

IMPLEMENTASI *K-MEANS CLUSTERING* BERBASIS *WEB* UNTUK ANALISIS DATA PENJUALAN EKSPEDISI DI LION PARCEL CILEDUG TANGERANG

Febrian Rizky Pratama¹, Purwanto^{2*}

^{1,2}Teknik Informatika, Fakultas Teknik Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ¹2011500572@student.budiluhur.ac.id, ^{2*}Purwanto@budiluhur.ac.id
(* : corresponding author)

Abstrak- Kebutuhan jasa pengiriman barang meningkat di era globalisasi, mendorong persaingan ketat di antara perusahaan logistik. Mitra lion parcel ingin mengetahui sebagai acuan untuk melakukan diskon diarea provinsi Jawa Barat, Jawa Tengah dan Jawa timur. Penelitian dibuat bertujuan untuk mengimplementasikan suatu metode data mining dengan algoritma K-Means clustering berbasis web guna menganalisis data penjualan pengiriman ekspedisi di provinsi Jawa Barat, Jawa Tengah, dan Jawa Timur pada agen Lion Parcel Ciledug Tangerang atau Selindo Travel. Metode penelitian mencakup pengumpulan data dari sistem Lion Parcel, pra-pemrosesan data, implementasi algoritma K-Means, pengembangan sistem web, serta evaluasi dan analisis hasil clustering. Hasil penelitian menunjukkan pengelompokan data pengiriman ke dalam beberapa cluster, dengan evaluasi cluster yang menghasilkan nilai purity tertinggi 68.345% untuk atribut provinsi. Pengujian aplikasi web menghasilkan dua cluster utama pada data asli dan enam cluster pada data dummy. Analisis modus mengidentifikasi nilai yang sering muncul pada berbagai atribut, membantu perusahaan memahami pola pengiriman barang. Hasil ini diharapkan memberikan informasi strategis bagi agen Lion Parcel untuk meningkatkan layanan dan mengembangkan strategi pemasaran yang efektif.

Kata Kunci: Data Mining, K-Means Clustering, Pengiriman Ekspedisi, Lion Parcel, Analisa Penjualan

WEB-BASED IMPLEMENTATION OF K-MEANS CLUSTERING FOR EXPEDITION SALES DATA ANALYSIS AT LION PARCEL CILEDUG TANGERANG

Abstract- *The need for freight forwarding services is increasing in the era of globalization, encouraging fierce competition among logistics companies. Lion parcel partners want to know as a reference to make discounts in the provinces of West Java, Central Java and East Java. This study aims to implement a data mining method with a web-based K-Means clustering algorithm to analyze expedition shipment sales data in the provinces of West Java, Central Java, and East Java at Lion Parcel Ciledug Tangerang or Selindo Travel agents. The research method includes data collection from the Lion Parcel system, data pre-processing, implementation of the K-Means algorithm, web system development, and evaluation and analysis of clustering results. The results of the study showed the grouping of shipping data into several clusters, with the cluster evaluation resulting in the highest purity value of 68.345% for provincial attributes. Web application testing results in two main clusters on the original data and six clusters on dummy data. Mode analysis identifies values that often appear on various attributes, helping companies understand the patterns of goods delivery. This result is expected to provide strategic information for Lion Parcel agents to improve services and develop effective marketing strategies.*

Keywords: *Data Mining, K-Means Clustering, Expedition Delivery, Lion Parcel, Sales Analytics*

1. PENDAHULUAN

Di era globalisasi saat ini, kebutuhan untuk menggunakan jasa pengiriman berdampak pada cara masyarakat menjalankan bisnis. Semakin meningkatnya kebutuhan akan informasi dan komunikasi secara global telah mendorong persaingan yang semakin ketat antara bisnis untuk menguasai pasar. Dengan semakin banyaknya kompetitor yang memenuhi harapan konsumen, konsumen menjadi lebih berhati-hati dan cerdas dalam mempertimbangkan setiap produk yang dirilis di pasar. Dengan demikian, konsumen akan menjadi faktor utama yang menentukan keberhasilan atau kegagalan suatu perusahaan dalam memasarkan barang dan jasanya.

Salah satu kebutuhan yang saat ini tak kalah pentingnya adalah kebutuhan jasa untuk melakukan pengiriman barang. Sektor jasa pada saat ini sangat berperan penting dalam sektor perekonomian dunia, dengan pelayanan yang baik akan mempengaruhi konsumen dalam memilih suatu jasa. Kualitas pelayanan merupakan faktor terpenting, dikarenakan pelanggan memiliki sifat kritis dalam memilih sebuah perusahaan jasa pengiriman barang yang akan digunakan. Selindo Travel atau Lion Parcel Ciledug Tangerang merupakan sebuah mitra kerja dari

perusahaan Lion Parcel (PT Lion Express). *Lion Parcel* (PT Lion Express) Merupakan sebuah perusahaan logistik yang didirikan sejak 14 Februari 2013 yang melayani *first, middle* dan *last mile* untuk pengiriman barang domestik maupun *international*.

Menurut [1], data mining melibatkan dalam suatu pencarian pola atau informasi menarik dalam data yang akan digunakan. Dengan menggunakan metode *K-Means Clustering*, data mining juga dapat membantu dalam melakukan suatu identifikasi pola dan mengelompokkan pengiriman barang di wilayah Jawa Barat, Jawa Tengah, dan Jawa Timur. *K-Means Clustering* merupakan salah satu metode yang paling umum digunakan dalam data mining.

