

# PERBANDINGAN ALGORITMA K-MEANS DENGAN K-MEDOIDS PADA PENGELOMPOKAN ARMADA KENDARAAN TRUK BERDASARKAN PRODUKTIVITAS

Aceng Supriyadi<sup>1)</sup>, Agung Triayudi \*<sup>2)</sup>, Ira Diana Sholihati<sup>3)</sup>

<sup>1, 2, 3)</sup> Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional

Jl. Sawo Manila, Pasar Minggu, Jakarta Selatan, Daerah Khusus Ibukota Jakarta, 12520

e-mail: [acengsupriyadi137@gmail.com](mailto:acengsupriyadi137@gmail.com)<sup>1)</sup>, [agungtriayudi@civitas.unas.ac.id](mailto:agungtriayudi@civitas.unas.ac.id)<sup>2)</sup>, [iradiana2803@gmail.com](mailto:iradiana2803@gmail.com)<sup>3)</sup>

\*Penulis Korespondensi

## ABSTRAK

Armada kendaraan truk merupakan salah satu aset utama dalam bisnis di bidang jasa transportasi. Evaluasi kinerja armada secara cepat dan akurat diperlukan untuk mendukung tercapainya produktivitas armada secara maksimal sehingga target perusahaan dapat mudah tercapai. Proses evaluasi kinerja armada yang masih dilakukan secara manual menyebabkan rumitnya proses pengolahan dan kurang akuratnya hasil evaluasi yang diperoleh, sehingga diperlukan suatu teknik pengolahan data secara cepat dan lebih akurat salah satunya dengan namun menerapkan teknik data mining menggunakan metode clustering. Metode Clustering akan digunakan untuk mengelompokkan setiap armada kendaraan berdasarkan produktivitas kerjanya. Pada Penelitian ini memkomparasikan penerapan Algoritma K-Means dan K-Medoids, yang kemudian dilakukan uji validasi terhadap hasil cluster yang terbentuk. Davies Bouldin Index sebagai metode dalam analisis kluster menghasilkan nilai validitas sebesar 0,67 untuk K-Means clustering dan 1,78 untuk K-Medoids. Berdasarkan nilai validitas yang dihasilkan Algoritma K-Means dipilih untuk diimplementasikan pada pembuatan aplikasi clustering armada kendaraan berbasis web karena paling relevan dengan nilai validitas DBI yang lebih rendah dari pada K-Medoids. Pengujian yang telah dilakukan terhadap hasil clustering pada aplikasi web didapatkan persentase kesesuaian sebesar 97 % baik dengan tool Rapidminer maupun dengan perhitungan secara manual.

**Kata Kunci:** Armada Kendaraan, Clustering, K-Means, K-Medoids.

## ABSTRACT

Truck fleet is one of main business assets in transportation services. A fast and accurate evaluation of fleet performance is required to support the achievement for maximum productivity; thus, the company can be easily achieving its goal. The manual process of fleet management creates a complex management system and the evaluation results obtained are less accurate, so it requires a faster and more accurate data processing technique, one of which is by applying data mining techniques using clustering method. Clustering method will be used in clustering each vehicles fleet based on its productivity and performance. This study differentiates the application of K-Means and K-Medoids Algorithm, which then a validation testing is carried out on clustering results. Davies Bouldin Index was used in cluster analysis produces a validity value of 0.67 for K-Means clustering and 1.78 for K-Medoids. Based on the validity value produced, K-Means algorithm was chosen to be implemented in making a web-based vehicle fleet clustering application since it is most relevant to the DBI validity value which is lower than K-Medoids. Tests that have been carried out on clustering results on web applications obtained the percentage of conformity of 97% both with the Rapidminer tool and manual calculations.

**Keywords:** Clustering, Fleet of Vehicle, K-Means, K-Medoids.

## I. PENDAHULUAN

Sebagai salah satu perusahaan penyedia jasa transportasi khususnya untuk jenis kendaraan truk yang kegiatan bisnisnya ialah melakukan pengiriman bahan baku material pembuatan semen dari lokasi muat menuju lokasi pembongkaran, PT.Lintas Harapan Mandiri senantiasa berupaya melakukan proses evaluasi terhadap kinerja armadanya. Namun proses evaluasi terhadap kinerja armada yang sudah berjalan masih belum sepenuhnya dilakukan secara maksimal serta belum mencapai tingkat akurasi terbaik dalam melakukan penentuan unit armada berdasarkan produktivitas kerjanya dikarenakan proses perhitungan masih bersifat manual dengan teknik sederhana yaitu hanya dengan melakukan pemeringkatan terhadap unit armada berdasarkan pendapatan setiap armada dikurangi dengan biaya perbaikan masing masing armada saja. Selain memerlukan waktu yang lama, proses pengolahan secara manual juga rumit serta butuh ketelitian tinggi dalam proses pengolahan data nya [1].

Berdasarkan permasalahan yang dihadapi maka sudah seharusnya ada pemanfaatan teknik pengelompokan armada berdasarkan kerjanya menggunakan teknik pengolahan data yang tepat dengan menerapkan teknologi informasi yang saat ini berkembang dengan sangat pesat sekali. Salah satu teknik perhitungan yang dapat

diterapkan untuk menangani permasalahan tersebut diatas adalah dengan memanfaatkan teknik data mining dengan menggunakan metode Clustering. Data mining adalah teknik untuk menggali, menemukan dan mencari informasi penting atau wawasan pengetahuan (*insight*) tersembunyi dari database besar [2]. Salah satu tujuan data mining adalah untuk mencari pola atau *trend* tertentu yang hendak di capai sehingga dapat digunakan untuk memprediksi dan mendukung dalam hal pengambilan keputusan di kemudian waktu [3]. Metode clustering adalah suatu teknik memisahkan sekumpulan titik data menjadi kelompok atau cluster tertentu. Titik data yang terdapat kemiripan karakteristik yang satu sama dengan yang lain berada dalam satu klaster dan tentunya berbeda dengan data yang ada pada klaster lain [4].

Penelitian ini bertujuan untuk mendapatkan pemetaan dan gambaran secara jelas dan akurat dari setiap unit armada truk yang beroperasi terkait produktivitas kerjanya dengan melakukan teknik pengolahan data secara cepat, tepat dan akurat dalam rangka mengantisipasi terhadap ketatnya persaingan dalam bisnis serta untuk menjaga loyalitas terhadap pelanggan, supaya perusahaan dapat melakukan pengambilan keputusan dengan tepat dalam menentukan dan mengatur strategi bisnis kedepannya [5]. Adapun manfaat yang diperoleh adalah dapat mengetahui kinerja setiap armada truk yang dikategorikan sebagai armada dengan produktivitas kerja tinggi, sedang dan terutama produktivitas rendah secara cepat dan akurat. Untuk teknik clustering pada penelitian ini menggunakan perbandingan penerapan antara algoritma K-Means Clustering dengan K-medoids, yang mana hasil perhitungan terbaik dan relevan antara kedua algoritma tersebut akan di pakai pada perancangan aplikasi clustering berbasis web.

