

**CLUSTERING PENGUNJUNG MALL MENGGUNAKAN METODE
K-MEANS DAN PARTICLE SWARM OPTIMIZATION**

**SKRIPSI
(JALUR SCIENTIST)**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh
TEUKU MUHAMMAD DISTA
19.11.2840

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2022

HALAMAN PENGESAHAN

SKRIPSI

**CLUSTERING PENGUNJUNG MALL MENGGUNAKAN METODE
K-MEANS DAN PARTICLE SWARM OPTIMIZATION**

yang disusun dan diajukan oleh

Teuku Muhammad Dista

19.11.2840

Telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Oktober 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Winda Mega P.D., M.Kom

NIK. 190302185

Erni Seniwati, S.Kom., M.Cs

NIK. 190302231

Ferlan Fauzi Abdullah, M.Kom

NIK. 190302276



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 7 November 2022

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, S.Kom., M.Kom.

NIK. 190302096

IDENTITAS JURNAL

Nama Jurnal : JURNAL MEDIA INFORMATIKA BUDIDARMA

Volume : 6

Nomor : 3

Halaman : 10 Halaman

Akreditasi : Sinta 3

Tahun Terbit : Juli 2022, Page 1339-1348

Judul : CLUSTERING PENGUNJUNG MALL MENGGUNAKAN
METODE K-MEANS DAN PARTICLE SWARM
OPTIMIZATION

Nama Penulis : Teuku Muhammad Dista, Ferlan Fauzi Abdulloh





HASIL REVIEW PROPOSAL

Jalur Non Skripsi

Sleman, 30-09-2022 19:42:09

Kepada Teuku Muhammad Dista,

Yang bertanda tangan di bawah ini:

Nama : **Windha Mega Pradnya D, M.Kom.**
NIK : **190302185**
Jabatan : **Ketua Program Studi S1 Informatika**

Tim Reviewer Proposal Non-Skripsi prodi Informatika telah melakukan uji kelayakan atas proposal dengan rincian sebagai berikut

Nim : 19.11.2840
Nama : Teuku Muhammad Dista
Jalur : Scientist
Poin Review : 95.41666666666667
Kesimpulan :

Direkomendasikan. Silakan mengajukan judul dengan melampirkan surat rekomendasi.

Demikian Surat ini dibuat untuk dapat digunakan sebagaimana mestinya.

Hormat kami,

Ketua Program Studi Informatika



Windha Mega P.D.,M.Kom.



JURNAL MEDIA INFORMATIKA BUDIDARMA

eISSN 2548-8368 / pISSN 2614-5278

Sekretariat : UNIVERSITAS BUDI DARMA | Jl. Sisingamangaraja No. 338, Medan, Sumatera Utara

Website: <https://ejournal.stmik-budidarma.ac.id/index.php/mib>

Email: mib.stmikbd@gmail.com

Medan, 7 Juni 2022

No : 544/MIB/LOA/VI/2022
Lamp : -
Hal : Surat Penerimaan Naskah Publikasi Jurnal

Kepada Yth,
Bapak/Ibu **Teuku Muhammad Dista**
Di Tempat

Terimakasih telah mengirimkan artikel ilmiah untuk diterbitkan pada **Jurnal Media Informatika Budidarma** (eISSN 2548-8368 / pISSN 2614-5278), dengan judul:

Clustering Pengunjung Mall Menggunakan Metode K-Means dan Particle Swarm Optimization

Penulis: **Teuku Muhammad Dista, Ferlan Fauzi Abdulloh***()

Berdasarkan hasil review dari reviewer, artikel tersebut dinyatakan **DITERIMA** untuk dipublikasikan pada **Volume 6, Nomor 3, Juli 2022**.

Sebagai informasi QR-Code digunakan untuk melihat link LOA Jurnal Media Informatika Budidarma, **Volume 6, Nomor 3, Juli 2022** yang telah dikeluarkan. Mohon segera untuk mengirimkan Copyright Transfer Form ke Email Jurnal MIB.

Demikian informasi yang kami sampaikan, atas perhatiannya kami ucapkan terimakasih.



Hormat Kami,

Surya Darma Nasution, M.Kom
Ketua Editor Jurnal MIB

Tembusan:

1. Author
2. Files



Clustering Pengunjung Mall Menggunakan Metode K-Means dan Particle Swarm Optimization