Pada penelitian ini, metode data mining diterapkan untuk melakukan *clustering* pada data penjualan barang di Lion Parcel Ciledug Tangerang atau Selindo Travel yang beroperasi di wilayah Jawa Barat, Jawa Tengah, dan Jawa Timur. Algoritma *K-Means Clustering* memiliki keunggulan karena mudah digunakan dan mudah dipahami. Menurut [2], *K-Means Clustering* adalah metode *clustering non-hirarkis* yang mengelompokkan data ke dalam satu atau lebih cluster. Metode ini bertujuan untuk mengelompokkan objek data ke dalam kelompok yang memiliki karakteristik serupa. Perbandingan antara penelitian ini dengan penelitian sebelumnya pada penelitian [3] algoritma yang digunakan ialah *Rough Set*, dengan fokus pada pembuatan sebuah aturan umum yang bertujuan untuk mengidentifikasi kelemahan-kelemahan dalam proses pengiriman. Kemudian pada penelitian [4] algoritma yang digunakan ialah *Breadth-First Search*, dengan fokus untuk mengoptimalkan rute pengiriman agar waktu tempuh menjadi yang tercepat, yang secara langsung berhubungan dengan efisiensi pengiriman barang kepada pelanggan. Algoritma *Breadth-First Search* digunakan untuk mencari rute paling optimal dari satu titik ke titik lainnya dalam jaringan pengiriman. Selanjutnya pada penelitian [5] algoritma yang digunakan ialah *Artificial Neural Network* dan Genetika, dengan fokus untuk memprediksi pendapatan agen ekspedisi dengan menghasilkan nilai yang akurat. Kemudian pada penelitian [6] algoritma yang digunakan ialah *Agile*, dengan fokus pada pendataan dan pengiriman ekspedisi dengan pendekatan *Extreme Programming* untuk meningkatkan akurasi pendataan, *tracking* barang yang hilang dan mengurangi biaya. Pada penelitian [7] algoritma yang digunakan ialah *K-Means* dan *Hierarchical Clustering*, dengan fokus untuk membantu perusahaan logistik dalam memahami pola pengiriman barang dan menentukan strategi pemasaran yang efektif berdasarkan klasterisasi data pengiriman.

Perbedaan utama penelitian ini dibandingkan dengan penelitian-penelitian sebelumnya adalah pada fokus pengembangan sistem berbasis web yang mengimplementasikan *K-Means Clustering*. Ini memungkinkan pengguna, seperti mitra Lion Parcel, untuk dengan mudah melakukan analisis data penjualan ekspedisi dan mengidentifikasi pola pengiriman dengan antarmuka yang ramah pengguna. Selain itu, penelitian ini juga memanfaatkan atribut-atribut spesifik yang relevan dengan industri ekspedisi, seperti tanggal pemesanan, lokasi pengiriman, produk yang digunakan, kelurahan, kecamatan, kabupaten, dan provinsi penerima untuk mencapai hasil yang lebih terfokus pada kebutuhan mitra Lion Parcel.

Metode *K-Means clustering* digunakan untuk mengelompokkan data menjadi beberapa *cluster* berdasarkan karakteristik yang sama sehingga memudahkan analisis dan pengambilan keputusan. Penelitian ini akan menerapkan algoritma ini untuk mengelompokkan pengiriman barang di wilayah Jawa Barat, Jawa Tengah, dan Jawa Timur, dengan harapan mendapatkan hasil yang optimal.

2. METODE PENELITIAN

2.1 Pengumpulan Data

Tahapan penelitian yang dilakukan oleh peneliti ialah mengumpulkan suatu data yang diperlukan, tahapan ini biasanya dikenal sebagai tahap pengumpulan data. Pada tahap pengumpulan data didapatkan dari hasil wawancara dengan pemilik mitra Lion Parcel Ciledug Tangerang. Data penjualan yang didapatkan adalah data penjualan selama 6 bulan dimulai dari 01 November 2023 sampai dengan 30 April 2024 dengan total 56 *coloum* dan 2009 *record* data penjualan kemudian diseleksi menjadi 10 *coloum* yang masih bersifat data mentah atau data belum siap olah.

2.2 Algoritma K-Means

K-Means merupakan sebuah metode data *clustering* yang memungkinkan setiap objek yang bukan variabel untuk dimasukkan dalam pengelompokan ke dalam (k) *cluster* atau *non-hirarki* yang dimana dalam pengelompokannya data yang berbentuk satu atau lebih kelompok atau *cluster* [2]. Menurut [8] Algoritma *K-Means* termasuk suatu metode *Partitioning Method*, yaitu membangun berbagai partisi dan kemudian mengevaluasi suatu partisi dengan beberapa kriteria. *K-Means* dapat diartikan sebagai suatu teknik *clustering* yang di dapatkan atau diperoleh dari setiap titik terhadap pusat cluster dengan melakukan langkah atau suatu tindakan yang di ulang-ulang untuk memperbaiki dan menambahkan suatu hasil atau *iteratve*[9].

1. Total data *cluster* yang dibutuhkan
2. Jenis *attribute* yang digunakan ialah *numeric*. Proses untuk melakukan *clustering* menggunakan algoritma K-Means ialah :
 - a. Tentukan nilai *k* (klaster) dalam total cluster yang akan dibentuk.
 - b. Pilih nilai centroid pada cluster yang berawal dari *k* (klaster).
 - c. Menghitung setiap jarak data input dan setiap data centroid menggunakan sebuah rumus jarak *euclidean* untuk menemukan sebuah jarak terdekat diantara setiap titik suatu data dengan centroid. Untuk menghitung setiap jarak data input atau pada setiap data *centroid* dapat menggunakan teori perhitungan dengan rumus teori jarak *euclidean* dapat digunakan untuk menghitung jarak dari semua data kepusat setiap *cluster*. Berikut merupakan rumus perhitungan jarak *euclidean* antara lain :

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

Keterangan :

x : Titik suatu data yang akan diperhitungkan jaraknya.

c_j : Suatu nilai centroid ke- j

x_i : Suatu komponen ke- i dari data ke- j

c_{ji} : Suatu Komponen ke- i dari centroid ke- j

- d. Memperbarui *centroid* yang sudah dihitung pada point 3 (tiga) setelah semua titik data yang dikelompokkan, *centroid* baru dihitung dengan mengambil perhitungan rata-rata dari semua titik data dalam klaster. Berikut rumus untuk memperbarui *centroid* antara lain :

$$\mu_j = \frac{1}{|C_j|} \sum_{x \in C_j} x \quad (2)$$

Keterangan rumus :

μ_j : Centroid baru untuk *cluster* ke- j

$|C_j|$: Jumlah titik data dalam *cluster* C_j .