*Davies Bouldin Index (DBI)* digunakan untuk mengevaluasi terhadap hasil klaster yang sudah terbentuk. DBI merupakan teknik mengukur tingkat validitas secara internal yang dilakukan dengan cara mengukur seberapa baik tingkat klasterisasi yang di hasilkan dengan menghitung jumlah banyaknya kriteria atau fitur lainnya dari dataset yang ada [6]. Tujuan dari pendekatan dalam pengukuran ini adalah untuk memaksimalkan jarak antara objek klaster yang satu dengan yang objek data pada klaster lainnya dan juga secara bersamaan meminimalkan jarak antara objek dalam sebuah klaster. Hasil Pengklasteran dengan nilai DBI minimum dinilai sebagai Pengklasteran dengan jumlah klaster yang terbaik [7]. Maka berdasarkan nilai validasi dengan DBI terendah antara algoritma K-Means dan K-Medoids akan dipilih dan digunakan dalam pembuatan aplikasi clustering berbasis website nantinya.

Beberapa penelitian terkait dengan penerapan algoritma K-means maupun K-medoids terdahulu yang telah banyak dilakukan antara lain: Penelitian sebelumnya terkait perbandingan penerapan K-means dan K-Medoids dalam hal Klasterisasi Data Bongkar Muat Provinsi Riau [8]. Diperoleh kesimpulan bahwa dalam pengolahannya, algoritma K-Means lebih cepat dan menghasilkan nilai DBI pada K-Means yang dinilai lebih rendah baik dengan jumlah klaster K=3 dan K=10. Begitu juga pada penelitian yang telah dilakukan di puskesmas DIY untuk melakukan klasterisasi puskesmas tersebut berdasarkan tingkat kelayakannya dapat disimpulkan bahwa algoritma K-Means memiliki hasil klasterisasi yang jauh lebih baik pula, hal ini mengacu terhadap nilai rasio simpangan baku yang dihasilkan pada K-means relative lebih kecil dibanding K-Medoids yaitu 54,69% untuk K-Means dan K medoids sebesar 55,3% [9].

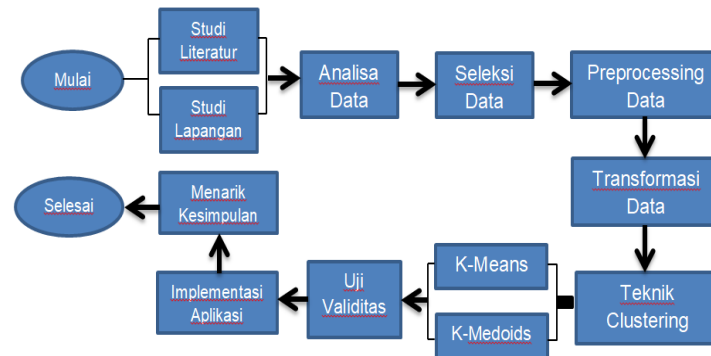
Algoritma K-Means juga dinilai lebih baik pada Penelitian sebelumnya yaitu pada penelitian dengan jurnal yang berjudul segmentasi pelanggan pada Bank XYZ [10]. Hasil penelitian ini menunjukkan bahwa metode K-Means Clustering mengungguli metode K-Medoids berdasarkan jarak intra cluster (AWC). Sementara berdasarkan indeks Davies Bouldin, K-Means Clustering memiliki kinerja sedikit lebih baik daripada K-Medoids. Penelitian lainnya juga yang pernah dilakukan terkait perbandingan algoritma K-means dan Agglomerative Hierarchical Clustering [11]. Penelitian ini dapat menghasilkan kesimpulan bahwa k-means lebih cocok untuk set data yang lebih besar karena waktu eksekusi yang lebih rendah dan tingkat perubahan yang lebih rendah dalam pemanfaatan memori.

Penelitian lainnya yang telah dilakukan yaitu pada jurnal yang berjudul komparasi antara K-means, K-medoids dan ISODATA Clustering pada layanan mesin pencari berbasis Cloud [12]. Hasil penelitian tersebut metode Iso-data jauh mengungguli K-Means dan K-Medoids, namun K-Means lebih cepat dan tinggi untuk tingkat performancenya dibanding K-Medoids walaupun selisihnya tidak terlalu signifikan terlihat. Pada penelitian terdahulu lainnya Implementasi algoritma K-Medoids justru dinilai lebih baik dan relevan daripada K-Means seperti pada penelitian Pengelompokan Wilayah Sebaran Cacat pada Anak [13]. Berdasarkan nilai validitas memakai metode *silhouette coefisien* pada K-Medoids menunjukan nilai yang jauh lebih besar.

Penelitian lainnya juga telah dilakukan terkait perbandingan antara algoritma K-Means dengan algoritma lainnya dalam metode clustering yaitu dan DBSCAN pada segmentasi rumah kost mahasiswa kelurahan Tembalang Semarang. Hasil penelitian dapat disimpulkan bahwa metode K-Means lebih baik dari metode DBSCAN berdasarkan hasil perhitungan indeks *silhouette coefisien* dengan nilai 0,463 untuk K-Means dan 0,281 dengan metode DBSCAN [14].

## II. METODE PENELITIAN

Prosedur penelitian merupakan salah satu proses penting. Pada prosedur penelitian ini tahapan demi tahapan harus dilakukan secara terperinci, terstruktur dan secara sistematis, supaya target penelitian tercapai sesuai yang diharapkan. Adapun untuk kerangka penelitian ini seperti pada gambar 1.



Gambar 1. Tahapan dalam Penelitian

Berdasarkan gambar 1 dapat diuraikan bahwa pada penelitian ini mencakup delapan tahapan yang dilakukan secara berurutan. Dimulai dari tahap pengumpulan data yaitu mencakup studi literatur dan studi lapangan, kemudian analisa data, seleksi data, preprocessing data, proses pengolahan atau teknik clustering menggunakan perbandingan algoritma K-Means dengan K-Medoids, kemudian uji validasi cluster yang dihasilkan dari kedua algoritma tersebut. Selanjutnya implementasi aplikasi clustering berbasis web menggunakan salah satu algoritma yang nilai validasi nya terendah dan paling relevan berdasarkan hasil evaluasi hasil cluster sebelumnya dan terakhir adalah menarik kesimpulan.

### A. Pengumpulan Data

Pada proses pengumpulan data dilakukan kegiatan studi literatur dengan cara mengumpulkan pengetahuan atau referensi dari berbagai macam sumber literatur seperti referensi dari jurnal ilmiah dan karya ilmiah yang berkaitan dengan topik yang akan diteliti [1]. Kemudian dilanjutkan dengan pengumpulan beberapa sample data yang dibutuhkan dan akan digunakan dalam proses pengolahan data mining. Sumber data yang digunakan berasal dari data perusahaan PT Lintas Harapan Mandiri yang terdiri dari 3 data sampel yang dipilih yaitu meliputi data armada truk terpilih, data transaksi pengiriman dan operasional armada, serta data perbaikan armada. Data-data tersebut masih berupa data keseluruhan yang belum dilakukan proses seleksi menjadi dataset yang siap diolah dalam penelitian.