Teuku Muhammad Dista, Ferian Fauzi Abdulloh*

Informatika, Universitas AMIKOM Yogyakarta, Yogyakarta, Indonesia

Email: *teuku.dista@students.amikom.ac.id, ferian@amikom.ac.id

Email Penulis Korespondensi: ferian@amikom.ac.id

Abstrak—Penelitian ini bertujuan untuk mengcluster pengunjung mall. Hal ini dilatarbelakangi oleh pendapatan mall yang kian menurun sejak pandemi. Nantinya dari beberapa cluster tersebut kita dapat mengetahui karakteristik dari pengunjung mall tersebut. Karakteristik itulah yang nantinya akan kita manfaatkan untuk menambah pendapatan dari mall tersebut. Pada penelitian ini menggunakan dataset dari Kaggle yang bernama Pengunjung_mall berbentuk CSV yang nantinya di olah dengan menggunakan Bahasa Python di Jupiter Notebook menggunakan metode K-Means. Untuk memastikan seberapa akurat metode K-Means tersebut dilakukan optimasi dengan metode PSO (Particle Swarm Optimization). Setelah dilakukan Clustering dan optimasi menggunakan Jupiter Notebook, selanjutnya hasil akan di evaluasi dengan DBI (Davies Bouldin Index) di Microsoft Excel untuk mengetahui seberapa baik Clustering yang di Hasilkan. Hasil Clustering yang di dapatkan di gunakan sebagai acuan menentukan karakteristik pengunjung mall yang menjadi salah satu strategi untuk meningkatkan keuntungan Mall. Hasilnya, kami telah berhasil membagi pelanggan mall menjadi 5 cluster berdasarkan pendapatan tahunan yang mereka dapatkan dan skor pengeluaran. Cluster tersebut sudah di optimasi dengan PSO dan berhasil meningkatkan cluster hasil metode K-Means yang di buktikan dengan metode Davies Bouldin Index. Penelitian ini telah disimpulkan bahwa pelanggan yang memiliki tingkat pendapatan tinggi dan memiliki skor pengeluaran tinggi adalah target dengan tingkat prioritas utama bagi mall.

Kata Kunci: Clustering; Algoritma; K-Means; Particle Swarm Optimization; Davies Bouldin Index; Jupyter Notebook

Abstract—This research aims to cluster mall visitors. This is motivated by the mall's income which has decreased since the pandemic. Later from these several clusters we can find out the characteristics of the mall's visitors. These characteristics will be used later to increase the income from the mall. In this research, we use a dataset from Kaggle named Pengunjung_mall in CSV format which will later be processed using Python language on Jupiter Notebooks using the K-Means method. To ensure how accurate the K-Means method is, optimization is carried out using the PSO (Particle Swarm Optimization) method. After performing clustering and optimization using Jupyter Notebook, the results will then be evaluated with DBI (Davies Bouldin Index) in Microsoft Excel to find out how well the Clustering is generated. The Clustering results obtained are used as a reference to determine the characteristics of mall visitors which is one strategy to increase Mall profits. As a result, we have succeeded in dividing mall customers into 5 clusters based on their annual earned income and expense scores. The cluster has been optimized with PSO and has succeeded in increasing the cluster resulting from the K-Means method which is proven by the Davies Bouldin Index method. This research has concluded that customers who have high income levels and have high spending scores are the targets with the highest priority level for malls.

Keywords: Clustering; Algorithm; K-Means; Particle Swarm Optimization; Davies Bouldin Index; Jupyter Notebook

I. PENDAHULUAN

Pada zaman seperti saat ini, persaingan pasar kian meningkat. Akibatnya, memiliki keunggulan kompetitif yang berkelanjutan sangat penting bagi Sebagian besar perusahaan yang berusaha untuk mempertahankan atau memperluas posisi mereka dan mengoptimalkan probabilitas serta pertumbuhan laba/keuntungan. Ada banyak bentuk keunggulan, tetapi keunggulan yang baik adalah keunggulan yang dapat menarik klien/pelanggan. Oleh karena itu, perusahaan harus memiliki pengetahuan tentang kliennya sehingga dapat menyesuaikan aktivitas bisnisnya untuk mengoptimalkan utilitas yang diperoleh dari pelanggannya dan kepuasan yang diberikannya kepada mereka. Secara garis besar pelanggan adalah aset yang sangat berharga bagi perusahaan. Oleh karena itu, perusahaan harus melakukan upaya khusus untuk mempertahankan pelanggan. Di sisi lain, pelanggan di perusahaan tidak semua menguntungkan. Pelanggan yang kurang menguntungkan cenderung tidak berkontribusi pada profitabilitas perusahaan dan bahkan dapat menghasilkan kerugian yang signifikan serta peristiwa negatif lainnya, sehingga klien/pelanggan ini harus ditargetkan dengan skema yang berbeda. Oleh karena itu, misi penting dari auditor internal perusahaan harus melakukan pemeriksaan apakah proses bisnis saat ini secara efisien dan konsisten mempertahankan pelanggan yang diinginkan. Selain itu perusahaan harus mengendalikan kerugian yang timbul dari pelanggan yang kurang menguntungkan. Dengan demikian, perusahaan dapat meningkatkan posisi pasarnya dalam industrinya.

Untuk mencapai misi ini, segmentasi pelanggan sangat penting. Clustering adalah teknik analisis data yang dapat digunakan dengan tepat untuk segmentasi pelanggan. Teknik ini menciptakan kelompok objek target berdasarkan informasi yang ditemukan dalam data yang membedakan objek dan hubungan di antara mereka. Kondisi dalam cluster tertentu mirip satu sama lain dan berbeda dari kondisi di cluster lain. Dengan mengelompokkan pelanggan ke dalam suatu Cluster, pelanggan yang lebih disukai dan yang kurang disukai dapat dibedakan. Setelah segmentasi pelanggan, auditor dapat memeriksa efisiensi dan efektivitas kegiatan bisnis perusahaan saat ini dalam hal memberikan layanan yang sesuai untuk berbagai jenis pelanggan dengan harga yang sesuai. Pada hal ini kami menyelidiki masalah audit internal yang terkait dengan protokol untuk menawarkan



diskon kepada berbagai pelanggan di suatu Mall. Tujuan utamanya adalah untuk memeriksa apakah diskon telah ditawarkan dengan tepat untuk berbagai jenis pelanggan. Untuk mencapai tujuan ini, teknik pengelompokan /yang sering di sebut *Clustering* digunakan untuk mengelompokkan pelanggan. Setelah dataset memasuki tahap preprocessing barulah data dapat di *Clustering*. *Clustering* pada tahap ini menggunakan metode *K-Means Clustering* yang di optimasi dengan PSO (*Particle Swarm Optimization*) dan di evaluasi dengan DBI (*Davies Bouldin Index*).

