

Perbandingan Algoritma K-Means dan K-Medoids Dalam Penclusteringan Data Penjualan PT. United Teknologi Integrasi

*Comparison between K-Means and K-Medoids Algorithm in the Collection of Sales
Data of PT. United Teknologi Integrasi*

Agung Lesmana¹, Wawan Gunawan²

^{1,2}Fakultas Ilmu Komputer, Universitas Mercu Buana

E-mail: ¹41517120044@student.mercubuana.ac.id, ²wawan.gunawan@mercubuana.ac.id

Abstrak

Pencatatan transaksi penjualan dalam sebuah perusahaan merupakan hal yang sangat penting untuk dilakukan. Dengan pencatatan transaksi penjualan setiap harinya perusahaan dapat mengetahui peningkatan tingkat penjualan barang. Data transaksi penjualan berisi data barang yang telah terjual, jumlah barang, nama barang, harga barang serta nama pelanggan yang melakukan pembelian barang. Data penjualan ini dapat dimanfaatkan untuk membantu perusahaan dalam meningkatkan penjualan perusahaan. Penelitian ini memanfaatkan teknik data mining yaitu clustering data dengan algoritma K-Means dan K-Medoids pada data penjualan. Pada penelitian ini dilakukan clustering data penjualan, data yang diambil dimulai dari tahun 2018 – 2020 di PT United Teknologi Integrasi. Data diambil dari tahun 2018 dikarenakan sistem pencatatan yang diterapkan di perusahaan ini mulai digunakan pada sejak tahun 2018. Penentuan clustering ditentukan dengan menggunakan metode elbow yang menunjukkan hasil cluster yang paling optimal dibentuk adalah sebanyak tiga cluster, yaitu produk yang terjual paling laris, laris dan kurang laris. Dari hasil tersebut dimanfaatkan untuk memberikan promosi yang lebih banyak kepada produk yang ada pada cluster kurang laris untuk dapat meningkatkan penjualan. Kemudian dilakukan perbandingan hasil clustering algoritma dengan uji kualitas cluster menggunakan metode Silhouette Index yang menghasilkan nilai maksimum Index 0.404 pada algoritma K-Means sedangkan 0.376 pada algoritma K-Medoids .

Kata kunci: algoritma k-medoids, algoritma k-means, data mining, clustering, data penjualan

Abstract

Recording sales transactions in a company is a very important thing to do. By recording sales transactions everyday the company can find out the increase in the level of sales of goods. Sales transaction data contains data on goods that have been sold, the number of goods, the name of goods, the price of goods, and the name of customers who make purchases of goods. This sales data can be utilized to help the company in increasing the company's sales. This research utilizes data mining techniques that cluster data with K-Means and K-Medoids algorithms on sales data. This study conducted clustering sales data, data taken starting from 2018 - 2020 at PT United Teknologi Integrasi. Data were taken from 2018 because the recording system implemented in this company began to be used in 2018. The determination of clustering is determined using the elbow method that shows the most optimal cluster results formed are as many as three clusters, namely products that sell best-selling, best-selling, and less in-demand. These results are used to provide more promotion to existing products in the cluster less in demand to be able to increase sales. Then a comparison of the results of clustering algorithms with cluster quality tests using the Silhouette Index method which produces a maximum index value of 0.404 on the K-Means algorithm while 0.376 on the K-Medoids algorithm.

Keywords: k-medoids algorithm, k-means algorithm, data mining, clustering, sales data

1. PENDAHULUAN

PT. United Teknologi Integrasi adalah perusahaan yang bergerak pada bidang usaha supplier barang IT. Penentuan strategi penjualan dalam perusahaan merupakan hal yang sangat penting untuk dilakukan. Dengan adanya strategi penjualan yang tepat dapat meningkatkan hasil penjualan perusahaan [1]. Perusahaan ini setiap hari melakukan transaksi penjualan barang IT kepada banyak customer yang setiap data penjualannya sudah tercatat pada sebuah sistem. Banyak data transaksi penjualan barang kepada customer yang telah dihasilkan dari tahun ke tahun tetapi data tersebut tidak dimanfaatkan. Data transaksi ini dapat memiliki potensi untuk dapat dimanfaatkan untuk menentukan strategi penjualan.

Selama ini data transaksi penjualan hanya digunakan untuk melihat report penjualan per periode. Pemanfaatan pengolahan data transaksi dapat memberikan manfaat untuk strategi penjualan yang lebih efektif serta efisien. Dengan memanfaatkan data mining, data yang diolah dapat memberikan informasi tersembunyi yang ada dalam data tersebut [2].

Salah satu metode pada data mining adalah clustering yang berguna untuk mengelompokkan data. Teknik clustering ini digunakan untuk melakukan pengelompokan data yang menghasilkan data barang yang kurang laris, laris dan terlaris dijual. Algoritma K-Means dan K-Medoids merupakan algoritma clustering yang dapat digunakan untuk melakukan pengelompokan data ke dalam cluster dengan karakteristik yang mirip atau sama [3].

Pada penelitian ini, dilakukan penerapan algoritma K-Means dan K-Medoids pada dataset transaksi penjualan untuk menentukan segmentasi produk. Kemudian dilakukan perbandingan kedua algoritma untuk mengetahui algoritma mana yang paling efektif untuk diterapkan pada pengolahan data penjualan. Tujuan dari penelitian ini adalah untuk melakukan perbandingan antara Algoritma K-Means dan K-Medoids dalam pengelompokan data penjualan dan memberikan rekomendasi strategi penjualan yang efektif dan efisien berdasarkan hasil pengelompokan data yang terbentuk.

Berbagai penelitian yang telah dilakukan terkait dengan pengelompokan data penjualan maupun yang bukan pengelompokan data penjualan menggunakan teknik clustering. Pada bagian ini disajikan beberapa tinjauan literatur yang berkaitan dengan penelitian ini.

Pada penelitian dengan judul pengelompokan data pembelian tinta dengan menggunakan metode k-means [4] membahas tentang pemanfaatan data mining dengan algoritma k-means. Data yang diolah adalah data pembelian tinta pada PT Mayer Indah Indonesia dari hasil penelitian diperoleh pengelompokan data pembelian tinta dengan menggunakan atribut barang masuk, barang keluar dan stok barang dari jumlah sampel sebanyak 42. Dari hasil iterasi dapat dikelompokkan barang yang paling banyak keluar, barang yang keluaranya sedang dan barang yang sedikit keluaranya. Metode K-Means dapat digunakan untuk memudahkan bagian pembelian dalam menentukan pengelompokan data pembelian tinta dengan tepat dan cepat. Pada penelitian lainnya yang berjudul Penerapan Data Mining Menggunakan Algoritma K-Means untuk Mengetahui Minat Customer di Toko Hijab [1], data pada penelitian ini diperoleh dari hasil observasi berdasarkan kisaran rata-rata brand hijab yang paling diminati oleh muslimah di Indonesia. Penelitian menerapkan algoritma k-means untuk mengetahui minat customer terhadap produk yang mereka jual. Pada perhitungan yang telah dilakukan dengan menggunakan dua buah parameter yaitu transaksi dan jumlah penjualan dengan hasil iterasi mendapatkan nilai rasio sebesar 0,374324132, iterasi dua mendapatkan rasio 0,54301832.

Pada penelitian lainnya yang berjudul Penerapan Algoritma K-Medoids untuk Menentukan Segmentasi [2]. Penelitian ini dilakukan dengan mencakup data transaksi penjualan. Data diolah dengan teknik clustering menggunakan Algoritma K-Medoids untuk menentukan segmentasi pelanggan. Pada penelitian ini dilakukan juga uji kualitas cluster menggunakan Silhouette Index dan Davies Boulbin Index dilakukan untuk menentukan jumlah cluster yang paling optimal. Hasil penelitian ini menunjukkan bahwa jumlah cluster optimal adalah 3 (tiga) cluster dengan nilai maksimum Silhouette Index adalah 0,375 dan nilai minimum

Davies Doublin Index adalah 1,030. Pada penelitian yang berjudul Teknik Data Mining Dalam Clustering Produksi Susu Segar Di Indonesia Dengan Algoritma K-Means [5]. Penelitian ini dilakukan dengan menggunakan metode clustering Algoritma K-Means. Penelitian ini menggunakan data provinsi di Indonesia dimulai dari tahun 2000 sampai tahun 2017 yang diperoleh dari Badan Pusat Statistik. Penelitian ini memiliki tujuan untuk melakukan clustering daerah penghasil susu yang terbagi atas dua cluster. Hasil dari penelitian ini adalah di Cluster berdasarkan 2 kelompok daerah penghasil susu, yaitu daerah penghasil susu tinggi dan daerah penghasil susu rendah. Dari 27 data produksi susu segar di Indonesia dapat diketahui, 2 provinsi cluster tingkat tinggi yakni : Jawa Barat dan Jawa Timur. Dan 25 lainnya serta di tambah dengan 7 provinsi yang tidak mengikuti perhitungan Algoritma K-Means Clustering termasuk ke dalam cluster tingkat rendah.