### B. Analisis dan Pengolahan Data

Tahapan berikutnya adalah melakukan analisis data sampai perhitungan dengan memakai tahapan KDD (*Knowledge Discovery in Database*). Tahapan KDD merupakan proses membongkar, menganalisis, dan memecah sekumpulan data yang berukuran besar sehingga menghasilkan informasi yang bersifat penting atau pengetahuan yang berguna. Tahapan KDD dimulai dari mulai analisis data atau *preprocessing*, pengolahan data sampai menghasilkan pengetahuan (*knowledge*) yang berupa informasi penting dan berguna serta tahap evaluasi.

#### 1) Seleksi Data

Tahapan KDD pertama kali pada penelitian ini adalah tahapan proses penyeleksian data ini, dari sekumpulan data yang ada kemudian dilakukan proses pemilihan dan penentuan jenis dan kategori data yang akan digunakan dan dibutuhkan pada penelitian. Karena tidak seluruh data akan dipakai pada proses perhitungan. Hanya data armada, transaksi pelayanan pengiriman dan data perbaikan armada kendaraan yang akan digunakan.

#### 2) Preprocessing Data

Pada tahap preprocessing ini dilakukan proses data cleaning. Data cleaning adalah menghapus atribut yang tidak akan dipakai dan kolom isian yang tidak bernilai (*null*) yang tidak diperlukan pada proses perhitungan nantinya, sehingga data yang akan diolah benar-benar data relevan [15]. Kemudian setelah tahapan cleaning seluruh sub-dataset, selanjutnya dilakukan proses integration data atau penggabungan beberapa subdata set sehingga menjadi satu dataset utuh yang siap dilakukan proses perhitungan clustering

### 3) *Transformasi Data*

Setelah tahapan preprocessing dilakukan kemudian dilanjutkan pada Proses transformasi data. Transformasi data yaitu proses perubahan beberapa subdataset menjadi satu format yang sama untuk mempermudah proses data mining [15]. Selain itu juga dilakukan perubahan pada nama atribut menjadi kode tertentu untuk mempermudah dalam penyajian dan pengolahan datanya.

### 4) *Data Mining*

Data Mining adalah teknik atau proses menemukan pola atau penggalian informasi unik dan menarik dari dalam dataset yang terpilih dengan menggunakan metode atau algoritma tertentu sesuai dengan maksud atau tujuan dari proses KDD secara keseluruhan [16]. Dalam Data Mining terdapat tiga teknik yang umum dipakai yaitu regresi, klasifikasi dan klustering. Metode yang digunakan pada penelitian ini sendiri adalah metode Clustering dengan menggunakan perbandingan antara algoritma K-Means dan K-Medoids.

#### a) Perhitungan Algoritma K-Means Clustering

K-means clustering merupakan metode dalam machine learning yang tanpa pengawasan (unsupervised learning) yang secara efektif digunakan untuk memecah data set yang dikelompokkan ke dalam grup k atau kluster k, di mana k mewakili jumlah kluster. Dalam pengelompokan k-means, setiap kluster diwakili oleh titik pusat (*centroid*) atau rata-rata yang dan dihitung sebagai nilai rata-rata data dalam kluster tersebut [17]. Algoritma k-means merupakan metode clustering yang paling sederhana dan paling umum. Ini karena K-means memiliki kemampuan untuk mengelompokkan data dalam jumlah besar dengan waktu komputasi yang relatif cepat dan efisien [18].

#### b) Perhitungan Algoritma K-Medoids Clustering

Algoritma K-Medoids atau istilah lainnya Partitioning Around Medoids (PAM), merupakan jenis lain dari metode K-Means. Perhitungan pada K-Medoids tidak mengacu terhadap nilai rata-rata (mean) seluruh data pada setiap cluster, tujuannya adalah untuk mengurangi *outlier* atau sensitivitas dari kluster yang dihasilkan pada dataset [19]. Algoritma K-Means dan K-Medoid keduanya bekerja dengan cara meminimalkan jarak antara titik data yang dilabeli untuk berada dalam satu kluster dan titik data yang ditetapkan sebagai pusat kluster tersebut. Pada algoritma K-Medoid memilih titik data sebagai pusat cluster (medoid) dilakukan secara acak pada titik data tiap tiap cluster bukan berdasarkan nilai rata-ratanya [20].

### 5) *Interpretation/Evaluasi*

Pada tahap ini dilakukan penerjemahan dan analisa pola-pola atau pengetahuan (knowledge) tertentu yang dihasilkan dari perhitungan dalam data mining. Proses Evaluasi ini dilakukan dengan cara pengujian apakah pola, informasi atau knowledge yang diperoleh sesuai dengan fakta atau hipotesa sebelumnya atau justru malah bertentangan. Informasi yang diperoleh kemudian dipresentasikan ke dalam bentuk visualisasi..

#### C. *Aplikasi Clustering berbasis web*

Aplikasi web clustering menggunakan salah satu algoritma antara K-Means dan k-Medoids. hasil evaluasi yang paling terbaik dan relevan akan digunakan pada aplikasi web clustering ini. Hasil clustering dari aplikasi web ini kemudian dikomparasikan dengan hasil yang didapat melalui penghitungan secara manual dan dengan Tools Rapidminer . Sehingga keakuratan dan persentase kesesuaian hasil clustering dapat diketahui, Harapannya Aplikasi web clustering yang telah dibuat nantinya dapat dipergunakan secara berkelanjutan untuk mengelompokkan armada kendaraan truk PT Lintas Harapan Mandiri, sehingga membantu dalam proses pengambilan keputusan.

## III. HASIL DAN PEMBAHASAN

### A. *Analisis Data*

Sebelum pengolahan data dilaksanakan terlebih dahulu dilakukan tahapan analisis data yang terdiri dari proses penyeleksian data terhadap sejumlah dataset yang sudah terkumpul dengan cara melakukan pemilahan data sesuai kebutuhan penelitian. Data yang dipilih bersumber dari PT Lintas Harapan Mandiri yang terdiri dari data armada, data histori operasional armada dan data histori perbaikan armada selama satu tahun yaitu periode 1 januari 2020 sampai 31 desember 2020 yang masing -masing subdataset berformat excel. Kemudian berlanjut ke tahap Preprocessing data yang terdiri dari tahapan cleaning data yaitu memilih atribut-atribut dan isian data yang akan digunakan dan membuang atribut beserta data yang tidak diperlukan dalam pengolahan kemudian ketiga dataset hasil proses cleaning tersebut dilakukan penggabungan atau integration data menjadi satu dataset utuh. Langkah selanjutnya dilakukan proses transformasi data sesuai keperluan perhitungan. Pada tahapan ini dilakukan

pengubahan beberapa nama atribut dan isian data ke dalam kode numerik tertentu untuk memudahkan dalam proses perhitungan Pada tabel 1 berikut disajikan atribut- atribut yang akan digunakan yang merupakan hasil integrasi dari ketiga subdataset sebelumnya dan sudah dilakukan inisialisasi agar memudahkan perhitungan.