Penelitian terdahulu dilakukan oleh Musthofa Galih Pradana dan Hoang Thi Ha yang berjudul "Memaksimalkan Strategi Peningkatan Segmentasi Pelanggan Mall Menggunakan *K-Means Clustering*"[1]. Berdasarkan penelitian ini difokuskan untuk mengetahui perilaku pelanggan dengan melihat hasil clustering. Penelitian ini membuktikan bahwa machine learning dapat diterapkan di segmentasi industri distrik perbelanjaan. Tetapi dengan asumsi pembelajaran mesin dapat melakukan pengelompokan dengan akurasi yang cukup akurat.

Penelitian terdahulu dilakukan oleh Mardalius dan Tika Christy, yang berjudul "Pemetaan potensi pelanggan sebagai strategi promosi Pakaian menggunakan algoritma *K-Means Clustering*"[2]. Hal ini dilatar belakangi oleh rencana promosi yang dilakukan tanpa konsep yang jelas sehingga dapat menimbulkan kerugian jika tidak dilakukan tepat sasaran. Berdasarkan penelitian ini bertujuan untuk memetakan calon pelanggan berdasarkan wilayah atau kecamatan yaitu cluster 1 memiliki 3 kecamatan, cluster 2 memiliki 7 kecamatan dan cluster 3 memiliki 13 distrik. Hasil tersebut diperkuat dengan pengujian software RapidMiner dengan akurasi data mengikuti hasil perhitungan dari 25 data.

Penelitian terdahulu dilakukan oleh Jun Dai, Paul Byrnes dan Miklos Vasarhelyi yang berjudul "Apakah Pelanggan Ditawari Diskon yang Sesuai? Studi Eksplorasi Menggunakan Teknik *Clustering* dalam Audit Internal" [3]. Berdasarkan penelitian ini bertujuan untuk penerapan metode *K-Means* Untuk menentukan kebijakan diskon bagi nasabah. Untuk mencapai tujuan penelitian ini teknik clustering di gunakan untuk menentukan karakteristik pelanggan. Setelah reduksi dimensi, penskalaan dan normalisasi data, selanjutnya penerapan algoritma *K-Means* untuk membuat cluster berdasarkan 6 atribut terpenting. Akibatnya 7 pemisahan yang sangat baik dari kelompok pelanggan kartu kredit dari mulai yang paling di sukai sampai paling tidak di sukai. Oleh karena itu, kebijakan diskon yang lebih relevan dan masuk akal harus dibuat untuk membedakan diskon yang ditawarkan kepada berbagai jenis pelanggan.

Penelitian terdahulu dilakukan oleh Rina Yuliana Sari, Hardian Oktavianto, Henny Wahyu Sulisty yang berjudul "Algoritma *K-Means* Dengan Metode *Elbow* Untuk Mengelompokkan Kabupaten/Kota Di Jawa Tengah Berdasarkan Komponen Pembentuk Indeks Pembangunan Manusia" [4]. Penelitian ini bertujuan untuk membantu pemerintah untuk dapat mengetahui permasalahan serta mempertimbangkan pengambilan kebijakan pada wilayah kabupaten/kota di provinsi Jawa Tengah berdasarkan variabel variabel IPM dengan memanfaatkan metode *Clustering*. Penelitian ini menggunakan metode *K-Means*, yang merupakan algoritma efektif untuk menganalisis data dalam jumlah besar.

Penelitian terdahulu dilakukan oleh Augusto Luis Ballardini yang berjudul "Tutorial tentang *Particle Swarm Optimization Clustering*"[5]. Berdasarkan penelitian ini mengusulkan penggunaan algoritma *Particle Swarm Optimization* untuk memperbaiki cluster yang di bentuk oleh Algoritma *K-Means*. Hasil penelitian ini menunjukkan bahwa algoritma PSO berkinerja lebih baik dibandingkan dengan algoritma *K-Means*. Algoritma PSO bekerja lebih lambat tetapi kesalahan kuantisasi lebih rendah sedangkan algoritma *K-Means* bekerja lebih cepat tetapi kesalahan kuantisasi lebih tinggi.

Pada penelitian ini penggunaan metode *Particle Swarm Optimization* untuk optimasi cluster yang di hasilkan oleh Metode *K-Means* terbukti tepat. Pasalnya metode *Particle Swarm Optimization* membuat cluster yang di hasilkan semakin baik. Hal ini di buktikan dengan adanya evaluasi sebelum dan setelah optimasi menggunakan metode *Davies Bouldin Index*. Diharapkan adanya penelitian ini dapat digunakan sebagai penyempurnaan dipenelitian sebelumnya dan dapat digunakan sebagai acuan penelitian selanjutnya.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