Pada penelitian lainnya yang berjudul Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering Di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo [6]. Pada penelitian ini menggunakan data transaksi pelayanan pasien di Rumah Sakit Anwar Medika. Hasil dari penelitian ini menghasilkan kolom 4 cluster yang terdiri dari kecamatan, diagnosa penyakit, usia dan jenis kelamin. Pengelompokan data rekam medis pasien ini menghasilkan informasi baru mengenai pola pengelompokan penyebaran penyakit di setiap kecamatan berdasarkan data rekam medis pasien dari rumah sakit anwar medika sebanyak 534 data dengan waktu penyelesaian sebanyak 0.06 detik. Penelitian lainnya yang berjudul Akurasi Pemberian Insentif Menggunakan Algoritma K-Medoids Terhadap Tingkat Kedisiplinan Pegawai [7]. Penelitian ini juga membahas terakrit penerapan metode clustering dengan Algoritma K-Medoids. Data yang digunakan untuk diolah pada penelitian ini berupa data rekapitulasi kehadiran sebanyak 25 orang pegawai sebagai sampel. Hasil dari penelitian ini berupa pengelompokan pegawai menjadi 3 kelompok yang mempunyai tingkat kedisiplinan baik berjumlah 12 pegawai, tingkat kedisiplinan cukup berjumlah 8 pegawai, dan tingkat kedisiplinan kurang berjumlah 5 pegawai. Berdasarkan hasil pengelompokan yang telah dihasilkan dapat menjadi pertimbangan bagi pimpinan untuk menentukan jumlah insentif untuk pegawai.

Dari berbagai literatur yang di review di atas, terdapat beberapa algoritma penerapan metode clustering yang digunakan diantaranya adalah Algoritma K-Means dan Algoritma K-Medoids. Dengan dasar ini, maka dilakukan penelitian untuk membandingkan Algoritma K-Means dan Algoritma K-Medoids menggunakan metode silhouette index. Penerapan algoritma ini menggunakan dataset transaksi data penjualan PT. United Teknologi Integrasi yang dikumpulkan dari tanggal 1 Januari 2018 – 31 Desember 2020. Dataset transaksi penjualan dilakukan pengolahan dan analisis untuk menghasilkan clustering produk yang terlaris, laris dan kurang laris yang bermanfaat untuk menentukan strategi penjualan.

2. METODE PENELITIAN

Tahapan penelitian merupakan langkah-langkah yang akan dilakukan peneliti dalam melakukan penelitian. Skema alur tahapan yang digunakan pada penelitian ini ditunjukkan pada Gambar 1. Penelitian ini menggunakan bahasa pemrograman python dengan tool Anaconda Jupyter Notebook untuk melakukan pengolahan data. Adapun tahap penelitian penulis melakukan tahapan-tahapan penelitian dari awal sebagai berikut:



Gambar 1. Skema Alur Tahapan Penelitian

2.1 Mendefinisikan Ruang Lingkup Masalah

Pada tahap mendefinisikan ruang lingkup untuk mendefinisikan rumusan dan batasan masalah yang akan diteliti, ruang lingkup penelitian ini akan berfokus pada penggunaan algoritma clustering K-Means dan K-Medoids serta melakukan perbandingan antara kedua algoritma tersebut pada data penjualan. Perbandingan dilakukan dengan membandingkan akurasi clustering yang telah terbentuk menggunakan metode Silhouette Coefficient.

2.2 Menganalisa Masalah

Pada tahap menganalisa masalah dilakukan pemahaman masalah yang telah ditentukan ruang lingkup atau batasannya, didapatkan analisa masalah yang akan dibahas pada penelitian ini adalah melakukan clustering data penjualan yang selama ini hanya digunakan sebagai report penjualan. Data ini dapat dimanfaatkan untuk mendapatkan informasi terkait barang yang terjual dengan melakukan clustering. Kemudian dilakukan juga perbandingan antara dua algoritma guna mengetahui keefektifan algoritma tersebut dan algoritma mana yang paling tepat digunakan dalam melakukan clustering pada data penjualan.

2.3 Menemukan Tujuan

Pada tahap menemukan tujuan ini berdasarkan pemahaman dari masalah yang dibahas, maka ditentukan tujuan yang akan dicapai dari penelitian ini adalah melakukan perbandingan algoritma clustering K-Means dan K-Medoids pada data penjualan perusahaan, mendapatkan hasil akhir clustering data penjualan beserta perbandingan kinerja algoritma clustering dengan metode silhouette index, kemudian menentukan algoritma yang paling tepat untuk digunakan

pada clustering data penjualan.

2.4 Mempelajari Literatur

Pada tahap mempelajari literatur, dilakukan pembelajaran dari literatur yang telah ada untuk dapat dijadikan dasar atau rujukan dalam penelitian ini. Di penelitian ini menggunakan beberapa studi literatur yang didapat berupa bahan-bahan yang diterbitkan secara rutin contohnya jurnal, buku dan literatur lainnya guna mendukung dasar dan sebagai rujukan dalam penelitian ini

2.5 Mengumpulkan Data

Pada tahap pengumpulan data dilakukan observasi yaitu pengamatan secara langsung ditempat penelitian sehingga permasalahan yang ada dapat diketahui dengan jelas. Kemudian dilakukan wawancara yang bertujuan untuk mendapatkan informasi atau data yang dibutuhkan. Selain itu, juga dilakukan studi kepustakaan yaitu dengan membaca buku-buku yang menunjang dalam melakukan penganalisaan terhadap data dan informasi yang didapat. Diketahui bahwa data yang dimiliki perusahaan hanya diolah sebagai laporan penjualan dan belum dimanfaatkan untuk mendapatkan informasi produk terlaris, laris dan kurang laris. Data yang digunakan untuk clustering adalah data transaksi penjualan PT. United Teknologi Integrasi pada periode tahun 2018 – 2020 sebanyak 7.158 record data.

2.6 Melakukan Persiapan Data

Pada tahap persiapan data dilakukan pembersihan data, pemilihan data, dan transformasi data. Tahap pembersihan data pada data awal digunakan untuk memastikan setiap dataset telah lengkap untuk dapat dipersiapkan agar dapat diolah sesuai dengan kebutuhan. Tahapan pembersihan data dimaksudkan untuk memastikan tidak ada kolom yang bernilai kosong ataupun tidak lengkap. [2] Kemudian tahap selanjutnya yaitu tahap pemilihan atribut data, dari jumlah atribut awal sebanyak 14 atribut dipilih 6 atribut (dengan simbol ✓) yang digunakan untuk keperluan pengolahan data. Pemilihan atribut yang digunakan ditampilkan pada Tabel 1 Pemilihan Atribut, Data tersebut sudah siap untuk digunakan sebagai pengolahan data.

Tabel 1. Pemilihan Atribut

Atribut	Penggunaan	Atribut	Penggunaan
No	x	Quantity	✓
SOD No	x	Unit	x
Date	x	Price	x
Customer POD	x	Discount	x
Curr	x	Total	x
Item	x	Nama Perusahaan	x

Tahapan transformasi data dilakukan untuk mengubah data menjadi sesuai dengan format untuk diolah dalam proses data mining.[2] Pengubahan data ini dilakukan dengan proses konversi data yang masih berupa label menjadi numerik dikarenakan pada komputasi clustering hanya dapat melakukan pengolahan data yang bersifat numerik. Kemudian setelah tahapan transformasi data, dilakukan normalisasi data menggunakan normalisasi MinMax agar data memiliki nilai antara 0 sampai 1 sehingga tidak memiliki nilai yang terlalu jauh.