TABEL 1  
ATRIBUT PADA DATASET

Nama atribut data	Inisialisasi Unit
Nomor Polisi Kendaraan	No.Unit
Jumlah Biaya Pengiriman	X01
Jumlah Tonase (beban muatan)	X02
Jumlah Omset	X03
Jumlah Ritase	X04
Jumlah Jarak Tempuh	X05.
Tahun Produksi Kendaraan	X06
Banyaknya Perbaikan	X07
Jumlah Biaya Perbaikan	X08

Adapun hasil akhir proses analisis data ini yaitu diperoleh satu dataset utuh. Dataset tersebut terdiri dari 66 data armada truk dengan 8 atribut. Berikut tabel dataset sebagai data sampel penelitian seperti terlihat pada tabel 2.

TABEL 2.  
DATA SAMPEL PENELITIAN

Urutan Data	No.Unit	Biaya Pengiriman	Jumlah Perbaikan	Biaya Perbaikan
Data ke-01	B01	318,738,100	20	10,053,847
Data ke-02	B02	313,530,233	27	24,871,440
Data ke-03	B03	365,320,500	22	14,957,739
.....	.....	.....	.....	.....
Data ke-20	D21	313,848,000	47	38,820,958
Data ke-21	D22	243,432,000	44	24,450,647
Data ke-22	D23	324,081,000	54	44,646,414
.....	.....	.....	.....	.....
Data ke-64	L05	168,263,500	8	1,577,230
Data ke-65	L06	291,375,000	13	4,634,081
Data ke-66	L07	255,660,000	8	2,065,051

## B. Pengolahan Data

### a) Perhitungan Algoritma K-Means

Perhitungan clustering pertama yang dilakukan adalah dengan menerapkan algoritma K-Means secara manual. Adapun Langkah langkah nya adalah .

1) Menentukan banyaknya cluster yang akan di bentuk

Jumlah cluster yang akan dibentuk adalah sebanyak 3 buah cluster ( $k=3$ ), sehingga terbentuk cluster dengan produktivitas kinerja tinggi, sedang dan rendah

2) Menentukan pusat cluster (*centroid*) awal secara acak

Lakukan dengan memilih 3 data secara bebas dari data sampel yang berjumlah 66 objek data sesuai dengan jumlah klaster yang akan di bentuk yaitu 3 data. Penentuan *centroid* awal untuk *centroid* 1, 2, dan 3 berturut turut adalah urutan data ke-02 Data ke-39 dan Data ke 64 seperti terlihat pada tabel 3 berikut

TABEL 3  
INISIALISASI *CENTROID* AWAL K-MEANS

Nama Atribut/Variabel	C1(Centorid 1)	C2( <i>Centroid</i> 2)	C3( <i>Centroid</i> 3)
NoUnit	B02	D50	L05
Biaya Pengiriman	313,530,233	327,358,000	168,263,500
Jumlah Tonase	5,980	10,769	1,597
Jumlah Omzet	723,596,033	688,007,696	373,202,564
Jumlah Ritase	224	322	140
Jarak Tempuh	45,436	71,495	40,226
Tahun Produksi	2018	2017	2019
Jumlah Perbaikan	27	28	8
Biaya Perbaikan	24,871,440	18,579,655	1,577,230

3) Kemudian hitung nilai jarak dari setiap objek data terhadap masing-masing pusat cluster (*centroid*) Persamaan yang digunakan adalah jarak matriks euclidean (*euclidean distance*) seperti pada persamaan (1)

$$d_{ij} = \sqrt{(x_i - y_j)^2 + \dots + (x_i - y_j)^2} \quad (1)$$

Untuk  $d_{ij}$  adalah nilai jarak terkecil antara data dengan *centroid*. Untuk  $x_i$  adalah nilai data urutan ke- $i$  pada dataset. Dan  $y_i$  adalah Nilai *centroid* urutan ke- $i$ . Dan misal untuk menghitung objek data ke-1 dengan *centroid* ke-1 adalah sebagai berikut:

$$d_{1,1} = \sqrt{(318.738.100 - 313.530.233)^2 + \dots + (10.053.847 - 24871440)^2}$$

$$d_{1,1} = 48.161.045$$

Untuk perhitungan objek data ke-1 dengan *centroid* ke-2 adalah sebagai berikut

$$d_{1,2} = \sqrt{(318.738.100 - 327.358.000)^2 + \dots + (10.053.847 - 18.579.655)^2}$$

$$d_{1,2} = 82.017.443$$

dan untuk perhitungan objek data ke 1 dengan *centroid* ke-3 adalah

$$d_{1,3} = \sqrt{(318.738.100 - 168.26.500)^2 + \dots + (10.053.847 - 1.577.230)^2}$$

$$d_{1,3} = 423.636.999$$

- 4) Kemudian lanjutkan perhitungan pada keseluruhan objek data sampai data terakhir dengan menggunakan rumus *euclidean distance* seperti cara perhitungan pada data ke-1 diatas lakukan perhitungan sampai diperoleh nilai jarak setiap objek data dengan masing masing *centroid*.
- 5) Setelah nilai jarak masing-masing objek data dengan masing-masing pusat cluster (*centroid*) diperoleh kemudian carilah nilai jarak euclidean terendah atau terdekat diantara ketiga hasil perhitungan jarak antara *centroid* dengan dataset tersebut (nilai jarak terendah dari C1,C2, dan C3). Berdasarkan nilai terendah itulah objek data akan terklasterisasi ke dalam cluster 1, cluster 2 ,atau cluster 3. Sehingga dapat disimpulkan hasil clustering pada iterasi pertama, jumlah data tiap-tiap cluster adalah seperti terlihat pada tabel 4.

TABEL 4  
JUMLAH ANGGOTA TIAP CLUSTER ITERASI PERTAMA

Cluster	Jumlah Anggota
Cluster ke-01	35
Cluster ke-02	22
Cluster ke-03	9

Berdasarkan tabel 4 diatas hasil perhitungan pada iterasi ke-1 terlihat bahwa cluster 1 adalah cluster dengan tingkat produktivitas tinggi dengan jumlah armada sebanyak 35 unit, untuk cluster 2 adalah kelompok armada dengan produktivitas sedang dengan anggota 22 armada dan pada cluster 3 adalah kelompok armada dengan produktivitas rendah dengan 9 unit armada. Kemudian Perhitungan dilanjutkan pada iterasi kedua dan diawali dengan menentukan kembali nilai *centroid* baru. Penentuan *centroid* baru pada iterasi ke-2 ini dilakukan dengan cara menghitung nilai rata-rata (means) dari masing-masing nilai data pada tiap atribut seluruh anggota dari setiap cluster iterasi ke-1 sebelumnya. Setelah nilai *centroid* baru diperoleh, kemudian ulangi langkah perhitungan ke-3 dan ke 4 seperti pada iterasi pertama sebelumnya. Maka diperoleh data anggota hasil clustering setiap cluster pada iterasi ke-2 seperti terlihat pada tabel 5.