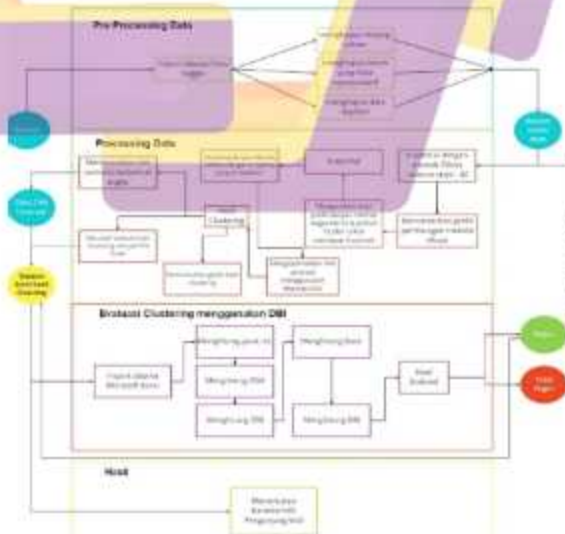
Pada penelitian ini dilakukan berdasarkan kerangka kerja penelitian. Kerangka kerja penelitian yang kami lakukan sebagai berikut :



Gambar 1. Kerangka Kerja Penelitian

Berdasarkan kerangka kerja penelitian pada **Gambar 1** diatas, maka dapat diuraikan pembahasan masing masing tahapan penelitian yaitu:

1. **Studi Literatur**
 Pada tahap ini penulis melakukan pencarian terhadap landasan landasan teori yang di peroleh dari berbagai sumber seperti buku, jurnal ilmiah dan refrensi lainnya untuk melengkapi penelitian baik mengenai konsep dan teori sehingga memiliki acuan yang baik dan relevan bagi penelitian.
2. **Menganalisis dataset**
 Pada tahap ini penulis menganalisis dataset dari Kaggle sebagai bahan penelitian. Dataset berupa file CSV yang nantinya akan di proses di tahap selanjutnya. Dataset yang telah di analisis sesuai dengan tema yang di bawakan dalam penelitian yaitu mengenai pelanggan mall.
3. **Identifikasi masalah**
 Pada tahap ini penulis mengidentifikasi hal apa saja yang mungkin menjadi masalah yang menghambat mall untuk memperoleh keuntungan. Sehingga penulis dapat menentukan rencana seperti apa yang akan dituju untuk menetapkan strategi guna memperbesar keuntungan mall.
4. **Tahapan clustering dan evaluasi cluster**
 Pada tahap clustering dan evaluasi cluster dilakukan dengan beberapa tahapan. Adapun alurnya dapat di lihat di **Gambar 2**



Gambar 2 . Alur Proses Clustering Dan Evaluasi Cluster



Berikut ini penjelasan mengenai tahapan-tahapan pada alur proses pada **Gambar 2** :

A. Preprocessing Data

Preprocessing data adalah proses dimana pada tahap ini dilakukan untuk mengubah data mentah ke dalam bentuk data yang mudah untuk di olah atau di pahami. Proses ini penting dalam memulai processing data karena data mentah sering kali memiliki format yang tidak teratur. Pada preprocessing data ini di lakukan penghapusan missing value, menghapus kolom yang tidak representatif atau tidak di pakai, menghapus data yang duplikat. Setelah itu di simpan lagi ke format yang sama dan data siap untuk di masukkan di proses selanjutnya

B. Processing Data

Processing data merupakan proses dimana pada tahap ini dilakukan proses clustering data yang sudah melalui proses preprocessing. Proses ini di mulai dengan mencari k optimal dengan metode elbow untuk mencari tahu seberapa banyak cluster yang baik dan ideal. Setelah k optimal di dapatkan proses selanjutnya yaitu clustering menggunakan metode *K-Means* dan dilanjutkan pengoptimalan titik centroid menggunakan metode PSO. Setelah proses clustering selesai kemudian hasil dari clustering tersebut di rubah menjadi file excel untuk memasuki proses selanjutnya.

C. Evaluasi cluster menggunakan DBI

Pada proses ini hasil clustering yang sudah di dapatkan akan di evaluasi dengan DBI . apabila hasil perhitungan DBI menunjukkan hasil angka mendekati 0 maka cluster tersebut sudah di katakana baik. Hasil evaluasi akan di bagi menjadi 2 hasil yaitu clustering setelah menggunakan metode *K-Means* dan setelah di optimasi dengan metode PSO. Dari kedua hasil evaluasi tersebut akan di bandingkan cluster mana yang paling baik.

D. Hasil

Hasil pada proses ini adalah berupa file excel yang siap untuk di analisis dan di tentukan kriterianya.

5. Menentukan karakteristik cluster

Karakteristik pelanggan akan di analisis berdasarkan kriteria berikut :

- Pendapatan tahunan yang diperoleh (gaji)
- Pengeluaran yang dikeluarkan untuk berbelanja di mall

6. Menetapkan diskon yang tepat bagi setiap cluster

Menetapkan diskon yang tepat bagi pelanggan adalah suatu service yang harus di lakukan oleh perusahaan guna meningkatkan penjualan. Pada tahap ini peneliti menganalisis hasil dari karakteristik pelanggan. Karakteristik pelanggan dapat di bedakan melalui pelanggan yang paling di sukai – pelanggan yang tidak di sukai. Tentunya peneliti harus pintar pintar untuk memberi diskon bagi masing masing karakteristik.

7. Hasil

Hasil dari clustering final akan menjadi acuan untuk mengetahui karakteristik pelanggan mall. Karakteristik pelanggan tersebut akan di manfaatkan untuk memberi service bagi pelanggan berupa diskon yang di harapkan nantinya penjualan dapat meningkat dan menambah keuntungan bagi mall tersebut.

8. Pembuatan laporan

Setelah semua tahapan penelitian dilakukan, selanjutnya yaitu akan dibuat laporan sebagai dokumentasi penelitian agar dapat dimanfaatkan dikemudian hari baik oleh peneliti maupun peneliti lainnya.