2.7 Clustering Algoritma

Clustering merupakan pengelompokan terhadap record, yang berfungsi memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Clustering dapat juga dikatakan sebagai identifikasi kelas objek yang memiliki kemiripan. Dengan menggunakan teknik clustering kita bisa lebih lanjut mengidentifikasi kepadatan dan jarak daerah dalam objek ruang dan dapat menemukan secara menyeluruh pola distribusi dan korelasi antara atribut. Penerapan metode clustering menghasilkan pengetahuan berupa penentuan beberapa cluster catatan data

yang memiliki kemiripan atribut. [8]

Terdapat berbagai metode yang dapat digunakan untuk mengukur nilai kesamaan antar objek-objek yang dibandingkan. Salah satunya ialah dengan Euclidean Distance. [9] Euclidean Distance menghitung jarak dua buah point dengan mengetahui nilai masing-masing atribut pada kedua poin tersebut. Jarak adalah pendekatan yang umum dipakai untuk menentukan kesamaan atau ketidaksamaan dua vektor fitur yang dinyatakan dengan ranking. Apabila nilai ranking yang dihasilkan semakin kecil nilainya maka semakin dekat/tinggi kesamaan antara kedua vektor tersebut. Teknik pengukuran jarak dengan metode Euclidean menjadi salah satu metode yang paling umum digunakan. Pengukuran jarak dengan metode Euclidean dapat dituliskan dengan persamaan berikut(1):[5][10][11]

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (1)$$

p,q = dua titik di ruang-n Euclidean
 q_i, p_i = vector Euclidean, dimana dari asal ruang (titik awal)
 n = ruang-n

2.7.1 Algoritma K-Means

Algoritma K-Means adalah metode clustering berbasis jarak yang membagi data ke dalam sejumlah cluster dan algoritma ini hanya bekerja pada atribut numerik. Algoritma K-Means termasuk partitioning clustering yang memisahkan data ke k daerah bagian yang terpisah. K-Means merupakan salah satu dari beberapa metode data clustering non hierarki dengan sistem kerja mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/kelompok. Pembagian data ke dalam cluster/kelompok pada metode ini menggunakan data dengan karakteristik yang sama yang dikelompokkan ke dalam satu cluster yang sama.

Adapun tahapan yang dilakukan untuk melakukan pengelompokan dengan metode K-Means Clustering yaitu :[5][10][11][12]

- a. Tentukan k sebagai jumlah cluster yang ingin dibentuk
- b. Inisialisasi k centroid (titik pusat Cluster) awal secara random.
- c. Alokasikan setiap data atau objek ke Cluster terdekat. Jarak antar objek dan jarak antara objek dengan Cluster tertentu ditentukan. Jarak antara data dengan pusat Cluster menggunakan teori jarak euclidean yang dirumuskan sebagai persamaan berikut(2):[13]

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (2)$$

- d. Jarak pusat Cluster dihitung kembali dengan keanggotaan Cluster yang sekarang. Pusat Cluster adalah rata-rata dari semua data atau objek dalam Cluster tertentu, jika dikehendaki dapat juga digunakan nilai median dari Cluster tersebut.
- e. Ulangi langkah ketiga hingga hasil iterasinya bernilai sama dengan iterasi sebelumnya atau ketika tidak ada lagi data yang berpindah Cluster.

2.7.2 Algoritma K-Medoids

Algoritma K-Medoids atau sering dikenal sebagai PAM (Partitioning Around Medoids), merupakan algoritma yang menerapkan objek sebagai perwakilan di setiap cluster yaitu medoid. Algoritma K-Medoids dapat meminimalkan jumlah perbedaan antara titik data dalam cluster dengan titik data terpilih di setiap cluster sebagai pusat (medoid). K-Medoids memiliki karakteristik dimana pusat cluster berada di antara titik-titik datanya. Penggunaan Euclidean Distance untuk menghitung jarak antara kedekatan objek dengan pusat (medoid) sehingga kemungkinan besar objek yang dipilih secara acak mirip dengan objek medoidnya. Tahapan penyelesaian K-Medoids adalah sebagai berikut: [2][7]

- a. Melakukan inisialisasi pusat cluster sebanyak jumlah cluster (k).
- b. Mengalokasikan setiap objek ke cluster terdekat menggunakan Euclidean Distance dengan rumus persamaan (3).

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \tag{3}$$

- c. Memilih objek secara acak pada masing-masing cluster sebagai calon medoid yang baru.
- d. Menghitung jarak setiap objek pada masing-masing cluster dengan calon medoid yang baru.
- e. Total simpangan (S) didapatkan dengan menghitung nilai total jarak yang baru dikurangi total jarak yang lama. Jika nilai S kurang dari 0, maka dilakukan penukaran objek dengan data cluster untuk membentuk sekumpulan jumlah cluster objek yang baru sebagai medoid.
- f. Mengulangi tahap c sampai dengan e apabila masih terjadi perubahan medoid, jika tidak ada perubahan maka diperoleh cluster serta anggota dari cluster masing-masing.

2.8 Analisis Hasil Clustering

Pada tahap analisis hasil clustering ini, dilakukan analisis terhadap hasil cluster yang terbentuk untuk mengetahui produk yang terlaris, laris dan kurang laris terjual. Serta dilakukan evaluasi hasil cluster menggunakan metode silhouette index. SI adalah teknik validasi untuk mengukur seberapa dekat setiap objek dalam satu cluster dan objek di cluster lain serta mengetahui seberapa baik objek terletak dalam clusternya. Nilai SI berada di antara -1 hingga +1 dan hasil nilai maksimum menunjukkan jumlah cluster yang optimal.[2][14]

2.9 Kesimpulan

Pada tahap kesimpulan ini, ditarik kesimpulan akhir dari hasil clustering yang telah terbentuk oleh kedua algoritma yang digunakan, serta perbandingan antara kedua algoritma yang telah dilakukan. Sehingga menghasilkan kesimpulan berupa algoritma mana yang paling sesuai untuk digunakan dalam melakukan clustering data penjualan serta produk yang telah dikelompokkan terjual terlaris, laris dan kurang laris..

2. HASIL DAN PEMBAHASAN

Dataset transaksi penjualan sebanyak 7.158 record data yang telah dikumpulkan dilakukan persiapan data dengan membersihkan data yang bernilai null. Hasil dari eksperimen menunjukkan hasil data transaksi penjualan tidak memiliki nilai yang bernilai null tersaji pada Tabel 2.

Tabel 2. Sumber Data

	Quantity	Total	Nama Perusahaan	Item
0	2	16218917.0	PT. AA INTERNATIONAL INDONESIA	Alibaba Cloud CDN
1	1	143550.0	PT. AA INTERNATIONAL INDONESIA	Service1
2	1	640712.0	PT. AA INTERNATIONAL INDONESIA	Alibaba Cloud CDN
3	1	550000.0	PT. AA INTERNATIONAL INDONESIA	Installation
4	1	143550.0	PT. ABM INVESTAMA TBK	Service1
5	1	12298000.0	PT. ABM INVESTAMA TBK	NB Dell Latitude 3480 (Core i5)
6	1	3542000.0	PT. ABM INVESTAMA TBK	Microsoft Visio Std 2016 SNGL OLP NL

7	1	6732000.0	PT. ABM INVESTAMA TBK	Acrobat Profesional 2017 Multiple Platfrom
8	1	3850000.0	PT. ABM INVESTAMA TBK	Printer Epson WF100
9	1	1859000.0	PT. ABM INVESTAMA TBK	Battery Dell PRI, 42WHR, 3C

Pada tahap selanjutnya dilakukan pemilihan data yaitu pemilihan atribut-atribut pada dataset untuk dapat menghasilkan data yang siap untuk dilakukan pengolahan. Dataset transaksi penjualan yang akan diolah memiliki dimensi 7.158 baris x 6 kolom. Data yang digunakan terdiri dari data quantity produk, total harga produk, nama perusahaan, dan item. Quantity produk adalah data yang menyajikan jumlah barang yang terjual, Total harga adalah total dari harga setiap item barang, Nama perusahaan adalah data nama perusahaan yang membeli item produk, dan Item adalah nama produk yang dibeli oleh customer. Tahap transformasi data dilakukan untuk mengubah data sesuai dengan format untuk diolah agar dapat digunakan dalam proses data mining. Transformasi data diawali dengan proses konversi data yang masih berupa kategori label menjadi kode numerik dilakukan pada atribut Nama Perusahaan dan Item. Hasil dari tahapan transformasi data transaksi penjualan disajikan pada Tabel 3. Kemudian dilakukan tahapan normalisasi data, dalam penelitian ini menggunakan normalisasi Min-Max scaler dari sklearn. Hasil normalisasi data merupakan hasil akhir dari tahapan persiapan data disajikan pada Tabel 3.

Tabel 3. Transformasi Data

	Quantity	Total	Nama Perusahaan	Item
0	2	16218917.0	41	103
1	1	143550.0	41	2000
2	1	640712.0	41	103
3	1	550000.0	41	793
4	1	143550.0	42	2000
5	1	12298000.0	42	1329
6	1	3542000.0	42	1213
7	1	6732000.0	42	78
8	1	3850000.0	42	1732
9	1	1859000.0	42	170

Setelah dilakukan normalisasi data, data tersebut akan digunakan dalam proses clustering dengan menggunakan algoritma K-Means dan K-Medoids.