$\sum_{x \in C_j} x$: Penjumlahan semua titik data x yang dialokasikan ke *cluster* C_j .

$\frac{1}{|c_j|}$: Rata-rata dari semua titik data dalam *cluster* C_j .

- e. Lakukanlah sebuah perulangan pada bagian point ke 3 hingga point ke 4 pada setiap cluster agar tetap tidak berubah.
- f. Menurut [10] *purity* ialah suatu ukuran evaluasi sederhana dan transparan. *Purity* dihitung dengan cara menjumlahkan suatu data atau dokumen yang berada di *cluster* yang digunakan lalu dibagi dengan jumlah keseluruhan dokumen. Nilai dari *purity* berkisar antara 0 (nol) dan 1 (satu), yang dimana nilai 1 (satu) menunjukkan bahwa pada setiap cluster hanya mengandung data dari satu kategori saja, dan nilai 0 menunjukkan bahwa cluster berisi data dari berbagai kategori yang terhitung secara merata. Berikut rumus perhitungan *purity*.

$$Purity = \frac{1}{N} \sum_{k=1}^K \max_j |C_k \cap L_j| \quad (3)$$

Keterangan :

N : Merupakan jumlah total data.

K : Merupakan jumlah dari cluster.

C_k : Merupakan cluster ke- k .

L_j : Merupakan kategori ke- j .

$|C_j \cap L_j|$: Merupakan jumlah data didalam suatu cluster ke- k yang berasal dari kategori ke- j .

\max_j : Merupakan jumlah data terbanyak dari 1 (satu) kategori didalam cluster ke- k

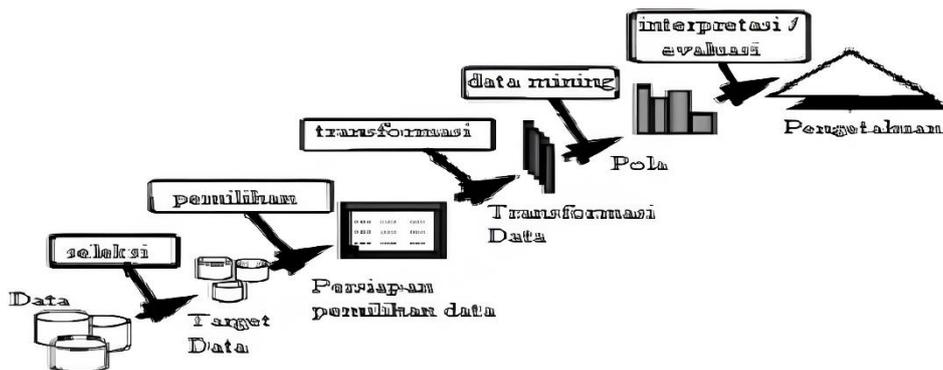
2.3 Clustering

Salah satu teknik data mining yang sering digunakan adalah *clustering*, yaitu proses pengelompokan satu set objek data ke dalam kelompok-kelompok yang disebut *cluster* [9]. *Clustering* adalah proses membagi data ke

dalam beberapa kelompok yang memiliki kesamaan karakteristik yang lebih besar di dalam kelompok tersebut dibandingkan dengan kelompok lainnya [2].

2.4 Knowledge Discovery in Database

Menurut [11] *Knowledge discovery in database* (KDD) merupakan suatu proses yang bertujuan untuk menganalisis dan menggali sebanyak besarnya suatu himpunan data serta mengekstrak informasi serta pengetahuan yang berguna. Menurut [12] Tahapan pada *data mining* dengan metode KDD atau *Knowledge Discovery in Database* ini mempunyai 6 point utama yaitu :



Gambar 1. Tahapan Proses KDD Dalam Data Mining

1. Selection Data

Pada tahapan *selection* data perlunya melakukan pemilihan data atau pemilihan *attribute* yang diperlukan. Dikarenakan tidak semua *attribute* data yang ada dalam *database* tersebut dibutuhkan.

2. Preprocessing Data

Pada tahapan *preprocessing* data yang digunakan ialah *cleansing* data atau *data cleaning*. *Data cleaning* ialah suatu langkah yang dalam prosesnya untuk menghilangkan *data noise* atau *missing value*. *Data cleaning* bertujuan ialah untuk menghindari dari suatu data yang memiliki nilai *null* atau tidak sesuai dalam setiap data atau *attribute* yang ada.

3. Transformation Data

Pada tahapan *transformation* data proses yang dilakukan ialah proses pengnormalisasian.. *Transformation* data bertujuan untuk mengubah suatu bentuk nilai pada setiap *attribute* yang sesuai dengan algoritma yang akan digunakan. Menurut [13] *transformation* ialah merubah data kedalam bentuk yang sesuai menggunakan format pengolahan data pada algoritma data mining yang digunakan. Setiap algoritma data mining dapat diolah dengan format tertentu.