2.2 Machine Learning

Kita sering melihat penerapan machine learning di berbagai bidang di sekitar kita, misalnya di Facebook, machine learning membantu kita mengenali diri sendiri dan teman teman kita bahkan di Youtube dapat merekomendasikan video berdasarkan hal-hal yang anda minati. Secara umum machine learning dikategorikan menjadi 2 jenis yaitu *Supervised learning* dan *Unsupervised learning*. Pembelajaran terbimbing biasanya digunakan oleh seorang analis data untuk memecahkan masalah seperti klasifikasi dan regresi [6][7], yang berarti bahwa dalam hal ini data terdapat label target yang ingin Anda prediksi di masa depan, misalnya memprediksi nilai siswa atau jumlah pengeluaran bulanan. Sebaliknya pada *unsupervised learning*, penggunaannya tidak selalu memiliki label atau target khusus untuk diprediksi, misalnya clustering, berdasarkan model matematisnya, algoritma pada *unsupervised learning* tidak memiliki target variable [8]. Misalnya kita ingin mengelompokkan siswa berdasarkan kebiasaan belajarnya atau membuat cluster berdasarkan jumlah pembelian suatu produk.

Industri pemasaran, terutama mal, memiliki persaingan yang ketat untuk meningkatkan pelanggan dan menghasilkan keuntungan yang besar. Untuk mencapai tugas ini, pembelajaran mesin sudah diterapkan oleh banyak toko dan pasar lainnya [9], mal atau pusat perbelanjaan memanfaatkan data yang mereka dapatkan saat bertransaksi dengan pelanggannya dan memanfaatkannya dengan mengembangkan model ML untuk menasar yang tepat [10]. Hal ini tidak hanya meningkatkan penjualan & jumlah pengunjung yang datang tetapi juga meningkatkan efisiensi dalam berbisnis.

2.3 Metode Clustering

Clustering dikenal sebagai metode untuk mengidentifikasi kelompok umum dalam kumpulan data. Entitas dalam setiap grup secara komparatif lebih mirip dengan entitas dari grup tersebut daripada entitas grup lainnya. Sejak tahun 70-an, segmentasi berbasis kluster sangat sering digunakan dalam berbagai penelitian yang melibatkan data,



terutama dalam bidang pemasaran, bahwa *clustering* bukanlah metode analisis data yang terstruktur, meskipun memiliki fleksibilitas yang baik, sangat tergantung pada data atau sampel yang digunakan [11].

2.4 Metode Elbow

Metode *Elbow* adalah metode untuk menghasilkan informasi mengenai K optimal cluster. Informasi tersebut di dapat dengan cara melihat perbandingan hasil antara jumlah cluster yang akan membentuk suatu titik terakhir [4]. K Optimal nantinya akan digunakan sebagai model data untuk mendapatkan cluster terbaik. Berikut ini merupakan rumus SSE :

$$SSE = \sum_{i=1}^k \sum_{x \in S_i} ||x_i - C_k||^2 \quad (1)$$

Keterangan rumus :

X_i : Nilai atribut dari data ke- i

C_k : Nilai atribut titik pusat Cluster ke- i

2.5 K-Means Clustering

K-means merupakan salah satu metode data *clustering* non *hirarki* untuk clustering dokumen yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/kelompok berdasarkan atribut menjadi k partisi, dimana $k < n$ [12], sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain [13]. Algoritma ini dinilai cukup efektif dalam pengolahan data mining yang digunakan dalam pengelompokan (*clustering*) data yang akan dianalisa [14]. Algoritma ini juga bertujuan untuk menemukan grup dalam data, dengan jumlah grup yang diwakili oleh variabel K . Variabel K sendiri adalah jumlah cluster yang diinginkan. Pada setiap cluster terdapat titik pusat (*centroid*). Dimulai dari menentukan cluster dengan metode elbow, selanjutnya menghitung jarak data dengan titik pusatnya menggunakan jarak Euclidean berikut

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

Keterangan rumus :

$d(x, y)$ - jarak

x_i - data training

y_i - data testing

i - variable data

n - dimensi data

Proses pengelompokan *K-means* dapat dihentikan ketika salah satu dari kriteria berikut terpenuhi yaitu ketika jumlah maksimum iterasi telah terlampaui, ketika ada sedikit perubahan vektor *centroid* selama iterasi atau ketika tidak ada perubahan keanggotaan cluster. Algoritma dihentikan ketika jumlah iterasi yang ditentukan pengguna telah terlampaui.

2.6 Particle Swarm Optimization

Particle swarm Optimization (PSO) adalah proses pencarian stokastik berbasis populasi, dimodelkan setelah perilaku sosial, di mana setiap partikel mewakili solusi potensial untuk masalah optimasi. Tujuan dari PSO adalah untuk menemukan posisi partikel yang menghasilkan dari fungsi fitness objektif yang diberikan. Setiap partikel mewakili posisi di N_d dimensional space dan melalui pencarian multi dimensi. Setiap partikel mempertahankan informasi yang mempunyai posisi terbaik dan posisi paling atas di lingkungannya.