3.1 Penerapan Algoritma K-Means Clustering

1. Tentukan k sebagai jumlah cluster yang ingin dibentuk.

Langkah awal metode k-means yaitu menentukan jumlah cluster yang di inisiasikan dengan “k”. Dalam penelitian ini ditentukan jumlah k adalah 3 sebagai titik pusat cluster.

2. Inisiasi k centroid (titik pusat cluster) awal secara random.[15]

Dilakukan pemilihan pusat cluster untuk melakukan iterasi pertama, titik pusat cluster disajikan pada Tabel 4.

Tabel 4. Titik pusat cluster awal

	Quantity	Total	Nama Perusahaan	Item
0	2	16218917.0	41	103
1	1	6732000.0	42	78
2	1	143550	41	2000

3. Alokasikan setiap data atau objek ke cluster terdekat.
Perhitungan jarak dilakukan dengan menggunakan metode Euclidean Distance sesuai pada persamaan nomor 2 yang tersaji pada Tabel 5.

Tabel 5. Peritungan jarak Iterasi 1 Algoritma K Means

Perhitungan Jarak	
<p>Iterasi 1 Cluster 1</p> $D_{(1,1)} = \sqrt{(2-2)^2 + (16218917 - 16218917)^2 + (41-41)^2 + (103-103)^2} = 0$ $D_{(1,2)} = \sqrt{(1-2)^2 + (143550 - 16218917)^2 + (41-41)^2 + (200-103)^2} = 16075367.11$ $D_{(1,3)} = \sqrt{(1-2)^2 + (640712 - 16218917)^2 + (41-41)^2 + (103-103)^2} = 15578205$ $D_{(1,4)} = \sqrt{(1-2)^2 + (550000 - 16218917)^2 + (41-41)^2 + (793-103)^2} = 15668917.02$ $D_{(1,5)} = \sqrt{(1-2)^2 + (143550 - 16218917)^2 + (41-41)^2 + (2000-103)^2} = 16075367.11$ $D_{(1,6)} = \sqrt{(1-2)^2 + (12298000 - 16218917)^2 + (42-41)^2 + (1329-103)^2} = 39201917.192$ $D_{(1,7)} = \sqrt{(1-2)^2 + (3542000 - 16218917)^2 + (42-41)^2 + (1213-103)^2} = 12676917.05$ $D_{(1,8)} = \sqrt{(1-2)^2 + (6732000 - 16218917)^2 + (42-41)^2 + (78-103)^2} = 9486917$ $D_{(1,9)} = \sqrt{(1-2)^2 + (3850000 - 16218917)^2 + (42-41)^2 + (1732-103)^2} = 12368917.11$ $D_{(1,10)} = \sqrt{(1-2)^2 + (1690000 - 16218917)^2 + (42-41)^2 + (170-103)^2} = 14528917$	<p>Iterasi 1 Cluster 2</p> $D_{(1,1)} = \sqrt{(2-1)^2 + (16218917 - 6732000)^2 + (41-42)^2 + (103-78)^2} = 9486917$ $D_{(1,2)} = \sqrt{(1-1)^2 + (143550 - 6732000)^2 + (41-42)^2 + (2000-78)^2} = 6588450.28$ $D_{(1,3)} = \sqrt{(1-1)^2 + (640712 - 6732000)^2 + (41-42)^2 + (103-78)^2} = 6091288$ $D_{(1,4)} = \sqrt{(1-1)^2 + (550000 - 6732000)^2 + (41-42)^2 + (793-78)^2} = 6182000.041$ $D_{(1,5)} = \sqrt{(1-1)^2 + (143550 - 6732000)^2 + (41-42)^2 + (2000-78)^2} = 6588450.28$ $D_{(1,6)} = \sqrt{(1-1)^2 + (12298000 - 6732000)^2 + (42-42)^2 + (1329-78)^2} = 5566000.141$ $D_{(1,7)} = \sqrt{(1-1)^2 + (3542000 - 6732000)^2 + (42-42)^2 + (1213-78)^2} = 3190000.202$ $D_{(1,8)} = \sqrt{(1-1)^2 + (6732000 - 6732000)^2 + (42-42)^2 + (78-78)^2} = 0$ $D_{(1,9)} = \sqrt{(1-1)^2 + (3850000 - 6732000)^2 + (42-42)^2 + (1732-78)^2} = 2882000.475$ $D_{(1,10)} = \sqrt{(1-1)^2 + (1690000 - 6732000)^2 + (42-42)^2 + (170-78)^2} = 5042000.001$
<p>Iterasi 1 Cluster 3</p> $D_{(1,1)} = \sqrt{(2-1)^2 + (16218917 - 143550)^2 + (41-41)^2 + (103-2000)^2} = 16075367.11$ $D_{(1,2)} = \sqrt{(1-1)^2 + (143550 - 143550)^2 + (41-41)^2 + (2000-2000)^2} = 0$ $D_{(1,3)} = \sqrt{(1-1)^2 + (640712 - 143550)^2 + (41-41)^2 + (103-2000)^2} = 497165.6191$ $D_{(1,4)} = \sqrt{(1-1)^2 + (550000 - 143550)^2 + (41-41)^2 + (793-2000)^2} = 406451.7922$ $D_{(1,5)} = \sqrt{(1-1)^2 + (143550 - 143550)^2 + (41-41)^2 + (2000-2000)^2} = 0$ $D_{(1,6)} = \sqrt{(1-1)^2 + (12298000 - 143550)^2 + (42-41)^2 + (1329-2000)^2} = 12154450.02$ $D_{(1,7)} = \sqrt{(1-1)^2 + (3542000 - 143550)^2 + (42-41)^2 + (1213-2000)^2} = 3398450.091$ $D_{(1,8)} = \sqrt{(1-1)^2 + (6732000 - 143550)^2 + (42-41)^2 + (78-2000)^2} = 6588450.28$ $D_{(1,9)} = \sqrt{(1-1)^2 + (3850000 - 143550)^2 + (42-41)^2 + (1732-2000)^2} = 3706450.01$ $D_{(1,10)} = \sqrt{(1-1)^2 + (1690000 - 143550)^2 + (42-41)^2 + (170-2000)^2} = 1546451.083$	

Didapatkan hasil Perhitungan jarak dilakukan dengan menggunakan metode Euclidean Distance, dengan cara menggunakan data yang telah diolah. Menghitung dengan melakukan pengurangan data pusat cluster dengan setiap data sebaran cluster kemudian dijumlahkan seluruhnya dan diakar pangkat dua sesuai dengan rumus Euclidean Distance data yang digunakan adalah data quantity, Total dan Item pusat cluster dikurangi dengan quantity sebaran data lainnya. Setelah itu dilakukan perhitungan jarak dengan menggunakan Euclidean Distance dan mendapatkan hasil alokasi data ke pusat cluster terlampir pada Tabel 6.

Tabel 6. Alokasi Data ke Cluster Berdasarkan Jarak Setiap Data dengan Pusat Cluster

	Quantity	Total	Nama Perusahaan	Item	Cluster
0	2	16218917.0	41	103	1
1	1	143550.0	41	2000	3
2	1	640712.0	41	103	3
3	1	550000.0	41	793	3
4	1	143550.0	42	2000	3
5	1	12298000.0	42	1329	1
6	1	3542000.0	42	1213	2
7	1	6732000.0	42	78	2
8	1	3850000.0	42	1732	2
9	1	1859000.0	42	170	3

Dari alokasi data yang ada pada Tabel 5 dilakukan perhitungan titik cluster baru yang terbentuk dengan menghitung rata rata nilai setiap cluster. Pada Tabel 6. Disajikan hasil dari alokasi setiap sebaran data sehingga menghasilkan anggota cluster baru berdasarkan jarak sebaran data dengan pusat cluster yang terdekat. Dari hasil tersebut kemudian dihitung kembali rata rata nilai setiap cluster dengan pusat cluster baru yang terbentuk. Kemudian terbentuk titik cluster baru yang terjadi pada Tabel 7. Cluster baru yang terbentuk ini berguna untuk melakukan perhitungan jarak data ke pusat cluster pada iterasi 2.