4. Data Mining

Menurut [14] *data mining* dapat disebut juga sebagai *knowledge discovery in database* adalah sebuah kegiatan yang meliputi suatu pengumpulan, pemakaian data, historis untuk menemukan sebuah keteraturan, pola atau suatu hubungan didalam set data yang berukuran besar. Menurut [15] *data mining* adalah suatu metode pengolahan data yang bertujuan untuk menemukan suatu pola-pola tersembunyi dalam suatu data. x

5. Evaluation Data

Pada tahapan *evaluation* data, perlunya melakukan analisis terhadap suatu hasil dari suatu proses pembelajaran data. Pada tahapan ini perlunya penggunaan suatu metode, yaitu metode *Purity*, metode ini memiliki tujuan untuk mengukur seberapa baik *cluster* yang dihasilkan oleh algoritma K-Means mencerminkan label asli dari data.

6. Knowledge

Setelah hasil analisa yang diperlukan telah didapatkan, maka tahap selanjutnya ialah hasil dari penelitian kita dapat menjadi sebuah *knowledge* atau pengetahuan bagi orang-orang yang menggunakan analisa kita.

2.5 Pengujian Hasil

Pada tahapan pengujian hasil, pada tahap ini merupakan sebuah tahapan akhir yang memiliki tujuan serta hasil dari *output* mengenai analisis dari sistem ini dengan proses perhitungan secara manual terdapat sebuah kesamaan. Dalam hal ini untuk mendapatkan suatu perbandingan dengan hasil yang didapat dari aplikasi berbasis

web yang dirancang dengan hasil proses perhitungan manual menggunakan algoritma *K-Means* dengan metode *clustering*.

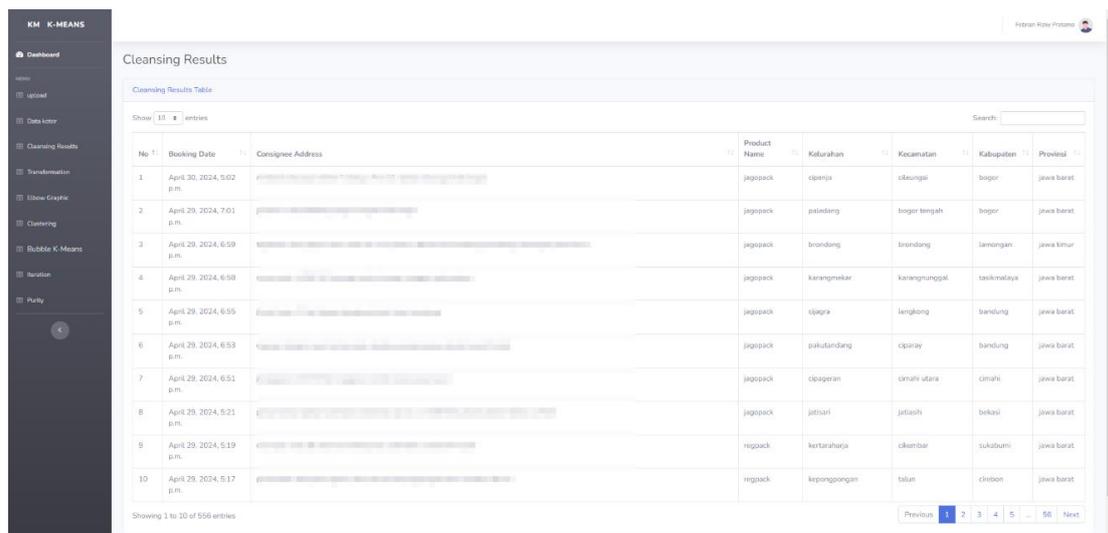
3. HASIL DAN PEMBAHASAN

Tabel 1 merupakan hasil dari *labeling* dan *transformation* yang dirubah dari data yang sudah di bersihkan pada tahap *pre-processing* yang nantinya akan diproses dengan cara memilih *centroid* secara random. 9 data yang akan digunakan merupakan data yang diambil. Setiap *attribute* diambil dari **Gambar 2**, yang dimana *attribute booking date* merupakan tanggal pemesanan, *consignee address* merupakan alamat penerima, *product name* merupakan product pengiriman yang digunakan, kelurahan, kecamatan, kabupaten dan provinsi merupakan area pengiriman dari penerima.

Tabel 1. Data Transformation

Id Shipping	Booking Date	Consignee Address	Product Name	Product			
				Kelurahan	Kecamatan	Kabupaten	Provinsi
1	6	131	3	201	584	131	11
2	6	131	3	202	554	131	11
3	6	105	3	203	554	169	33
4	6	137	3	204	554	137	11
5	6	130	3	205	700	130	11
6	6	131	3	206	554	130	11
7	6	136	3	207	970	136	11
8	6	134	3	208	740	134	11
9	6	132	2	209	554	132	11
556	1	150	2	544	1078	150	33

Nilai yang terdapat pada **Tabel 1** merupakan hasil *transformation* dari data bersih atau *cleansing*, berikut pada **Gambar 2** merupakan data bersih atau *cleansing* sebelum di *transformation* seperti pada **Tabel 1**.