x_i : Posisi saat ini pada partikel

v_i : kecepatan saat ini pada partikel

y_i : posisi saat ini pada partikel

Menggunakan notasi di atas, posisi partikel disesuaikan berdasarkan :

$$v_{i,k}(t+1) = wv_{i,k}(t) + c1r1.k(t)(y_{i,k}(t) - x_{i,k}(t)) + c2r2.k(t)(\hat{y}_k(t) - x_{i,k}(t)) \quad (3)$$

$$X_i(t+1) = X_i(t) + V_i(t+1) \quad (4)$$

Dimana w adalah berat inerti dan $c1$ dan $c2$ adalah percepatan konstanta. Kecepatan dengan demikian dihitung berdasarkan tiga kontribusi yaitu sebagian kecil dari kecepatan sebelumnya, komponen kognitif yang merupakan fungsi dari jarak partikel dari posisi terbaik dan komponen sosial yang merupakan fungsi dari jarak partikel dari partikel paling tinggi yang ditemukan. Posisi personal best dari partikel dapat di hitung sebagai berikut



$$y_i(t+1) = \begin{cases} y_i(t) & \text{if } f(x_i(t+1)) \geq f(y_i(t)) \\ x_i(t+1) & \text{if } f(x_i(t+1)) < f(y_i(t)) \end{cases} \quad (5)$$

Pendekatan dasar dari PSO adalah berdasarkan interpretasi dari lingkungan partikel. Setiap partikel memperbarui posisinya sendiri berdasarkan pengalamannya sendiri Pbest (Personal Best) dan pengalaman partikel terbaik di swarm Gbest (Global Best)[15].

2.6.1 Particle Swarm Optimization Clustering

Dalam konteks *clustering* satu partikel mewakili N_c cluster centroid vector artinya setiap particle x_i terbentuk sebagai berikut

$$x_i = (m_{i1}, m_{i2}, \dots, m_{iN_c}) \quad (6)$$

di mana m_{ij} adalah centroid ke- j dalam partikel ke- i . Kawanan partikel mewakili beberapa solusi kandidat untuk masalah pengelompokan. Selanjutnya, di masing-masing iterasi PSO mencoba meningkatkan solusi kandidat. Kualitas *clustering* dapat diukur dengan menggunakan kesalahan kuantisasi sebagai berikut

$$f = \frac{\sum_{j=1}^{N_c} \sum_{p \in z^d(zp, m_j)} |z_p - m_j|^2}{N_c} \quad (7)$$

Di mana Z adalah vektor data yang akan di cluster, dan d adalah fungsi jarak *Euchidean*

2.7 Davies Bouldin Index

Evaluasi menggunakan *Davies Bouldin Index* ini adalah evaluasi yang berkerja di internal cluster, dimana baik atau tidaknya hasil cluster dilihat dari nilai DBI yang terkecil [16]. *Davies-Bouldin Index* merupakan salah satu metode yang digunakan untuk mengukur sejauh mana ketepatan suatu cluster pada suatu metode clustering [17]. Kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap titik pusat cluster dari cluster yang diikuti. Sedangkan separasi didasarkan pada jarak antar titik pusat cluster terhadap cluster-nya. Perhitungan evaluasinya di mulai dari menentukan nilai SSW, SSB dari titik centroid awal lalu akan menentukan nilai DBI nya.

Persamaan yang digunakan untuk memperoleh nilai *Sum of Square Within cluster* (SSW)

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (8)$$

Keterangan rumus :

m - jumlah data dalam cluster ke i

x - data pada cluster

c - centroid

$d(x,c)$ - jarak data dengan centroid

Sum of Square Between-cluster (SSB) bertujuan untuk mengetahui separasi antar cluster.

$$SSB_{i,j} = d(c_i, c_j) \quad (9)$$

Keterangan rumus :

C_i - cluster Saturday

C_j - cluster lain

$D(c_i, c_j)$ - jarak centroid dg centroid lain

Selanjutnya Untuk menghitung nilai rasio yang dimiliki oleh masing-masing cluster, digunakan persamaan berikut

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (10)$$

R_{ij} adalah ukuran rasio seberapa baik nilai perbandingan antara cluster ke i dengan cluster ke j . Cluster yang baik adalah cluster dengan kohesi sekecil mungkin dan separasi sebesar mungkin. Persamaan DBI adalah sebagai berikut :

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{j \neq i} (R_{i,j}) \quad (11)$$

Keterangan rumus :

R_{ij} - rasio SSW dan SSB dari dua cluster

K - jumlah cluster terbentuk



- b. Cluster 2 : 55,0875 49,7125
- c. Cluster 3 : 25,72727273 79,36363636
- d. Cluster 4 : 26,30434783 20,91304348
- e. Cluster 5 : 86,53846154 82,12820513

3.1.3 Proses Optimasi Cluster

Hasil clustering yang sudah di dapat lalu di optimasi lagi dengan *Particle Swarm Optimization* agar hasil lebih akurat. Hasil dari clustering yang sudah di optimasi dengan PSO dapat dilihat di **Gambar 6**.

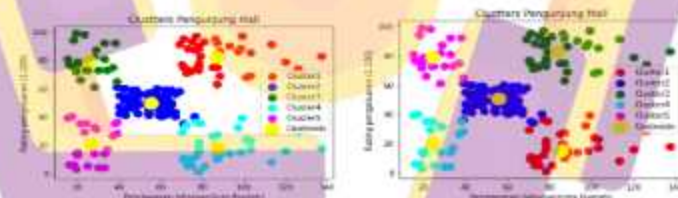


Gambar 6. Hasil Clustering setelah optimasi

Hasil clustering setelah optimasi kemudian sudah di ketahui titik centroidnya. Titik centroidnya sebagai berikut:

- a. Cluster 1 : 86,77240632 15,2726717
- b. Cluster 2 : 55,32666756 50,44306318
- c. Cluster 3 : 84,07620372 83,34551904
- d. Cluster 4 : 24,35252843 16,22639717
- e. Cluster 5 : 23,21183377 77,5881396

Dari hasil clustering yang sudah di optimasi dengan PSO, titik centroid bergeser. Warna dari cluster pun berubah tetapi tidak mempengaruhi hasil hanya penandaan saja. Hasil perbandingan cluster dapat di lihat di **Gambar 7**.