Tabel 7. Titik Cluster Baru yang Terbentuk

	Quantity	Total	Nama Perusahaan	Item	Cluster
0	1.5	14258458.5	41.5	716	1
1	1	4708000	42	1007.67	2
2	1	633562.4	41.2	1013.2	3

Ulangi langkah ketiga hingga hasil iterasinya bernilai sama dengan iterasi sebelumnya atau ketika tidak ada lagi data yang berpindah Cluster. Perhitungan jarak kembali dilakukan dengan menggunakan metode Euclidean Distance sesuai pada persamaan nomor 2 yang tersaji pada Tabel 8.

Tabel 8. Perhitungan jarak Iterasi 2 Algoritma K Means

Perhitungan Jarak	
Iterasi 2 Cluster 1	Iterasi 2 Cluster 2
$D_{(1,1)} = \sqrt{(2 - 1.5)^2 + (16218917 - 14258458.5)^2 + (41 - 41.5)^2 + (103 - 716)^2} = 1960458.596$	$D_{(1,1)} = \sqrt{(2 - 1)^2 + (16218917 - 4708000)^2 + (41 - 42)^2 + (103 - 1007.67)^2} = 11510917.04$
$D_{(1,2)} = \sqrt{(1 - 1.5)^2 + (143550 - 14258458.5)^2 + (41 - 41.5)^2 + (2000 - 716)^2} = 14114908.56$	$D_{(1,2)} = \sqrt{(1 - 1)^2 + (143550 - 4708000)^2 + (41 - 42)^2 + (2000 - 1007.67)^2} = 4564450.108$
$D_{(1,3)} = \sqrt{(1 - 1.5)^2 + (640712 - 14258458.5)^2 + (41 - 41.5)^2 + (103 - 716)^2} = 13617746.51$	$D_{(1,3)} = \sqrt{(1 - 1)^2 + (640712 - 4708000)^2 + (41 - 42)^2 + (103 - 1007.67)^2} = 4067288.101$
$D_{(1,4)} = \sqrt{(1 - 1.5)^2 + (550000 - 14258458.5)^2 + (41 - 41.5)^2 + (793 - 716)^2} = 13708458.5$	$D_{(1,4)} = \sqrt{(1 - 1)^2 + (550000 - 4708000)^2 + (41 - 42)^2 + (793 - 1007.67)^2} = 4158000.006$
$D_{(1,5)} = \sqrt{(1 - 1.5)^2 + (143550 - 14258458.5)^2 + (41 - 41.5)^2 + (2000 - 716)^2} = 14114908.56$	$D_{(1,5)} = \sqrt{(1 - 1)^2 + (143550 - 4708000)^2 + (41 - 42)^2 + (2000 - 1007.67)^2} = 4564450.108$

$D_{(1,6)} = \sqrt{(1-1.5)^2 + (12298000 - 14258458.5)^2 + (42 - 41.5)^2 + (1329 - 716)^2} = 1960458.596$	$D_{(1,6)} = \sqrt{(1-1)^2 + (12298000 - 4708000)^2 + (42 - 42)^2 + (1329 - 1007.67)^2} = 7590000.007$
$D_{(1,7)} = \sqrt{(1-1.5)^2 + (3542000 - 14258458.5)^2 + (42 - 41.5)^2 + (1213 - 716)^2} = 10716458.51$	$D_{(1,7)} = \sqrt{(1-1)^2 + (3542000 - 4708000)^2 + (42 - 42)^2 + (1213 - 1007.67)^2} = 1166000.018$
$D_{(1,8)} = \sqrt{(1-1.5)^2 + (6732000 - 14258458.5)^2 + (42 - 41.5)^2 + (78 - 716)^2} = 7526458.527$	$D_{(1,8)} = \sqrt{(1-1)^2 + (6732000 - 4708000)^2 + (42 - 42)^2 + (78 - 1007.67)^2} = 2024000.214$
$D_{(1,9)} = \sqrt{(1-1.5)^2 + (3850000 - 14258458.5)^2 + (42 - 41.5)^2 + (1732 - 716)^2} = 10408458.55$	$D_{(1,9)} = \sqrt{(1-1)^2 + (3850000 - 4708000)^2 + (42 - 42)^2 + (1732 - 1007.67)^2} = 858000.3057$
$D_{(1,10)} = \sqrt{(1-1.5)^2 + (1690000 - 14258458.5)^2 + (42 - 41.5)^2 + (170 - 716)^2} = 12568458.51$	$D_{(1,10)} = \sqrt{(1-1)^2 + (1690000 - 4708000)^2 + (42 - 42)^2 + (170 - 1007.67)^2} = 3018000.116$

Iterasi 2 Cluster 3

$D_{(1,1)} = \sqrt{(2-1)^2 + (16218917 - 633562.4)^2 + (41 - 41.5)^2 + (103 - 1013.2)^2} = 15585354.63$
$D_{(1,2)} = \sqrt{(1-1)^2 + (143550 - 633562.4)^2 + (41 - 41.5)^2 + (2000 - 1013.2)^2} = 490013.3936$
$D_{(1,3)} = \sqrt{(1-1)^2 + (640712 - 633562.4)^2 + (41 - 41.5)^2 + (103 - 1013.2)^2} = 7207.304922$
$D_{(1,4)} = \sqrt{(1-1)^2 + (550000 - 633562.4)^2 + (41 - 41.5)^2 + (793 - 1013.2)^2} = 83562.69013$
$D_{(1,5)} = \sqrt{(1-1)^2 + (143550 - 633562.4)^2 + (41 - 41.5)^2 + (2000 - 1013.2)^2} = 490013.3936$
$D_{(1,6)} = \sqrt{(1-1)^2 + (12298000 - 633562.4)^2 + (42 - 41.5)^2 + (1329 - 1013.2)^2} = 11664437.6$
$D_{(1,7)} = \sqrt{(1-1)^2 + (3542000 - 633562.4)^2 + (42 - 41.5)^2 + (1213 - 1013.2)^2} = 2908437.607$
$D_{(1,8)} = \sqrt{(1-1)^2 + (6732000 - 633562.4)^2 + (42 - 41.5)^2 + (78 - 1013.2)^2} = 6098437.672$
$D_{(1,9)} = \sqrt{(1-1)^2 + (3850000 - 633562.4)^2 + (42 - 41.5)^2 + (1732 - 1013.2)^2} = 3216437.68$
$D_{(1,10)} = \sqrt{(1-1)^2 + (169000 - 633562.4)^2 + (42 - 41.5)^2 + (170 - 1013.2)^2} = 1056437.937$

Dilakukan perhitungan jarak setiap data dengan menggunakan Euclidean Distance dan mendapatkan hasil alokasi data ke pusat cluster baru tersaji pada Tabel 9.

Tabel 9. Alokasi Data ke Cluster Berdasarkan Jarak Setiap Data dengan Pusat Cluster

	Quantity	Total	Nama Perusahaan	Item	Cluster	Status
0	2	16218917.0	41	103	1	Tetap
1	1	143550.0	41	2000	3	Tetap
2	1	640712.0	41	103	3	Tetap
3	1	550000.0	41	793	3	Tetap
4	1	143550.0	42	2000	3	Tetap
5	1	12298000.0	42	1329	1	Tetap
6	1	3542000.0	42	1213	2	Tetap
7	1	6732000.0	42	78	2	Tetap
8	1	3850000.0	42	1732	2	Tetap
9	1	1859000.0	42	170	3	Tetap

Hasil cluster pada iterasi yang baru menunjukkan hasil cluster yang sama dengan hasil sebelumnya maka iterasi cluster selesai,[16][17] proses clustering berakhir dan dilanjutkan untuk pengujian kualitas cluster.

3.2 Penerapan Algoritma K-Medoids Clustering

- Melakukan inisialisasi pusat cluster sebanyak jumlah cluster (k). Langkah awal metode k-means yaitu menentukan jumlah cluster yang di inisiasikan dengan “k”. Dalam penelitian ini ditentukan cluster 1 adalah produk dengan penjualan

terlaris, cluster 2 adalah produk dengan penjualan laris, dan cluster 3 adalah produk dengan penjualan kurang laris.

- Inisiasi k centroid (titik pusat cluster) awal secara random.
Dilakukan pemilihan pusat cluster untuk melakukan iterasi pertama, titik pusat cluster disajikan pada Tabel 10.

Tabel 10. Titik pusat cluster awal

	Quantity	Total	Nama Perusahaan	Item
0	2	16218917.0	41	103
1	1	6732000.0	42	78
2	1	143550	41	2000

- Mengalokasikan setiap objek ke cluster terdekat.
Perhitungan jarak dilakukan dengan menggunakan metode Euclidean Distance sesuai pada persamaan nomor 3 yang tersaji pada Tabel 11.