TABEL 5.  
JUMLAH ANGGOTA TIAP CLUSTER PADA ITERASI KEDUA

Cluster	Jumlah Anggota
Cluster ke-01	35
Cluster ke-02	21
Cluster ke-03	10

Berdasarkan tabel 5 diatas dapat terlihat bahwa pada iterasi ke-2 telah terjadi perpindahan atau perubahan anggota cluster dari hasil clustering pada proses iterasi pertama sebelumnya. Terlihat bahwa pada cluster 2 terdapat satu anggota yang berpindah cluster menjadi anggota cluster ke-3. Kemudian perhitungan K-Means di lanjutkan pada iterasi ke-3 dengan langkah-langkah yang sama seperti pada iterasi ke-2 sebelumnya. Perhitungan ini dilakukan untuk memastikan masih terjadi perpindahan anggota cluster kembali atau tidak. Adapun hasil perhitungan K-Means clustering pada iterasi ke-3 diperoleh hasil anggota klaster seperti terlihat pada tabel.6.

TABEL 6  
JUMLAH ANGGOTA TIAP CLUSTER PADA ITERASI KETIGA

Cluster	Jumlah Anggota
Cluster ke-01	35
Cluster ke-02	20
Cluster ke-03	11

Tabel 6 diatas menunjukkan bahwa hasil clustering pada iterasi ke-3 masih terjadi perpindahan anggota pada tiap cluster. Maka perhitungan dilanjutkan kembali pada iterasi ke-4 dengan langkah dan proses seperti pada iterasi ke-3 dan ke-2 sebelumnya. Berdasarkan hasil klasterisasi yang diperoleh pada iterasi ke-4, maka proses perhitungan pun dihentikan sampai pada iterasi ke-4 saja, karena hasil clustering adalah tetap dan tidak mengalami perubahan atau perpindahan anggota pada tiap-tiap clusternya. Sehingga susunan anggota pun masih sama seperti hasil pada iterasi ke-3 sebelumnya. Maka hasil akhir pada perhitungan clustering dengan algoritma K-Means secara manual adalah hasil akhir perhitungan yang didapatkan pada iterasi ke-3 dengan data anggota tiap –tiap cluster seperti pada tabel 7.

TABEL 7  
HASIL AKHIR KLASTERISASI PADA K-MEANS

Cluster	Jumlah Anggota	Data Anggota Cluster
Cluster ke-01	35	Selain data armada pada cluster 2 dan cluster 3 D21,D12,D15,D16,D23,D24,D25,D27,D28,D30
Cluster ke-02	20	D35,D36,D37,D38,D50,D20,D59,D60,D61,L07
Cluster ke-03	11	B06,D22,D26,D29,D31,D32,D33,D39,D58,L03,L05

#### b) Perhitungan Algoritma K-Medoids

Proses perhitungan menggunakan algoritma ke dua yaitu K-Medoids clustering adapun tahapan dan hasil pengolahan data dengan K-Medoids proses perhitungan secara sederhananya adalah sebagai berikut [21]:

- 1) Inisialisasi medoid awal dengan cara memilih secara acak sesuai dengan jumlah cluster yang akan dibentuk yaitu 3 klaster medoid awal yang dipilih adalah data ke-64,data 65,dan data ke 66
- 2) Kemudian Hitung pula matriks jarak menggunakan *euclidean distance* antara medoids dengan setiap objek data, sama hal nya seperti perhitungan mencari nilai jarak terdekat pada K-Means
- 3) Tentukan nilai jarak terendah atau terkecil diantara ke tiga nilai jarak yang diperoleh
- 4) Kemudian hitunglah total cost jarak terdekat keseluruhan jarak objek data ke *medoids* . Adapun total cost jarak terdekat iterasi pertama dan hasil klasterisasi yang terbentuk seperti pada tabel 8.

TABEL 8  
JUMLAH TIAP KLASTER PADA ITERASI PERTAMA K-MEDOIDS

Nama Cluster	Jumlah Anggota	Total cost jarak terdekat
Cluster 01	7	
Cluster 02	14	
Cluster 03	45	
Total	66	4.881.429.773

- 5) Lakukan perhitungan pada iterasi berikutnya diawali dengan memilih medoid baru atau pusat cluster pada masing-masing cluster yang terbentuk hasil iterasi sebelumnya dan lakukan secara acak atau bebas dalam memilih kandidat medoid baru.
- 6) Ulangi kembali langkah ke-2 sampai dengan 4 dan berikut hasil anggota klasterisasi yang terbentuk beserta total cost jarak terdekat pada iterasi ke-2 seperti pada tabel 9

TABEL 9  
JUMLAH TIAP KLASTER ITERASI KEDUA K-MEDOIDS

Nama cluster	Jumlah anggota	Total cost jarak terdekat
Cluster 01	7	
Cluster 02	13	
Cluster 03	46	
Total	66	4.942.128.062

- 7) Hitunglah nilai total simpangan (S) dengan cara mengurangkan nilai total jarak hasil perhitungan klaster baru dengan total jarak klaster lama. Jika didapatkan nilai simpangan kurang dari nol ( $S < 0$ ), maka lakukan iterasi perhitungan berikutnya.

- 8) Ulangi langkah ke 5 sebelumnya diatas sampai dengan langkah ke-7 hingga nilai simpangan lebih dari nol ( $S > 0$ ) dengan cara mengurangi total jarak baru dengan total jarak lama ,sehingga hasil clustering adalah berupa data data yang telah terclusterisasikan pada iterasi kedua sebelumnya atau pada iterasi yang lama. Adapun hasil perhitungan total simpangan ( $S$ ) nilai total jarak baru iterasi ke 2 dikurangi dengan total jarak lama pada iterasi pertama K-medoids pada penelitian ini seperti terlihat pada tabel 10.