Gambar 7. Perbandingan Cluster sebelum dan sesudah optimasi

Untuk selanjutnya hasil tersebut akan di bandingkan lagi, apakah hasilnya lebih baik sebelum optimasi atau sesudah optimasi di proses evaluasi.

3.2 Evaluasi Cluster

Evaluasi cluster dilakukan dengan menghitung *Davies Bouldin Index*. Yang berarti jika nilai DBI suatu cluster mendekati 0 maka sudah di katakan cluster yang baik. Untuk menghitung DBI melalui 3 proses diantaranya menghitung SSW, SSB, RATIO. Evaluasi di hitung 2 kali, untuk yang pertama menghitung DBI cluster yang belum di optimasi, untuk yang kedua menghitung DBI cluster yang sudah di optimasi. Hasilnya sebagai berikut:



Gambar 8. Sebelum optimasi



Gambar 9. Sesudah optimasi

Angka DBI menandakan hasil sesudah optimasi lebih mendekati 0 di bandingkan dengan hasil sebelum optimasi, hasil tersebut dapat di lihat di **Gambar 8 & 9**. Jadi cluster yang akan di gunakan dan ditentukan karakteristiknya adalah Hasil cluster setelah di optimasi.



3.3 Karakteristik Pelanggan

Berdasarkan hasil clustering yang sudah di dapatkan kita dapat membedakan pelanggan mall menjadi 5 karakteristik.

1. Karakteristik pertama yaitu kelompok berwarna merah menunjukkan bahwa kelompok ini adalah orang-orang berpendapatan tinggi dan mempunyai skor pengeluaran tinggi. Karakteristik pertama ini adalah target yang sangat ideal untuk mall. Mereka adalah aset utama bagi mall dan sumber keuntungan terbesar dari setiap karakter pengunjung.
2. Karakteristik kedua yaitu kelompok berwarna biru menunjukkan bahwa kelompok ini adalah orang-orang yang mempunyai tingkat pendapatan yang tinggi namun skor pengeluaran rendah. Karakteristik kedua ini adalah orang-orang yang unik. Mereka jarang berbelanja namun mempunyai pendapatan yang tinggi. Pelanggan dengan karakteristik kedua ini adalah target bagi mall untuk menawarkan fasilitas dan segala discount yang ada dengan harapan mereka akan tergoda dengan hal tersebut. Hal tersebut adalah strategi marketing yang bagus bagi mall guna menambah pendapatan.
3. Karakteristik ketiga yaitu kelompok berwarna hijau menunjukkan bahwa kelompok ini adalah orang-orang yang mempunyai pendapatan dan skor pengeluaran setara. Kita dapat berasumsi bahwa mereka adalah orang-orang yang tidak selalu membeli suatu produk tetapi memiliki tingkat kemauan yang tinggi untuk berbelanja meskipun terkadang mereka memiliki pendapatan yang kecil. Kelompok orang ini bukanlah kelompok yang memiliki potensi pendapatan tinggi untuk mall. Sebagai pengelola sebisa mungkin menghindari penargetan kelompok orang ini dalam strategi pasar. Namun, mereka masih dapat dipertimbangkan melalui teknik analisis data lain yang dapat meningkatkan tingkat pengeluaran mereka.
4. Karakteristik keempat yaitu kelompok berwarna cyan menunjukkan bahwa kelompok ini adalah orang-orang yang mempunyai pendapatan yang rendah tetapi mempunyai skor pengeluaran yang tinggi. Orang-orang seperti ini adalah tipe orang yang gemar berbelanja dengan asumsi mereka akan menghabiskan uang mereka untuk berbelanja tanpa memikirkan pendapatan yang mereka peroleh. Pengunjung dengan karakteristik keempat ini adalah target yang baik bagi mall karena mereka mudah tergoda dengan fasilitas dan discount yang di berikan oleh pihak mall.
5. Karakteristik kelima yaitu kelompok berwarna magenta menunjukkan bahwa kelompok ini adalah orang-orang yang mempunyai pendapatan yang rendah dan juga skor pengeluaran yang rendah. Hal ini wajar karena mereka akan belanja sesuai kebutuhan dan juga memikirkan pendapatannya. Bahkan apa yang mereka lakukan mungkin merupakan pilihan yang bijaksana dan baik berdasarkan kondisi mereka. Karakteristik keempat ini sebaiknya ditempatkan di prioritas terendah karena mereka cenderung berhemat dan tidak tertarik dengan adanya fasilitas dan discount yang di berikan oleh pihak mall.