Tabel 11. Perhitungan jarak Iterasi 1 Algoritma K Medoids

Perhitungan Jarak	
Medoids 1 Cluster 1	Medoids 1 Cluster 2
$D_{(1,1)}$ $= \sqrt{(2-2)^2 + (16218917 - 16218917)^2 + (41-41)^2 + (103-103)^2} = 0$	$D_{(1,1)}$ $= \sqrt{(2-1)^2 + (16218917 - 6732000)^2 + (41-42)^2 + (103-78)^2} = 9486917$
$D_{(1,2)}$ $= \sqrt{(1-2)^2 + (143550 - 16218917)^2 + (41-41)^2 + (200-103)^2} = 16075367.11$	$D_{(1,2)}$ $= \sqrt{(1-1)^2 + (143550 - 6732000)^2 + (41-42)^2 + (2000-78)^2} = 6588450.28$
$D_{(1,3)}$ $= \sqrt{(1-2)^2 + (640712 - 16218917)^2 + (41-41)^2 + (103-103)^2} = 15578205$	$D_{(1,3)}$ $= \sqrt{(1-1)^2 + (640712 - 6732000)^2 + (41-42)^2 + (103-78)^2} = 6091288$
$D_{(1,4)}$ $= \sqrt{(1-2)^2 + (550000 - 16218917)^2 + (41-41)^2 + (793-103)^2} = 15668917.02$	$D_{(1,4)}$ $= \sqrt{(1-1)^2 + (550000 - 6732000)^2 + (41-42)^2 + (793-78)^2} = 6182000.041$
$D_{(1,5)}$ $= \sqrt{(1-2)^2 + (143550 - 16218917)^2 + (41-41)^2 + (2000-103)^2} = 16075367.11$	$D_{(1,5)}$ $= \sqrt{(1-1)^2 + (143550 - 6732000)^2 + (41-42)^2 + (2000-78)^2} = 6588450.28$
$D_{(1,6)}$ $= \sqrt{(1-2)^2 + (12298000 - 16218917)^2 + (42-41)^2 + (1329-103)^2} = 39201917.192$	$D_{(1,6)}$ $= \sqrt{(1-1)^2 + (12298000 - 6732000)^2 + (42-42)^2 + (1329-78)^2} = 5566000.141$
$D_{(1,7)}$ $= \sqrt{(1-2)^2 + (3542000 - 16218917)^2 + (42-41)^2 + (1213-103)^2} = 12676917.05$	$D_{(1,7)}$ $= \sqrt{(1-1)^2 + (3542000 - 6732000)^2 + (42-42)^2 + (1213-78)^2} = 3190000.202$
$D_{(1,8)}$ $= \sqrt{(1-2)^2 + (6732000 - 16218917)^2 + (42-41)^2 + (78-103)^2} = 9486917$	$D_{(1,8)}$ $= \sqrt{(1-1)^2 + (6732000 - 6732000)^2 + (42-42)^2 + (78-78)^2} = 0$
$D_{(1,9)}$ $= \sqrt{(1-2)^2 + (3850000 - 16218917)^2 + (42-41)^2 + (1732-103)^2} = 12368917.11$	$D_{(1,9)}$ $= \sqrt{(1-1)^2 + (3850000 - 6732000)^2 + (42-42)^2 + (1732-78)^2} = 2882000.475$
$D_{(1,10)}$ $= \sqrt{(1-2)^2 + (1690000 - 16218917)^2 + (42-41)^2 + (170-103)^2} = 14528917$	$D_{(1,10)}$ $= \sqrt{(1-1)^2 + (1690000 - 6732000)^2 + (42-42)^2 + (170-78)^2} = 5042000.001$
Iterasi 1 Medoids 3	
$D_{(1,1)} = \sqrt{(2-1)^2 + (16218917 - 143550)^2 + (41-41)^2 + (103-2000)^2} = 16075367.11$	
$D_{(1,2)} = \sqrt{(1-1)^2 + (143550 - 143550)^2 + (41-41)^2 + (2000-2000)^2} = 0$	
$D_{(1,3)} = \sqrt{(1-1)^2 + (640712 - 143550)^2 + (41-41)^2 + (103-2000)^2} = 497165.6191$	
$D_{(1,4)} = \sqrt{(1-1)^2 + (550000 - 143550)^2 + (41-41)^2 + (793-2000)^2} = 406451.7922$	
$D_{(1,5)} = \sqrt{(1-1)^2 + (143550 - 143550)^2 + (41-41)^2 + (2000-2000)^2} = 0$	

$$D_{(1,6)} = \sqrt{(1-1)^2 + (12298000 - 143550)^2 + (42-41)^2} + (1329 - 2000)^2 = 12154450.02$$

$$D_{(1,7)} = \sqrt{(1-1)^2 + (3542000 - 143550)^2 + (42-41)^2} + (1213 - 2000)^2 = 3398450.091$$

$$D_{(1,8)} = \sqrt{(1-1)^2 + (6732000 - 143550)^2 + (42-41)^2} + (78 - 2000)^2 = 6588450.28$$

$$D_{(1,9)} = \sqrt{(1-1)^2 + (3850000 - 143550)^2 + (42-41)^2} + (1732 - 2000)^2 = 3706450.01$$

$$D_{(1,10)} = \sqrt{(1-1)^2 + (1690000 - 143550)^2 + (42-41)^2} + (170 - 2000)^2 = 1546451.083$$

Didapatkan hasil Perhitungan jarak dilakukan dengan menggunakan metode Euclidean Distance, dengan cara menggunakan data yang telah diolah. Menghitung dengan melakukan pengurangan data pusat cluster dengan setiap data sebaran cluster kemudian dijumlahkan seluruhnya dan diakar pangkat dua sesuai dengan rumus Euclidean Distance data yang digunakan adalah data quantity, Total dan Item pusat cluster dikurangi dengan quantity sebaran data lainnya. Setelah itu dilakukan perhitungan jarak dengan menggunakan Euclidean Distance dan mendapatkan hasil alokasi data ke pusat cluster terlampir pada Tabel 12.

Tabel 12. Alokasi Data ke Cluster Berdasarkan Jarak Setiap Data dengan Pusat Cluster

	Quantity	Total	Nama Perusahaan	Item	Cluster
0	2	16218917.0	41	103	1
1	1	143550.0	41	2000	3
2	1	640712.0	41	103	3
3	1	550000.0	41	793	3
4	1	143550.0	42	2000	3
5	1	12298000.0	42	1329	1
6	1	3542000.0	42	1213	2
7	1	6732000.0	42	78	2
8	1	3850000.0	42	1732	2
9	1	1859000.0	42	170	3

- Memilih objek secara acak pada masing-masing cluster sebagai calon medoid yang baru.

Tabel 13. Titik Pusat Cluster Medoid Baru

	Quantity	Total	Nama Perusahaan	Item
0	2	16218917.0	41	103
1	1	143550.0	41	2000
2	1	640712.0	41	103

Pada Tabel 13 ditampilkan pemilihan cluster secara acak sebagai medoid baru untuk melakukan iterasi kedua dengan titik cluster baru.

- Menghitung jarak setiap objek pada masing-masing cluster dengan calon medoid yang baru.

Perhitungan jarak kembali dilakukan dengan menggunakan metode Euclidean Distance sesuai pada persamaan nomor 3 yang tersaji pada Tabel 14.