The screenshot shows a web application interface with a sidebar menu on the left containing options like Dashboard, Upload, Data Input, Cleansing Results, Transformation, Elbow Graph, Clustering, Bubble K-Means, Scatter, and Purify. The main content area displays a table titled 'Cleansing Results Table' with 10 rows of data. The columns are: No, Booking Date, Consignee Address, Product Name, Kelurahan, Kecamatan, Kabupaten, and Provinsi. The data in the table is as follows:

No	Booking Date	Consignee Address	Product Name	Kelurahan	Kecamatan	Kabupaten	Provinsi
1	April 30, 2024, 5:02 p.m.	[Redacted]	jagoack	cipeper	cibeunung	bagor	jawa barat
2	April 29, 2024, 7:01 p.m.	[Redacted]	jagoack	padalung	bagor tengah	bagor	jawa barat
3	April 29, 2024, 6:59 p.m.	[Redacted]	jagoack	brondong	brondong	tanjungan	jawa timur
4	April 29, 2024, 6:58 p.m.	[Redacted]	jagoack	karangmekar	karangmunggil	tasikmalaya	jawa barat
5	April 29, 2024, 6:55 p.m.	[Redacted]	jagoack	ciager	lengkong	bandung	jawa barat
6	April 29, 2024, 6:53 p.m.	[Redacted]	jagoack	pakutandang	cisaray	bandung	jawa barat
7	April 29, 2024, 6:51 p.m.	[Redacted]	jagoack	ciageran	cihaur utara	cihaur	jawa barat
8	April 29, 2024, 5:21 p.m.	[Redacted]	jagoack	jatisari	jatisari	bekasi	jawa barat
9	April 29, 2024, 5:19 p.m.	[Redacted]	jagoack	kertaharja	ciembur	sukabumi	jawa barat
10	April 29, 2024, 5:17 p.m.	[Redacted]	jagoack	kepongongan	talun	cirobon	jawa barat

Gambar 2. Data Bersih Sebelum Di Transformation

Dari **Tabel 2** yang diambil untuk perhitungan hanya 9 data *records* dari 556 data pada *transformation*. Setelah mengambil 9 *record* data untuk perhitungan, kemudian menentukan *centroid* awalnya secara random. Pada perhiungan dibawah ini menggunakan $k=3$.

Tabel 2. *Centroid* Awal

<i>Centroid</i>	Booking Date	Consignee Address	Product Name	Kelurahan	Kecamatan	Kabupaten	Provinsi
C1	6	131	3	201	584	131	11
C2	6	132	2	209	554	132	11
C3	6	137	3	204	554	137	11

Setelah menentukan *centorid* awal pada **Tabel 2**, kemudian tahap selanjutnya adalah menghitung iterasi 0 atau iterasi awal. Berikut contoh cara melakukan perhitungannya.

Iterasi 0 :

Data 1:

Jarak ke *centroid* 1:

$$\sqrt{(6-6)^2 + (131-131)^2 + (3-3)^2 + (201-201)^2 + (584-584)^2 + (131-131)^2 + (11-11)^2} = \sqrt{0} = 0$$

Jarak ke *centroid* 2:

$$\sqrt{(6-6)^2 + (131-132)^2 + (3-2)^2 + (201-209)^2 + (584-554)^2 + (131-132)^2 + (11-11)^2} = \sqrt{0+1+1+64+900+1+0} = \sqrt{967} = 31.10$$

Jarak ke *centroid* 3 :

$$\sqrt{(6-6)^2 + (131-137)^2 + (3-3)^2 + (201-204)^2 + (584-554)^2 + (131-137)^2 + (11-11)^2} = \sqrt{0+36+0+9+900+36+0} = \sqrt{981} = 31.32$$

Hasil cluster: *centroid* 1

Pada contoh perhitungan data 1 dengan menghitung menggunakan 3 *centroid* awal yang dipilih random, kemudian hasil yang ditentukan diambil dari nilai terkecil dari 3 jarak antar *centroid*. Berikut hasil yang didapatkan dari perhitungan iterasi 0 pada **tabel 3**.

Tabel 3. Hasi Iterasi 0

Id shipping	C1	C2	C3	Hasil
1	0	31.10	31.32	C1
2	30.02	7.21	8.72	C2
3	59.24	51.17	50.33	C3
4	31.32	8.72	0	C3
5	116.09	146.06	146.33	C1
6	30.43	3.87	9.43	C2
7	384.88	416.04	416.02	C1
8	156.17	186.03	186.07	C1
9	31.10	0	8.72	C2

Tahap berikutnya adalah menghitung rata-rata jumlah *cluster* yang didapatkan dari data yang telah didapatkan untuk mendapatkan titik *centroid* baru atau iterasi pertama. Berikut perhitungan rata-rata pada *centroid* 1, *centroid* 2 dan *centroid* 3.

Rata-rata dari *centroid* 1 :

$$\text{Booking Date} : \frac{6+6+6+6}{4} = 6$$

Rata-rata dari *centroid* 2 :

$$\text{Booking Date} : \frac{6+6+6}{3} = 6$$

$$\text{Consignee Address} : \frac{131+130+136+134}{4} = 132.75$$

$$\text{Product Name} : \frac{3+3+3+3}{4} = 3$$

$$\text{Kelurahan} : \frac{201+205+207+208}{4} = 205.25$$

$$\text{Kecamatan} : \frac{584+700+970+740}{4} = 132.75$$

$$\text{Kabupaten} : \frac{131+130+136+134}{4} = 132.75$$

$$\text{Provinsi} : \frac{11+11+11+11}{4} = 11$$

$$\text{Consignee Address} : \frac{131+131+132}{3} = 131.33$$

$$\text{Product Name} : \frac{3+3+2}{3} = 2.67$$

$$\text{Kelurahan} : \frac{202+206+209}{3} = 205.67$$

$$\text{Kecamatan} : \frac{554+554+554}{3} = 554$$

$$\text{Kabupaten} : \frac{131+130+132}{3} = 131$$

$$\text{Provinsi} : \frac{11+11+11}{3} = 11$$

Centroid 1 baru : (6, 132.75, 3, 205.25, 748.5, 132.75, 11)

Centroid 2 baru : (6, 131.33, 2.67, 205.67, 554, 131, 11)

Rata-rata dari centroid 3 :

$$\text{Booking Date} : \frac{6+6}{2} = 6$$

$$\text{Consignee Address} : \frac{105+137}{2} = 121$$

$$\text{Product Name} : \frac{3+3}{3} = 3$$

$$\text{Kelurahan} : \frac{203+204}{2} = 203.5$$

$$\text{Kecamatan} : \frac{554+554}{2} = 554$$

$$\text{Kabupaten} : \frac{169+137}{2} = 153$$

$$\text{Provinsi} : \frac{33+11}{2} = 22$$

Centroid 3 Baru : (6, 121, 3, 203.5, 554, 153, 22)

Berikut pada **tabel 4** hasil *centroid* baru.