TABEL 10  
HASIL PERHITUNGAN NILAI SIMPANGAN

Total cost Jarak terdekat		Nilai Simpangan
Iterasi ke-2( baru)	Iterasi ke-1 (lama)	Jarak baru-jarak lama
4.942.128.062	4.881.429.773	60.698.289

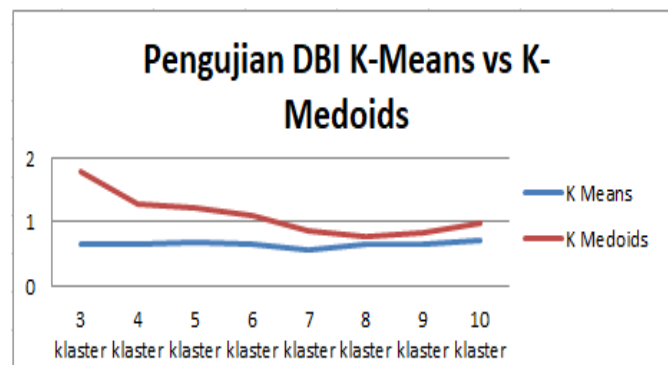
Berdasarkan tabel 10 diatas dapat disimpulkan bahwa nilai simpangan yang dihasilkan sudah lebih dari 0 ( $S > 0$ ). Maka iterasi pun dihentikan dan hasil klasterisasi adalah data anggota klaster pada iterasi pertama. Adapun sebaran data tiap cluster seperti ditunjukkan pada tabel 11 berikut

TABEL 11  
PERHITUNGAN TOTAL SIMPANGAN BERDASARKAN SELISIH COST JARAK TERDEKAT

Cluster	Jumlah Anggota	Data Anggota Cluster
Cluster ke-01	7	D26,D32,D33,D39,D58,L03,L05
Cluster Ke-02	14	B06,D15,D22,D29,D30,D31,D35, D36,D37,D38,D52,D59D60,D61,L07
Cluster ke-03	45	Selain data armada pada cluster 1 dan cluster 2

### C. Evaluasi Cluster atau Pengujian DBI K-Means dan K-Medoids

Setelah hasil clustering kedua algoritma tersebut diperoleh, kemudian dilakukan evaluasi cluster atau pengujian terhadap hasil clustering dengan menggunakan tools Rapidminer untuk mengetahui besarnya nilai *Davies Bouldin Index (DBI)*. DBI ini dipakai sebagai acuan jumlah klaster dan memperlihatkan klasterisasi terbaik atau paling relevan antara K-Means dan K-Medoids , Adapun hasil perbandingan evaluasi cluster dengan menguji nilai validasi *Davies Bouldin Index (DBI)* antara K-Means dan k-Medoids tersebut seperti ditunjukkan pada gambar 3.



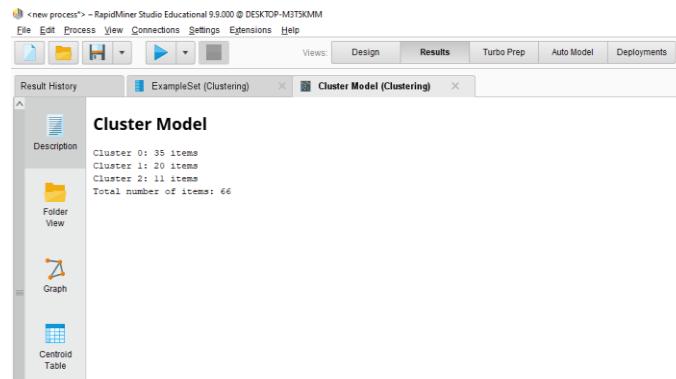
Gambar 2. Hasil Perbandingan Nilai DBI antara K-Means dengan K-Medoids

Berdasarkan hasil evaluasi seperti terlihat pada gambar 3 diatas dapat disimpulkan bahwa algoritma K-Means lebih relevan dan lebih baik untuk diimplementasikan pada aplikasi clustering berbasis web. Setelah mengacu terhadap hasil DBI pada K-Means yang lebih rendah pada seluruh percobaan baik pengelompokan dengan 3 klaster sampai 10 klaster dibandingkan dengan pada algoritma K-Medoids clustering. Perbedaan nilai DBI terjauh terlihat pada percobaan dengan menggunakan 3 klaster untuk K-Means dengan nilai DBI sebesar 0,67 dan K-Medoids sebesar 1,78-Medoids.

Sebelum mengimplementasikan algoritma K-Means pada pembuatan aplikasi berbasis web, terlebih dahulu dilakukan juga perhitungan K-Means dengan tools Rapidminer dan hasil yang diperoleh pada tools Rapidminer adalah terdapat kesesuaian hasil clustering sebesar 100% dengan perhitungan secara manual dimana jumlah cluster terbentuk kedalam 3 cluster dengan susunan dan anggota tiap cluster yang sama. Adapun hasil klasterisasi pada



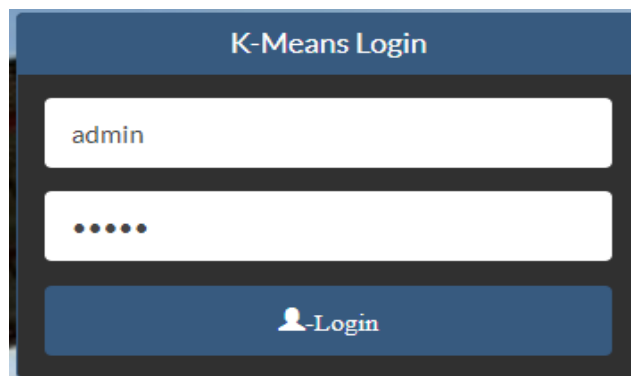
Rapidminer seperti terlihat pada gambar 3.



Gambar 3. Hasil Perhitungan K-means pada Tool Rapidminer

#### D. Implementasi Algoritma K-Means Pada Aplikasi Web

Kemudian penelitian ini dilanjutkan dengan implementasi metode K-Means tersebut pada aplikasi berbasis web yang akan dibuat. Berikut adalah tampilan hasil clustering yang dihasilkan pada aplikasi K-means clustering berbasis web. Untuk tampilan pertama aplikasi ini yaitu berupa halaman login seperti terlihat pada gambar 4.



Gambar 4. Tampilan Login Aplikasi K-Means

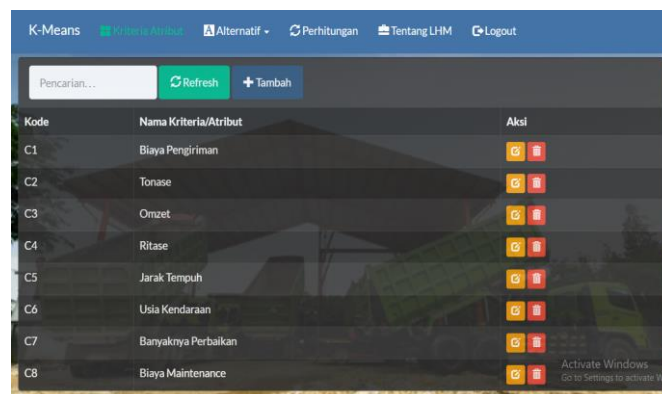
Setelah login berhasil, maka system akan menuju ke halaman beranda atau home. Pada halaman home terdapat beberapa menu yang berfungsi untuk melakukan masukan atau input data. Pada menu tersebut terdapat menu kriteria untuk menginput data atribut, menu alternatif untuk mengimport dataset dengan format file excel, menu perhitungan untuk melakukan proses K-Means clustering. Berikut tampilan halaman home seperti terlihat pada gambar 5.