Tabel 14. Perhitungan jarak Iterasi 2 Algoritma K Medoids

Perhitungan Jarak	
Medoids 2 Cluster 1	Medoids 2 Cluster 2
$D_{(1,1)} = \sqrt{(2-2)^2 + (16218917 - 16218917)^2 + (41-41)^2} + (103 - 103)^2 = 0$ $D_{(1,2)} = \sqrt{(1-2)^2 + (143550 - 16218917)^2 + (41-41)^2} + (2000 - 103)^2 = 16075367.11$	$D_{(1,1)} = \sqrt{(2-1)^2 + (16218917 - 143550)^2 + (41-41)^2} + (103 - 2000)^2 = 16075367.11$ $D_{(1,1)} = \sqrt{(1-1)^2 + (143550 - 143550)^2 + (41-41)^2} + (2000 - 2000)^2 = 0$

$D_{(1,3)} = \sqrt{(1-2)^2 + (640712 - 16218917)^2 + (41-41)^2 + (103-103)^2} = 15578205$ $D_{(1,4)} = \sqrt{(1-2)^2 + (550000 - 16218917)^2 + (41-41)^2 + (793-103)^2} = 15668917.02$ $D_{(1,5)} = \sqrt{(1-2)^2 + (143550 - 16218917)^2 + (41-41)^2 + (2000-103)^2} = 16075367.11$ $D_{(1,6)} = \sqrt{(1-2)^2 + (12298000 - 16218917)^2 + (42-41)^2 + (1329-103)^2} = 39201917.192$ $D_{(1,7)} = \sqrt{(1-2)^2 + (3542000 - 16218917)^2 + (42-41)^2 + (1213-103)^2} = 12676917.05$ $D_{(1,8)} = \sqrt{(1-2)^2 + (6732000 - 16218917)^2 + (42-41)^2 + (78-103)^2} = 9486917$ $D_{(1,9)} = \sqrt{(1-2)^2 + (3850000 - 16218917)^2 + (42-41)^2 + (1732-103)^2} = 12368917.11$ $D_{(1,10)} = \sqrt{(1-2)^2 + (1690000 - 16218917)^2 + (42-41)^2 + (170-103)^2} = 14528917$	$D_{(1,1)} = \sqrt{(1-1)^2 + (640712 - 143550)^2 + (41-41)^2 + (103-2000)^2} = 497165.6191$ $D_{(1,1)} = \sqrt{(1-1)^2 + (550000 - 143550)^2 + (41-41)^2 + (793-2000)^2} = 406451.7922$ $D_{(1,1)} = \sqrt{(1-1)^2 + (143550 - 143550)^2 + (41-41)^2 + (2000-2000)^2} = 0$ $D_{(1,1)} = \sqrt{(1-1)^2 + (12298000 - 143550)^2 + (42-41)^2 + (1329-2000)^2} = 12154450.02$ $D_{(1,1)} = \sqrt{(1-1)^2 + (3542000 - 143550)^2 + (42-41)^2 + (1213-2000)^2} = 3398450.091$ $D_{(1,1)} = \sqrt{(1-1)^2 + (6732000 - 143550)^2 + (42-41)^2 + (78-2000)^2} = 6588450.28$ $D_{(1,1)} = \sqrt{(1-1)^2 + (3850000 - 143550)^2 + (42-41)^2 + (1732-2000)^2} = 3706450.01$ $D_{(1,1)} = \sqrt{(1-1)^2 + (1690000 - 143550)^2 + (42-41)^2 + (170-2000)^2} = 1546451.083$
--	--

Medoid 2 Cluster 3

$$D_{(1,1)} = \sqrt{(2-1)^2 + (16218917 - 640712)^2 + (41-41)^2 + (103-103)^2} = 15578205$$

$$D_{(1,2)} = \sqrt{(1-1)^2 + (143550 - 640712)^2 + (41-41)^2 + (2000-103)^2} = 497165.6191$$

$$D_{(1,3)} = \sqrt{(1-1)^2 + (640712 - 640712)^2 + (41-41)^2 + (103-103)^2} = 0$$

$$D_{(1,4)} = \sqrt{(1-1)^2 + (550000 - 640712)^2 + (41-41)^2 + (793-103)^2} = 90714.6242$$

$$D_{(1,5)} = \sqrt{(1-1)^2 + (143550 - 640712)^2 + (41-41)^2 + (2000-103)^2} = 497165.6191$$

$$D_{(1,6)} = \sqrt{(1-1)^2 + (12298000 - 640712)^2 + (42-41)^2 + (1329-103)^2} = 11657288.06$$

$$D_{(1,7)} = \sqrt{(1-1)^2 + (3542000 - 640712)^2 + (42-41)^2 + (1213-103)^2} = 2901288.212$$

$$D_{(1,8)} = \sqrt{(1-1)^2 + (6732000 - 640712)^2 + (42-41)^2 + (78-103)^2} = 6091288$$

$$D_{(1,9)} = \sqrt{(1-1)^2 + (3850000 - 640712)^2 + (42-41)^2 + (1732-103)^2} = 3209288.413$$

$$D_{(1,10)} = \sqrt{(1-1)^2 + (1690000 - 640712)^2 + (42-41)^2 + (170-103)^2} = 1049288.002$$

Didapatkan hasil Perhitungan jarak dilakukan dengan menggunakan metode Euclidean Distance, dengan cara menggunakan data yang telah diolah. Menghitung dengan melakukan pengurangan data pusat cluster dengan setiap data sebaran cluster kemudian dijumlahkan seluruhnya dan diakar pangkat dua sesuai dengan rumus Euclidean Distance data yang digunakan adalah data quantity, Total dan Item pusat cluster dikurangi dengan quantity sebaran data lainnya. Setelah itu dilakukan perhitungan jarak dengan menggunakan Euclidean Distance dan mendapatkan hasil alokasi data ke pusat cluster terdekat yang disajikan pada Tabel 15.

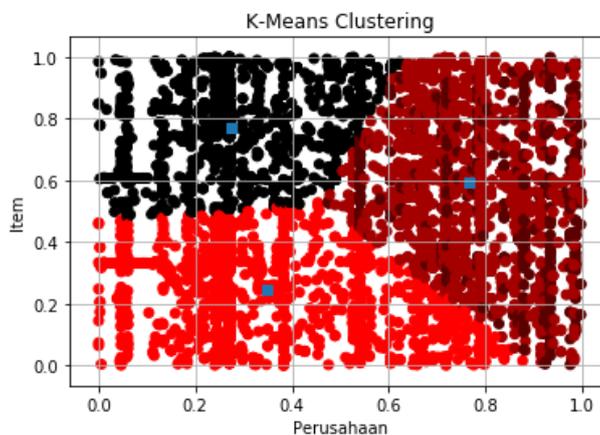
Tabel 15. Alokasi Data ke Cluster Berdasarkan Jarak Setiap Data dengan Pusat Cluster

	Quantity	Total	Nama Perusahaan	Item	Cluster
0	2	16218917.0	41	103	1
1	1	143550.0	41	2000	2
2	1	640712.0	41	103	3
3	1	550000.0	41	793	3
4	1	143550.0	42	2000	2
5	1	12298000.0	42	1329	1
6	1	3542000.0	42	1213	3
7	1	6732000.0	42	78	3
8	1	3850000.0	42	1732	3
9	1	1859000.0	42	170	3

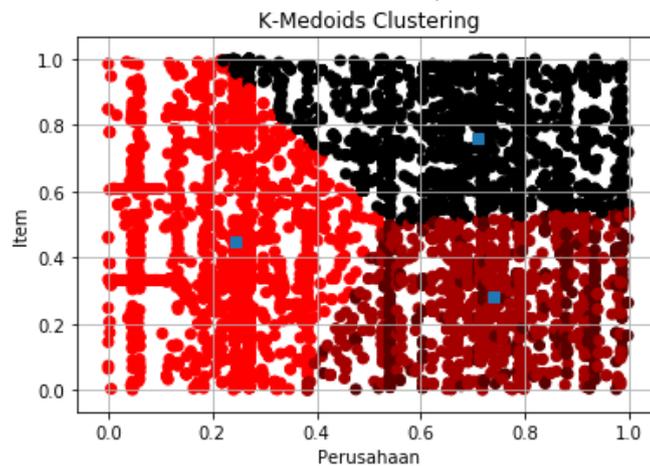
6. Menghitung Total simpangan (S)
Perhitungan dilakukan dengan melakukan pengurangan antara nilai total jarak yang baru dikurangi total jarak yang lama.
$$S = 17262784.44 - 12442986.36$$
$$= 4819798.082$$
Jika nilai S kurang dari 0, maka dilakukan penukaran objek dengan data *cluster* untuk membentuk sekumpulan jumlah *cluster* objek yang baru sebagai medoid.
 7. Mengulangi tahap 3 sampai dengan 5 apabila masih terjadi perubahan medoid.
Dalam penelitian ini menghasilkan nilai simpangan > 0 maka tidak dilakukan penukaran objek dan iterasi selesai.

3.3 Tampilan Cluster yang Terbentuk

Pada Gambar 1 a dan b menunjukkan hasil proses Clustering algoritma K Means dan clustering algoritma K Medoids.