Tabel 4. Nilai Centroid iterasi 1

Centroid	Booking Date	Consignee Address	Product Name	Kelurahan	Kecamatan	Kabupaten	Provinsi
C1	6	132.75	3	205.25	748.5	132.75	11
C2	6	131.33	2.67	205.67	554	131	11
C3	6	121	3	203.5	554	153	22

Setelah menghitung rata-rata untuk mendapatkan titik *centroid* baru, tahap selanjutnya kembali menghitung iterasi selanjutnya hingga hasilnya tidak berubah.

Tabel 5. Hasil Iterasi 1

Id shipping	C1	C2	C3	Hasil
1	164.61	30.34	40.12	C2
2	194.57	3.70	26.60	C2
3	200.68	51.22	25.16	C3
4	194.50	8.42	25.16	C2
5	48.22	146	148.49	C1
6	194.38	1.15	27.50	C2
7	222.02	416.16	416.70	C1
8	9.13	186.08	187.80	C1
9	194.56	3.61	26.73	C2

Pada penelitian ini iterasi terhenti pada iterasi ke-2 hasil perhitungan *clustering* data menunjukkan hasil setiap data tidak berpindah *cluster*, maka iterasi selanjutnya tidak perlu dilakukan. Berikut **tabel 6** nilai *centorid* baru untuk perhitungan iterasi 2.

Tabel 6. *Centroid* Baru Iterasi 2

<i>Centroid</i>	Booking Date	Consignee Address	Product Name	Kelurahan	Kecamatan	Kabupaten	Provinsi
C1	6	133.33	3	206.67	803.33	133.33	11
C2	6	132.4	2.8	204.4	560	132.4	11
C3	6	105	3	203	554	169	33

Setelah menentukan *centroid* baru untuk iterasi ke-2, tahap selanjutnya menghitung iterasi ke-2. Berikut **tabel 7** hasil dari perhitungan iterasi ke-2.

Tabel 7. Hasil Iterasi 2

Id shipping	C1	C2	C3	Hasil
1	218.92	24.32	59.24	C2
2	250.25	6.76	51.04	C2
3	255.03	51.05	0	C3
4	250.25	8.86	50.33	C2
5	103.42	140.01	154.73	C1
6	250.01	6.81	51.85	C2
7	167.02	410.05	419.08	C1
8	63.46	180.02	192.79	C1
9	250	7.62	51.18	C2

Setelah menghitung iterasi ke-2, selanjutnya menghitung kembali *centroid* baru untuk *centroid* ke-3. Berikut pada **tabel 8** hasil *centroid* baru iterasi ke-3.

Tabel 8. *Centroid* Baru Iterasi 3

<i>Centroid</i>	Booking Date	Consignee Address	Product Name	Kelurahan	Kecamatan	Kabupaten	Provinsi
C1	6	133.33	3	206.67	803.33	133.33	11
C2	6	132.4	2.8	204.4	560	132.4	11
C3	6	105	3	203	554	169	33

Pada **tabel 8** dan **tabel 6** dapat dilihat untuk hasil perhitungan iterasi ke-2 dan iterasi ke-1 hasilnya sama atau tidak berubah maka perhitungan diberhentikan. Berikut semua perhitungan iterasi yang telah dilakukan pada **tabel 9**.

Tabel 9. Hasil Semua Perhitungan Iterasi

Iterasi	Centroid	1	2	3	4	5	6	7
0	C1	6	132.75	3	205.25	748.5	132.75	11
0	C2	6	131.33	2.67	205.67	554	131	11
0	C3	6	121	3	203.5	554	153	22
1	C1	6	133.33	3	206.67	803.33	133.33	11
1	C2	6	132.4	2.8	204.4	560	132.4	11
1	C3	6	105	3	203	554	169	33
2	C1	6	133.33	3	206.67	803.33	133.33	11
2	C2	6	132.4	2.8	204.4	560	132.4	11
2	C3	6	105	3	203	554	169	33

Setelah mendapatkan hasil perhitungan iterasi yang sama atau tidak berubah, kemudian menentukan *cluster* yang didapatkan pada setiap *ide* atau data. Hasil *cluster* diambil dari perhitungan iterasi terakhir, pada penelitian iterasi berhenti pada iterasi ke 2. Berikut hasil *cluster* yang didapatkan pada **tabel 10**.

Tabel 10. Hasil Semua Perhitungan Iterasi terakhir

Id Shipping	Cluster
1	2
2	2
3	3
4	2
5	1
6	2
7	1
8	1
9	2

Setelah menghitung iterasi dan mendapatkan *cluster* dari hasil perhitungan, selanjutnya menghitung *purity*. *Purity* merupakan suatu metrik evaluasi yang digunakan untuk mengetahui kemurnian suatu cluster yang dihasilkan yang sesuai dengan asli dari data. Nilai *purity* berkisar 0 hingga 1, nilai *purity* dapat dikatakan bagus bila mendekati nilai 1. Berikut contoh perhitungan *purity* untuk menghitung *attribute* kecamatan sebagai berikut :