Gambar 5. Tampilan Home Aplikasi K-Means

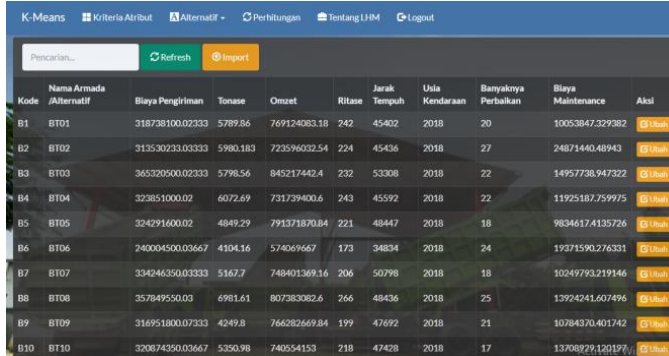
Pada halaman home ini user dapat memilih menu kriteria terlebih dahulu. Menu kriteria ini digunakan untuk menginputkan data kriteria atau atribut-atribut yang akan di pakai. Data atribut yang digunakan pada penelitian ini

sebanyak 8 atribut dan nama atribut disesuaikan dengan nama-nama pada dataset excel yang sudah disiapkan. Adapun tampilan halaman input kriteria seperti terlihat pada gambar 6



Gambar 6. Tampilan Menu Input Data Atribut

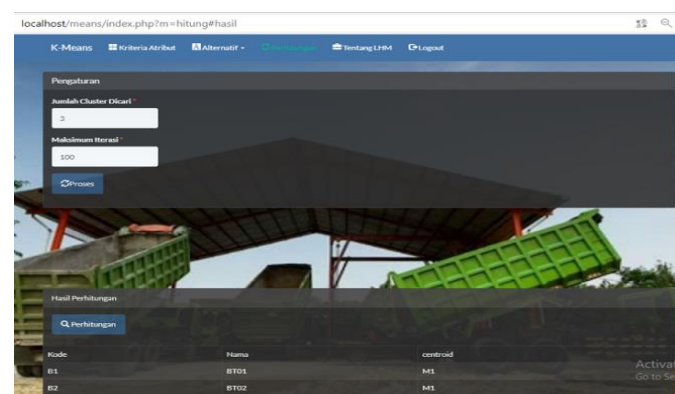
Setelah itu user dapat melakukan import file excel yang berupa dataset yang sudah disiapkan sebelumnya melalui menu nilai Alternatif. Kemudian pilih tombol import dan lakukan proses upload dataset. Adapun tampilan dari halaman untuk mengimport *setdata* tersebut seperti terlihat pada gambar 7.



Kode	Nama Armada/Alternatif	Biaya Pengiriman	Tonase	Omzet	Ritase	Jarak Tempuh	Usia Kendaraan	Banyaknya Perbaikan	Biaya Maintenance	Aksi
B1	BT01	318738100.02333	5789.86	769124083.18	242	45402	2018	20	10053847.329382	G H
B2	BT02	313530233.03333	5980.183	723596032.54	224	45436	2018	27	24871440.48943	G H
B3	BT03	365320500.02333	5798.56	845217442.4	232	53308	2018	22	14957738.947322	G H
B4	BT04	323851000.02	6072.69	731739400.6	243	45592	2018	22	11925187.759975	G H
B5	BT05	324291600.02	4849.29	791371870.84	221	48447	2018	18	9834617.4135726	G H
B6	BT06	240004500.03667	4104.16	574059667	173	34834	2018	24	19371590.276331	G H
B7	BT07	334246350.03333	5167.7	748401369.16	206	50798	2018	18	10249793.219146	G H
B8	BT08	357849550.03	6981.61	807383002.6	266	48436	2018	25	13924241.607496	G H
B9	BT09	316951800.07333	4249.8	766282669.84	199	47692	2018	21	10784370.401742	G H
B10	BT10	320874350.03667	5350.98	740554153	218	47428	2018	17	13708929.420197	G H

Gambar 7 Tampilan Menu Import Dataset

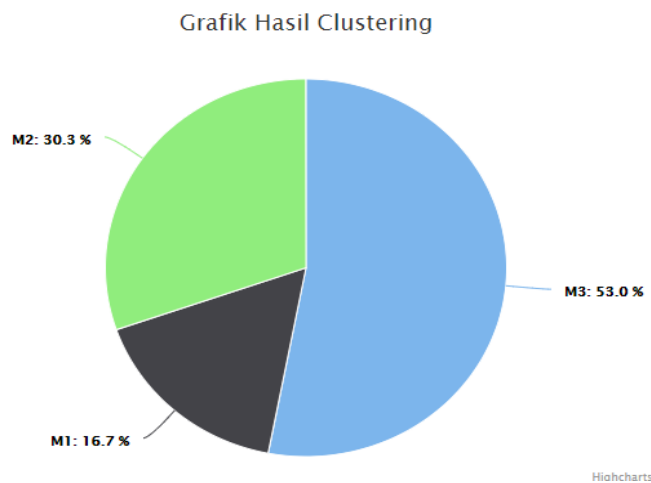
Kemudian untuk memulai perhitungan dengan K-Means user dapat langsung memilih submenu perhitungan. Sehingga sistem akan melakukan analisa data sesuai dataset yang telah di upload. Kemudian user diminta untuk menentukan jumlah kluster yang akan dicari dan jumlah maksimal iterasi yang akan di lakukan. Pada kasus ini jumlah cluster adalah 3 dan jumlah maksimal iterasi adalah 100. Berikut adalah tampilan halaman perhitungan seperti terlihat pada gambar 8.



Gambar 8 Tampilan Halaman Menu Proses Perhitungan

Pada Aplikasi K-Means berbasis web ini user juga dapat melihat tampilan visualisasi hasil proses clustering berdasarkan nilai persentase hasil clustering yang diperoleh. Tampilan visualisasi ini dalam bentuk diagram lingkaran yang terbagi kedalam 3 irisan data atau bagian sesuai dengan jumlah kluster yang dicari. Pada penelitian ini hasil clustering dalam bentuk visualisasi yang terbentuk terdiri dari 16,7% untuk cluster ke-1 dengan jumlah data

armada kendaraan truk sebanyak 10 armada, persentase sebesar 30,3% pada cluster ke-2 dengan jumlah armada truk sebanyak 21 armada, dan 53 % untuk cluster ketiga dengan jumlah armada truk sebanyak 34 armada. Berikut merupakan tampilan visualisasi seperti terlihat pada gambar 9