Gambar 1.a Hasil Clustering K-Means



Gambar 1.b Hasil Clustering K-Medoids

Gambar 1 a dan b merupakan hasil pengclusteringan dari data penjualan di mana terbagi atas tiga cluster yaitu barang terlaris, laris dan kurang laris. Hasil dari clustering menggunakan algoritma K-Medoids menunjukkan anggota yang terbentuk pada Cluster 0 (Terlaris) ditunjukkan dengan node berwarna hitam berjumlah sebanyak 2940 data, Cluster 1 (Laris) ditunjukkan dengan node berwarna merah muda berjumlah 1655 data dan Cluster 2 (Kurang Laris) ditunjukkan dengan node berwarna merah tua berjumlah sebanyak 2563 data. Hasil

Silhouette Score dari clustering algoritma K-Medoids adalah 0.376. Sedangkan untuk algoritma K-Means menunjukkan anggota yang terbentuk pada Cluster 0 (Terlaris) ditunjukkan dengan node berwarna merah muda berjumlah sebanyak 2196 data, Cluster 1 (Laris) ditunjukkan dengan node berwarna merah tua berjumlah 1838 data dan Cluster 2 (Kurang Laris) ditunjukkan dengan node berwarna hitam berjumlah sebanyak 2563 data. Hasil Silhouette Score dari clustering algoritma K-Means adalah 0.404

Tabel 16. Hasil Akhir Perbandingan Algoritma

Algoritma	Sillhouette Score
K-Means Clustering	0.404
K-Medoids Clustering	0.376

3. KESIMPULAN DAN SARAN

Dalam penelitian ini dilakukan penambahan data terhadap data transaksi penjualan pada periode tahun 2018-2020 untuk tujuan clustering produk paling laris, laris dan kurang laris terjual serta melakukan perbandingan algoritma K Means dan algoritma K Medoids. Hasil dari penelitian ini menyimpulkan bahwa Algoritma K-Means memiliki kualitas yang lebih baik dalam membentuk cluster. Sedangkan dari hasil clustering yang terbentuk dapat dimanfaatkan untuk rekomendasi pemberian promosi yang lebih banyak kepada produk yang ada pada cluster kurang laris untuk dapat meningkatkan penjualan.

DAFTAR PUSTAKA

- [1] Y. Yulianti, D. Y. Utami, N. Hikmah, and F. N. Hasan, 2019, "PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS UNTUK MENGETAHUI MINAT CUSTOMER DI TOKO HIJAB," *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 241–246, Sep. 2019, doi: 10.33480/pilar.v15i2.650, [Online]. Available: <https://ejournal.nusamandiri.ac.id>.
- [2] A. Ayu, D. Sulistyawati, and M. Sadikin, 2021, "SISTEMASI: Jurnal Sistem Informasi Penerapan Algoritma K-Medoids untuk Menentukan Segmentasi Pelanggan." *Sistemasi: Jurnal Sistem Informasi*, Nomor 3, Volume 10, 516 – 526, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [3] R. Kahfi and M. Syahrizal, 2020, "Penerapan Data Mining Untuk Memprediksi Penjualan Bahan Bakar Minyak Menggunakan Algoritma K-Means Clustering (Studi Kasus: PT. Anugerah Alam Semesta)," *Journal of Information Sistem Research (JOSH)*, No. 3, Volume 1, Hal: 189 – 198, [Online]. Available: <https://ejournal.seminar-id.com>.
- [4] H. Sumarno, H. Priyono, and N. Hikmah, 2019, "Pengelompokkan Data Pembelian Tinta Dengan Menggunakan Metode K-Means," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, Nomor 2, Volume 3, 381-392, [Online]. Available: <http://tunasbangsa.ac.id/ejurnal/index.php/jsakti>.
- [5] I. Safitra Damanik, S. Retno Andani, D. Sehendro, M. P. Studi, S. Informasi, and S. Tunas Bangsa, 2019, "Teknik Data Mining Dalam Clustering Produksi Susu Segar Di Indonesia Dengan Algoritma K-Means," *BRAHMANA: Jurnal Penerapan Kecerdasan Buatan*, vol. 1, no. 1, pp. 31–39, [Online]. Available: <https://tunasbangsa.ac.id/pkm/index.php/brahmana>.
- [6] A. Ali, 2019, "Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 19, no. 1, pp. 186–195, Nov. 2019, doi: 10.30812/matrik.v19i1.529, [Online]. Available: <https://journal.universitاسbumigora.ac.id>.

- [7] W. Robiansyah and G. W. Nurcahyo, 2021, “Akurasi Pemberian Insentif Menggunakan Algoritma K-Medoids Terhadap Tingkat Kedisiplinan Pegawai,” *Jurnal Informasi dan Teknologi*, pp. 139–144, Sep, doi: 10.37034/jidt.v3i3.125.
- [8] P. Alam Jusia, F. Muhammad Irfan, and S. Dinamika Bangsa Jambi Jl Jend Sudirman Thehok Jambi, 2019, “CLUSTERING DATA UNTUK REKOMENDASI PENENTUAN JURUSAN PERGURUAN TINGGI MENGGUNAKAN METODE K-MEANS,” *urnal IKRA-ITH Informatika*, No 3, Vol 3, 75 – 84, [Online]. Available: <https://journals.upi-yai.ac.id/index.php/ikraith-informatika/article/view/596>.
- [9] R. Novianto and L. Goeirmanto, 2019, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering untuk Menganalisa Bisnis Perusahaan Asuransi,” *Jurnal Teknik Informatika dan Sistem (JATISI)*, *Hal June25 th*, vol. 6, no. 1, 2019, [Online]. Available: <http://jurnal.mdp.ac.id>
- [10] W. Wahyuni and H. Fahmi, 2020, “Penerapan Data Mining Clustering Pada Siswa-Siswi SMK Swasta Jaya Krama Beringin Dalam Menerima Potongan Biaya Administrasi Sekolah Dengan Menggunakan Algoritma K-Means,” *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, No 2, Vol.3, pp 1-7; [Online]. Available: <https://garuda.kemdikbud.go.id/documents/detail/1825264>.
- [11] T. Syahputra, 2019, “PENERAPAN DATA MINING DALAM PENGELOMPOKAN KELAS MAHASISWA BERDASARKAN HASIL UJIAN SARINGAN MASUK DENGAN ALGORITMA K-MEANS,” *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 5, no. 2, pp. 161–166, Jun. 2019, doi: 10.33330/jurteksi.v5i2.350, [Online]. Available: <https://jurnal.stmikroyal.ac.id/index.php/jurteksi>.
- [12] D. Sri Wardiani and N. Merlina, 2019, “IMPLEMENTASI DATA MINING UNTUK MENGETAHUI MANFAAT RPTRA MENGGUNAKAN METODE K-MEANS CLUSTERING,” *Jurnal PILAR Nusa Mandiri, Maret*, vol. 15, no. 1, p. 125, [Online]. Available: <http://bsi.ac.id/ubsj>
- [13] D. Retno Lestari, 2019, “Implementasi Data Mining Untuk Pengaturan Layout Swalayan Delimas Lestari Kencana Dengan Menggunakan Dengan Menggunakan Algoritma K-Means Clustering,” *Jurnal Riset Komputer (JURIKOM)*, No. 2, Vol. 6, 516 – 526, [Online]. Available: <http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom/112>
- [14] H. Artanto, F. Marisa, and D. Purnomo, 2019 “Seminar Nasional Hasil Riset Prefix-RTR IMPLEMENTASI DAN KOMPARASI ALGORITMA FUZZY C-MEANS DAN K-MEANS UNTUK MENGELOMPOKKAN SISWA BERDASARKAN NILAI AKADEMIK DAN PERILAKU SISWA (DATA SURVEY)”, *Conference on Innovation and Application of Science and Technology (CIASTECH)*, Malang, 08 Nov 2019.
- [15] N. Hikmah, D. Ariyanti, and M. Sugesti, 2019, “Penerapan Teknik Data Mining untuk Clustering Armada pada PT. Siaga Transport Indonesia Menggunakan Metode k-Means.”, *Jurnal Explore STMIK Mataram* , No 1, Volume 9, 7 – 12, [Online]. Available: <https://garuda.kemdikbud.go.id/author/view/1005445>.
- [16] W. Purba and W. Siawin, 2019, “IMPLEMENTASI DATA MINING UNTUK PENGELOMPOKKAN DAN PREDIKSI KARYAWAN YANG BERPOTENSI PHK DENGAN ALGORITMA K-MEANS CLUSTERING,” *Jurnal Sistem Informasi Ilmu Komputer Prima (JUSIKOM PRIMA)*, vol. 2, no. 2, 2019, [Online]. Available: <http://jurnal.unprimdn.ac.id/index.php/JUSIKOM>.
- [17] E. Dewi *et al.*, 2019, “Clustering Wilayah Dan Pelanggaran Berkendaraan Menggunakan Algoritma K-Means Pada Data Satlantas Polres Tasikmalaya Kota,” *JURNAL SISTEM INFORMASI DAN TEKNOLOGI INFORMASI*, No. 1, Vol. 8, 1-11, [Online]. Available: <https://ejurnal.dipanegara.ac.id/index.php/jusiti>.