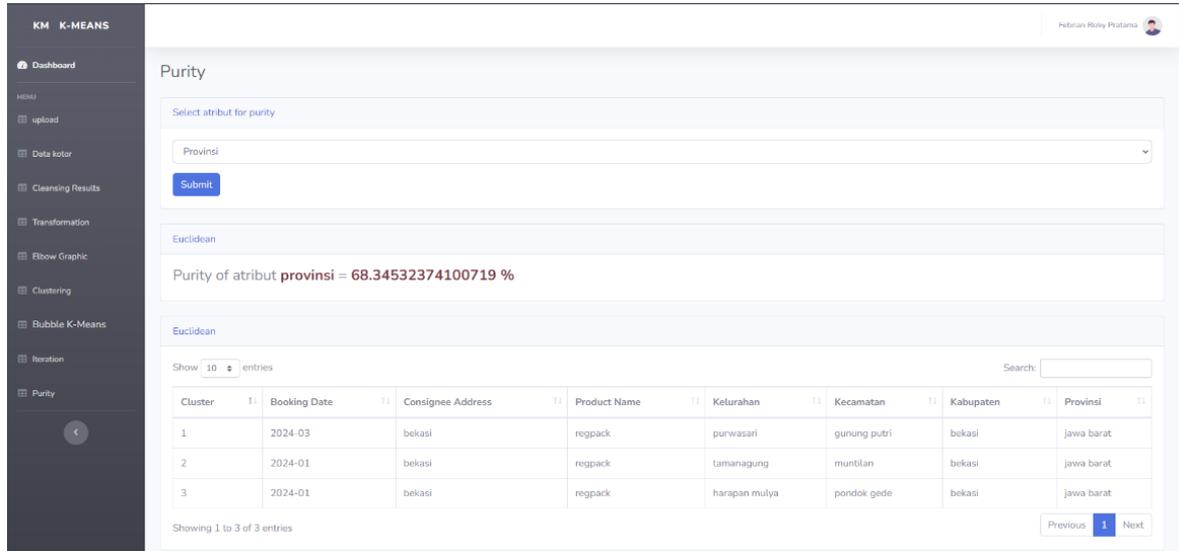
- Menghitung Jumlah total elemen (N)
N = 9
- Menentukan klaster dan kecamatan yang ada
Cluster : 1,2,3
Kecamatan : 584, 554, 700, 970, 740
- Menghitung jumlah elemen di setiap *cluster* berdasarkan kecamatan
Cluster 1 : Cluster 2 : Cluster 3 :
Kecamatan 700 : 1 Kecamatan 584 : 1 Kecamatan 554 : 1
Kecamatan 970 : 1 Kecamatan 554 : 4
Kecamatan 740 : 1 Kecamatan 132 : 1
- Menentukan Jumlah maksimum elemen dari kelas setiap yang sama disetiap klaster
Cluster 1 : max (1,1,1) = 1 Cluster 2 : max (1,4,1) = 5 Cluster 3 : max (1) = 1
- Menghitung Purity

$$Purity = \frac{1}{9} \times (1 + 4 + 1) = \frac{6}{9} = 0.6667 * 100 = 66.67\%$$

Setelah semua dihitung seperti kecamatan maka *purity* yang didapatkan pada setiap *attribute*, berikut rincian hasil lainnya antara lain :

- Booking date : 1 atau 100%
- Consignee address : 0.4444 atau 44.44%
- Product name : 0.8889 atau 88.89%
- Kelurahan : 0.2222 atau 22.22%
- Kabupaten : 0.4444 atau 44.44%
- Provinsi : 0.8889 atau 88.89%

Nilai *purity* yang tinggi menunjukkan bahwa klaster yang dihasilkan murni, yang berarti *cluster* tersebut didominasi oleh elemen-elemen dari satu kelas tertentu. Makna dari setiap klaster yang terbentuk ialah hasil dari klustersasi, terbentuk beberapa klaster yang mencerminkan pola pengiriman berdasarkan atribut tertentu. Setiap klaster memiliki makna tersendiri yang dapat diindikasikan melalui atribut yang sering muncul atau modus dalam klaster tersebut. Kesamaan di setiap anggota pada setiap klaster dan label klaster kesamaan diantara anggota setiap klaster dapat dilihat dari atribut yang memiliki modus terbanyak atau dominan. Misalkan, jika dalam suatu klaster atribut "provinsi" jawa barat dan "kecamatan" bekasi lebih dominan, maka kesamaannya diantara banyak nya anggota klaster tersebut adalah pengiriman yang dominan di wilayah "provinsi" jawa barat. Berdasarkan kesamaan ini, klaster tersebut dapat diberi label "Pengiriman Terbanyak atau Dominan Di Jawa Barat". Maka dengan ini label klaster mencerminkan pola yang sering muncul dalam data, berdasarkan kesamaan atribut diantara anggota klaster.



Purity

Select atribut for purity

Provinsi

Submit

Euclidean

Purity of atribut provinsi = 68.34532374100719 %

Euclidean

Show 10 entries

Cluster	Booking Date	Consignee Address	Product Name	Kelurahan	Kecamatan	Kabupaten	Provinsi
1	2024-03	bekasi	regpack	purwasari	gunung putri	bekasi	jawa barat
2	2024-01	bekasi	regpack	tamanagung	muntilan	bekasi	jawa barat
3	2024-01	bekasi	regpack	harapan mulya	pondok gede	bekasi	jawa barat

Showing 1 to 3 of 3 entries

Previous 1 Next

Gambar 3. Hasil Purity dan Hasil Modus Terbanyak Didalam Kluster

4. KESIMPULAN

Berdasarkan hasil analisis dan uji coba program sistem analisis data penjualan Lion Parcel Ciledug Tangerang menggunakan *K-Means Clustering* yang dilakukan oleh peneliti dapat disimpulkan, bahwa pada evaluasi *purity* dengan nilai tertinggi didapatkan oleh *attribute* provinsi sebesar 0.683 / 68.3%. Kemudian hasil *cluster* yang didapatkan menggunakan perhitungan $K=3$ dengan 556 *record* dengan mengambil titik *centroid* secara random didapatkan untuk *cluster* 1 (C1) 153 data, *cluster* 2 (C2) 103 data dan *cluster* 3 (C3) 300 data. Kemudian tingkat kemunculan tertinggi pada *cluster* 1 ialah 2024-03, bekasi, regpack, purwasari, gunung putri, bekasi, jawa barat. Kemudian untuk *cluster* 2 ialah 2024-01, bekasi, regpack, baktijaya, medan satria, bekasi, jawa barat. Dan pada *cluster* 3 ialah 2024-01, bekasi, regpack, tamanagung, muntilan, bekasi, jawa barat.