4. KESIMPULAN

Berdasarkan hasil clustering dengan menggunakan metode *K-Means* yang di optimasi dengan *Particle Swarm Optimization* dan telah di evaluasi menggunakan *Davies Bouldin Index* maka dapat di simpulkan bahwa hasil clustering menghasilkan 5 cluster. Berdasarkan evaluasinya 5 cluster yang sudah dioptimasi lebih baik di bandingkan dengan yang belum dioptimasi walaupun kinerja algoritmanya lebih lambat. Hasil clustering tersebut di analisis dan mendapatkan 5 karakteristik. Karakteristik tersebut dapat digunakan untuk membantu pihak manager serta seluruh kreatif marketing untuk dapat mensiasati strategi guna meningkatkan pendapatan mall. Penelitian ini membuktikan bahwa penggunaan metode *Particle Swarm Optimization* untuk optimasi cluster yang di hasilkan oleh metode *K-Means* terbukti tepat setelah di evaluasi dengan *Davies Bouldin Index*. Data yang sudah di peroleh tidak menutup kemungkinan sewaktu waktu akan berubah seiring berkembangnya zaman. Terlebih lagi sifat manusia yang dapat berubah ubah sesuai dengan lingkungannya.

REFERENCES

- [1] M. G. Pradana, "Memaksimalkan Strategi Peningkatan Segmentasi Pelanggan Mall Menggunakan K-means Clustering," vol. 2, no. 1, 2021.
- [2] S. Royal, "PEMETAAN POTENSI PELANGGAN SEBAGAI STRATEGI PROMOSI PAKAIAN MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING," 2020, doi: 10.33480/juk.v6i1.1414.
- [3] J. Dai, P. Byrnes, and M. Vasarhelyi, "Are Customers Offered Appropriate Discounts? An Exploratory Study of Using Clustering Techniques in Internal Auditing," *Rutgers Stud. Account. Anal. Audit Anal. Financ. Ind.*, pp. 59–69, 2019, doi: 10.1108/978-1-78743-085-320191005.
- [4] B. Komponen and P. Indeks, "1, 2, 3," pp. 2–3, 2018.
- [5] A. L. Ballardini, "A tutorial on Particle Swarm Optimization Clustering," 2018, [Online]. Available: <http://arxiv.org/abs/1809.01942>.
- [6] L. Kouhalvandi, O. Ceylan, and S. Ozoguz, "Automated Deep Neural Learning-Based Optimization for High Performance High Power Amplifier Designs," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 67, no. 12, pp. 4420–4433, 2020, doi: 10.1109/TCSI.2020.3008947.
- [7] A. Akmal, "Predicting Dropout on E-learning Using Machine Learning," *J. Appl. Data Sci.*, vol. 1, no. 1, pp. 29–34, 2020, doi: 10.47738/jads.v1i1.9.



- [8] S. Hidayat, M. Matsuoka, S. Baja, and D. A. Rampisela, "Object-based image analysis for sago palm classification: The most important features from high-resolution satellite imagery," *Remote Sens.*, vol. 10, no. 8, 2018, doi: 10.3390/RS10081319.
- [9] L. L. Rego, N. A. Morgan, and C. Fornell, "Reexamining the market share-customer satisfaction relationship," *J. Mark.*, vol. 77, no. 5, pp. 1–20, 2013, doi: 10.1509/jm.09.0363.
- [10] I. Markoulidakis, I. Rallis, I. Georgoulas, G. Kopsiaftis, A. Doulamis, and N. Doulamis, "A Machine Learning Based Classification Method for Customer Experience Survey Analysis," *Technologies*, vol. 8, no. 4, p. 76, 2020, doi: 10.3390/technologies8040076.
- [11] O. Dogan, C. Fernandez-Llata, and B. Oztaysi, *Process mining application for analysis of customer's different visits in a shopping mall*, vol. 1029. Springer International Publishing, 2020, doi: 10.1007/978-3-030-23756-1_20.
- [12] I. P. Sari, E. Fatkhayah, and J. Triyono, "Data Mining Menggunakan Algoritma K-Means Clustering Untuk Pengelompokan Produk Yang Paling Tidak Laku Terjual Pada Koperasi Mahasiswa Universitas Negeri Yogyakarta (Kopma Uny)," *J. Scr.*, vol. 6, no. 1, pp. 55–61, 2018, [Online]. Available: <https://ejournal.akprind.ac.id/index.php/script/article/view/620>
- [13] D. Triyansyah and D. Fitriana, "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing," *J. Telekomun. dan Komput.*, vol. 8, no. 3, p. 163, 2018, doi: 10.22441/incomtech.v8i3.4174.
- [14] L. A. Setiyo and I. F. B. Andoro, "PENERAPAN ALGORITMA K-MEANS UNTUK (Studi Kasus : Universitas Katolik Widya Mandala Kampus Kota Madiun)," pp. 1–8, 2021.
- [15] G. K. Patel, V. K. Dabhi, and H. B. Prajapati, "Clustering Using a Combination of Particle Swarm Optimization and K-means," *J. Intell. Syst.*, vol. 26, no. 3, pp. 457–469, 2017, doi: 10.1515/jisys-2015-0099.
- [16] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Proporsi Berdasarkan Potensi Desa," *J. Sains dan Manaj.*, vol. 9, no. 1, pp. 95–100, 2021, [Online]. Available: <https://ejournal.hsi.ac.id/ejurnal/index.php/evolusi/article/view/10428/4839>
- [17] M. D. Kartikasari, "Self-Organizing Map Menggunakan Davies-Bouldin Index dalam Pengelompokan Wilayah Indonesia Berdasarkan Konsumsi Pangan," *Sambura J. Math.*, vol. 3, no. 2, pp. 187–196, 2021, doi: 10.34312/jjom.v3i2.10942.
- [18] Dataset Kaggle by Bakri Siregar "Penganjung Mall" <https://www.kaggle.com/datasets/bakrisiregar/datasetpanganjunggmall>