Gambar 9. Tampilan Grafik Hasil Perhitungan K-Means

#### IV. KESIMPULAN

Berdasarkan penelitian yang telah berhasil dilakukan, maka dapat disimpulkan bahwa hasil evaluasi cluster untuk algoritma K-Means memiliki nilai validitas DBI yang jauh lebih rendah dibandingkan dengan nilai DBI algoritma K-Medoids, pada seluruh jumlah kluster yang diujicobakan dari mulai 3 kluster sampai 10 kluster percobaan. Maka algoritma K-Means dipilih dan diterapkan pada pembuatan aplikasi Clustering berbasis web ini. Hasil clustering K-Means melalui aplikasi web diperoleh kesesuaian jumlah dan susunan anggota cluster dengan perhitungan secara manual dan tool Rapidminer sebesar 100 % untuk cluster 1 yaitu sebagai cluster dengan produktivitas rendah dengan jumlah anggota sebanyak 11 armada, pada cluster ke-2 sebagai cluster dengan produktivitas sedang sebesar 96 % dengan jumlah anggota pada perhitungan melalui aplikasi web sebanyak 21 armada, sedangkan secara manual dan juga tool Rapidminer sebanyak 20 armada, untuk cluster ke 3 sebagai cluster dengan produktivitas tinggi. diperoleh persentase kesesuaian sebesar 97% dengan jumlah anggota 34 armada pada perhitungan aplikasi web dan sebanyak 35 armada pada perhitungan secara manual dan tool Rapidminer. Dengan total persentase kesesuaian hasil cluster sebesar 97%, maka aplikasi clustering berbasis web ini masih dinilai layak dan relevan untuk dapat dimanfaatkan dan dipakai dalam melakukan proses klusterisasi atau pengelompokan armada kendaraan truk berdasarkan tingkat produktivitasnya pada PT. Lintas Harapan Mandiri.

#### DAFTAR PUSTAKA

- [1] S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data TELKOMSEL Menggunakan Metode K-Means Clustering," *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.
- [2] D. S. Maylawati, T. Priatna, H. Sugilar, and M. A. Ramdhani, "Data Science For Digital Culture Improvement in Higher Education Using K-means Clustering and Text Analytics," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 5, pp. 4569–4580, 2020, doi: 10.11591/IJECE.V10I5.PP4569-4580.
- [3] D. F. Pramesti, Lahan, M. Tanzil Furqon, and C. Dewi, "Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 9, pp. 723–732, 2017.
- [4] M. T. Islam, P. K. Basak, P. Bhowmik, and M. Khan, "Data Clustering Using Hybrid Genetic Algorithm with k-Means and k-Medoids Algorithms," 2019, doi: 10.1109/ICSEC47112.2019.8974797.
- [5] Y. Siyamto, "Pemanfaatan Data Mining Dengan Metode Clustering Untuk Evaluasi Biaya Dokumen Ekspor di PT Winstar Batam," *Media Informatika Budidarma*, vol. 1, no. 2, pp. 28–31, 2017.
- [6] F. Tempola, M. Muhammad, and A. Mubarak, "Penggunaan Internet Dikalangan Siswa SD di Kota Ternate: Suatu Survey, Penerapan Algoritma Clustering dan Validasi DBI," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 6, pp. 1153–1160, 2020, doi: 10.25126/jtiik.2020722370.
- [7] A. Badruttamam, S. Sudarno, and D. A. I. Maruddani, "Penerapan Analisis Kluster K-Modes Dengan Validasi Davies Bouldin Index Dalam Menentukan Karakteristik Kanal Youtube di Indonesia (Studi Kasus: 250 Kanal Youtube Indonesia Teratas Menurut Socialblade)," *Jurnal Gaussian*, vol. 9, no. 3, pp. 263–272, 2020, doi: 10.14710/j.gauss.v9i3.28907.
- [8] I. Kamila, U. Khairunnisa, and M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 5, no. 1, pp. 119–125, 2019, doi: 10.24014/rmsi.v5i1.7381.
- [9] Y. H. Susanti and E. Widodo, "Perbandingan K-Means dan K-Medoids Clustering terhadap Kelayakan Puskesmas di DIY Tahun 2015," *Prosiding SI MaNis (Seminar Nasional Integrasi Matematika dan Nilai Islami)*, vol. 1, no. 1, pp. 116–122, 2017.
- [10] M. Aryuni, E. Didik Madyatmadja, and E. Miranda, "Customer Segmentation in XYZ Bank Using K-Means and K-Medoids Clustering," 2018,

- doi: 10.1109/ICIMTech.2018.8528086.
- [11] B. Karthikeyan, D. J. George, G. Manikandan, and T. Thomas, "A Comparative Study on K-Means Clustering and Agglomerative Hierarchical Clustering," *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 5, 2020, doi: 10.30534/ijeter/2020/20852020.
- [12] W. Reda, H. Elazhary, and E. Hassanein, "Comparing K-means, K-medoids and Isodata Clustering Algorithms for a Cloud Service Search Engine," *International Journal of Recent Technology and Engineering*, vol. 8, no. 3, 2019, doi: 10.35940/ijrte.C5632.098319.
- [13] D. Marlina, N. Lina, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak," *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*, vol. 4, no. 2, 2018, doi: 10.24014/coreit.v4i2.4498.
- [14] S. Budiman, D. Safitri, and D. Ispriyanti, "Perbandingan Metode K-Means Dan Metode DbSCAN Pada Pengelompokan Rumah Kost Mahasiswa Di Kelurahan Tembalang Semarang," *Jurnal Gaussian*, vol. 5, no. 4, pp. 757–762, 2016.
- [15] T. Thi Bi Dan, S. Widya Sihwi, and R. Anggrainingsih, "Implementasi Iterative Dichotomiser 3 Pada Data Kelulusan Mahasiswa S1 Di Universitas Sebelas Maret," *Jurnal Teknologi & Informasi ITSmart*, vol. 4, no. 2, 2016, doi: 10.20961/its.v4i2.1770.
- [16] M. Kadafi, "Penerapan Algoritma FP-Growth untuk Menemukan Pola Peminjaman Buku Perpustakaan UIN Raden Fatah Palembang," *MATICS*, vol. 10, no. 2, 2019, doi: 10.18860/mat.v10i2.5628.
- [17] D. Aggarwal and D. Sharma, "Application Of Clustering For Student Result Analysis," *International Journal of Recent Technology and Engineering*, vol. 7, no. 6, pp. 50–53, 2019.
- [18] M. M. Khairunnisa, A. Triayudi, and E. T. E. Handayani, "Application of K-Means Clustering on the Performance Evaluation of Lecturers Based on Student Questionnaire," *Manajemen, Teknologi Informatika dan Komunikasi (Mantik)*, vol. 4, no. 1, pp. 760–766, 2020.
- [19] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. Ilmi R.H.Zer, and D. Hartama, "Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 di Indonesia," *Jti (Jurnal Teknologi Informasi)*, vol. 4, no. 1, pp. 166–173, 2020.
- [20] N. Qona'ah, A. R. Devi, and I. M. G. M. Dana, "Laboratory Clustering using K-Means, K-Medoids, and Model-Based Clustering," *Indonesian Journal of Applied Statistics*, vol. 3, no. 1, pp. 64–77, 2020, doi: 10.13057/ijas.v3i1.40823.
- [21] E. Buulolo, R. Syahputra, and A. Fau, "Algoritma K-Medoids Untuk Menentukan Calon Mahasiswa Yang Layak Mendapatkan Beasiswa Bidikmisi di Universitas Budi Darma," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 3, 2020, doi: 10.30865/mib.v4i3.2240.