DAFTAR PUSTAKA

- [1] H. U. Rabbani and D. Anubhakti, "Implementasi Data Mining Untuk Prediksi Tangga Lagu Aplikasi Langit Musik," *Tek. Komput. Jar.*, pp. 29–34, 2020.
- [2] Y. Pratama, H. Hendrawan, E. Rasywir, B. T. Carenina, and D. R. Anggraini, "Penerapan Algoritma K-Means clustering Untuk Mengelompokkan Provinsi Berdasarkan Banyaknya Desa/Kelurahan Dengan Upaya Antisipasi/Mitigasi Bencana Alam," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1232–1240, 2022, doi: 10.47065/bits.v4i3.2549.
- [3] H. S. Simorangkir, "Analisa Data Ekspedisi Paket Dengan Menerapkan Algoritma Rough Set (Studi Kasus: Jne Agen Menteng Medan)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 228–234, 2019, doi: 10.30865/komik.v3i1.1593.
- [4] D. Rizky Deovalent Zudianta, "Pencarian Rute Terbaik Pengiriman Barang Menggunakan Algoritma Breadth-First Search," *J. Ilm. MIKA AMIK Al Muslim, J.*, vol. IV, no. 2, pp. 69–75, 2020.
- [5] A. Y. Naufal, M. Tafrikan, and A. K. Rachmawati, "Implementasi Backpropagation ANN dan Algoritma Genetika Terhadap Estimasi Pendapatan Agen Ekspedisi Pengiriman Barang," *Walisongo J. Inf. Technol.*, vol. 5, no. 1, pp. 65–78, 2023, doi: 10.21580/wjit.2023.5.1.14452.
- [6] A. Rinaldi, F. W. Utomo, and ..., "Implementasi Analisis Penelusuran Pengiriman Barang Paket Berbasis Android Dengan Metode Agile Pada Anandamaya Residences," *Bul. Ilm. Ilmu ...*, vol. 1, no. 4, pp. 555–569, 2023, [Online]. Available: <http://jurnalmahasiswa.com/index.php/biikma/article/view/688%0Ahttp://jurnalmahasiswa.com/index.php/biikma/article/download/688/406>.
- [7] R. Mahesh Prabhu, M. S. Hema, S. Chepure, and M. Nageswara Guptha, "Logistics optimization in supply chain management using clustering algorithms," *Scalable Comput.*, vol. 21, no. 1, pp. 107–114, 2020, doi: 10.12694/SCPE.V21I1.1628.
- [8] R. Hablum, A. Khairan, and R. Rosihan, "Clustering Hasil Tangkap Ikan Di Pelabuhan Perikanan Nusantara (Ppn) Ternate Menggunakan Algoritma K-Means," *JIKO (Jurnal Inform. dan Komputer)*, vol. 2, no. 1, pp. 26–33, 2019, doi: 10.33387/jiko.v2i1.1053.
- [9] H. Prastiwi, Jeny Pricilia, and Errissya Rasywir, "Implementasi Data Mining Untuk Menentuksn Persediaan Stok Barang

- Di Mini Market Menggunakan Metode K-Means Clustering,” *J. Inform. Dan Rekayasa Komputer(JAKAKOM)*, vol. 2, no. 1, pp. 141–148, 2022, doi: 10.33998/jakakom.2022.2.1.34.
- [10] R. Cahyanto, A. R. Chrismanto, and D. Sebastian, “Pengelompokan Komentar Dataset Sentipol dengan Modified K-Means Clustering,” *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 3, pp. 531–540, 2020, doi: 10.28932/jutisi.v6i3.3006.
- [11] I. K. J. Arta, G. Indrawan, and G. Rasben Dantes, “Data Mining Rekomendasi Calon Mahasiswa Berprestasi di STMIK Denpasar Menggunakan Metode Technique For Other Reference By Similarity to Ideal Solution,” *J. Ilmu Komput. Indones.*, vol. 4, no. 1, pp. 11–21, 2019.
- [12] T. K. Titus and M. Jajuli, “Clustering Data Kecelakaan Lalu Lintas di Kecamatan Cileungsi Menggunakan Metode K-Means,” *Gener. J.*, vol. 6, no. 1, pp. 1–12, 2022, doi: 10.29407/gj.v6i1.16103.
- [13] H. Syahputra, L. Mayola, and D. Guswandir, “Clustering Tingkat Penjualan Menu (Food and Beverage) Menggunakan Algoritma K-Means,” *J. KomtekInfo*, vol. 9, pp. 29–33, 2022, doi: 10.35134/komtekinfo.v9i1.274.
- [14] Feryanto, F. T. Kesuma, and S. P. Tamba, “Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering,” *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 2, no. 2, pp. 67–72, 2020, doi: 10.34012/jusikom.v2i2.376.
- [15] R. Riadi and Mesran, “Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Analisa Penjualan Parfume,” *J. Informatics, Electr. Electron. Eng.*, vol. 2, no. 4, pp. 138–145, 2023, doi: 10.47065/jieee.v2i4.1181.