialization Method For The K-Means Algorithm Using Neighborhood Model,” *Comput. Math. With Appl.*, Vol. 58, No. 3, Pp. 474–483, Aug. 2009.

#### Biodata Penulis



Ari Yunus Hendrawan, lahir di Kuala Kapuas 12 agustus 1988, Magister Komputer dari Universitas Dian Nuswantoro dan bekerja sebagai Dosen di STIMIK AKI Pati