

SEGMENTASI PELANGGAN PADA PT. TAB MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING

Eko Santoso ¹, Agus Umar Hamdani ².

¹ Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur,

¹ Jl. Ciledug Raya, Petukangan Utara, Jakarta Selatan, 12260. DKI Jakarta, Indonesia.

¹ 1912520010@student.budiluhur.ac.id, ² agus.umarhamdani@budiluhur.ac.id

Abstrak : Pelanggan adalah aset yang sangat berharga bagi perusahaan, mempertahankan dan memberikan pelayanan terbaik kepada pelanggan setia pada perusahaan dagang sangat diperlukan. Dengan mengetahui profil kelompok pelanggan berdasarkan pembeliannya diharapkan dapat membantu perusahaan dalam mengidentifikasi kelompok pelanggan yang sangat bernilai dan setia. Klasterisasi pelanggan merupakan sebuah cara yang dapat mengenali kelompok-kelompok pelanggan yang memiliki karakteristik yang sama sehingga hasilnya dapat dimanfaatkan perusahaan untuk mengembangkan program loyalitas dan pelayanan khusus untuk masing-masing kelompok pelanggan tersebut. Pada penelitian ini klasterisasi pelanggan yang akan dibentuk dengan menggunakan algoritma K-Means, dengan data yang diolah berasal dari hasil laporan transaksi penjualan periode 2019 sampai dengan 2022. Setelah melalui proses preprocessing, dataset yang telah disiapkan akan dinormalisasikan agar memiliki rentang data yang tidak terlalu jauh. Tujuan dari penelitian ini adalah untuk membentuk kelompok pelanggan, dengan metode K-Means dihasilkan tiga buah klaster dengan populasi 34,96% kelas Bronze, 34,96% kelas Silver dan kelas Gold 28,01%.

Kata Kunci : Data Mining, Klasterisasi Pelanggan, K-Means, Pengelompokan Pelanggan, Karakteristik Pelanggan, Segmentasi Pelanggan.

Abstract: Customers are valuable assets for any company, and maintaining and providing excellent service to loyal customers is crucial for a trading company. By understanding customer profiles based on their purchasing behavior, the company can identify high-value and loyal customer segments. Customer clustering is a method that helps recognize groups of customers with similar characteristics, allowing the company to develop loyalty programs and tailored services for each customer segment. In this study, customer segmentation will be performed using the K-Means algorithm, with data derived from sales transaction reports from the period of 2019 to 2022. After preprocessing, the prepared dataset will be normalized to ensure that the data range is not too wide. The objective of this research is to form customer clusters using the K-Means method, resulting in three clusters with a population distribution of 34.96% in the Bronze class, 34.96% in the Silver class, and 28.01% in the Gold class.

Keywords: Data Mining, Customer Clustering, K-Means, Customer Grouping, Customer Characteristics, Customer Segmentation.

PENDAHULUAN

Pelanggan merupakan salah satu aset yang berharga bagi perusahaan. Kebutuhan informasi yang lebih mendalam

diperlukan untuk mengetahui profil dan karakteristik pelanggan pada perusahaan trading. Proses identifikasi seberapa tingkat profit pelanggan kepada perusahaan dilakukan dengan

menggunakan seluruh data yang dimiliki oleh perusahaan. Segmentasi atau pengelompokan pelanggan dilakukan dengan membagi seluruh pelanggan menjadi beberapa kelompok sesuai dengan kemiripan perilaku pembelian diantara seluruh pelanggan yang dimiliki, disaat jumlah pelanggan dan jumlah transaksi semakin besar maka proses segmentasi ini menjadi satu hal yang menyulitkan ketika dilakukan secara manual. Penggunaan data mining akan lebih mudah mengenali pola-pola tertentu pada seluruh data yang diolah [1], dengan begitu setelah hasil segmentasi pelanggan terbentuk menjadi beberapa kelompok maka setiap kelompok akan lebih mudah dikenali perilaku pembeliannya dan akan membantu manajemen untuk dapat mengambil keputusan bisnis yang lebih cocok dengan potensi pembelian produk pada masing-masing kelompok pelanggan. Keberhasilan identifikasi pelanggan ini diharapkan bisa mempertahankan pelanggan dan mencegah pelanggan berpindah pembelian kepada perusahaan lain (pesaing). PT Tab adalah perusahaan yang bergerak dalam bidang penjualan barang-barang untuk peralatan restoran dan hotel. Mayoritas pelanggan perusahaan adalah institusi hotel dan restoran yang tersebar di seluruh Indonesia, selain itu ada juga pelanggan ritel yang membeli langsung lewat showroom perusahaan. Perusahaan ini telah memiliki aplikasi inventori yang telah digunakan sejak tahun 2005, sehingga data-data penjualan pelanggan telah tersedia sebagai bahan penelitian. Namun sayangnya data transaksi pembelian pelanggan belum dimanfaatkan dengan maksimal oleh perusahaan. Contoh pemanfaatan data tersebut dapat diolah untuk mengetahui mana pelanggan-pelanggan yang lebih potensial yang memungkinkan perusahaan untuk

memberikan pelayanan yang lebih baik kepada pelanggan-pelanggan potensial tersebut. Dengan berkembangnya teknologi informasi saat ini, banyak teknik-teknik penambangan data yang terlahir untuk menyelesaikan masalah segmentasi pelanggan. Salah satu teknik pengolahan data yang sering dipakai yaitu teknik data mining dengan metode K-means *Clustering*. Algoritma K-means *Clustering* ini diharapkan dapat membantu Perusahaan untuk mengidentifikasi kelompok pelanggan potensialnya. Berdasarkan latar belakang yang telah disebutkan diatas maka dapat dirumuskan masalah yang dibahas pada penelitian ini, yaitu bagaimana melakukan klasterisasi pelanggan di PT. Tab dan menemukan pelanggan potensial di PT. Tab. Algoritma K-means merupakan salah satu algoritma data mining yang sangat populer untuk mengelompokkan data dalam beberapa klaster berdasarkan kemiripan datanya. Algoritma K-means dipilih karena bekerja dengan baik pada data yang besar. Algoritma ini dapat mengelompokkan data dengan cepat dan cukup mudah untuk diimplementasikan.

KAJIAN PUSTAKA

Data Mining

Penambangan data (*data mining*) adalah pencarian pengetahuan dalam basis data pada proses identifikasi pola-pola yang valid, yang berpotensi bermanfaat dan dapat dipahami dengan mudah. *Data Mining* adalah proses penggalian informasi dan pola yang bermanfaat dari suatu data yang sangat besar. Proses data mining terdiri dari pengumpulan data, ekstraksi data, analisa data, dan statistik data. Ia juga umum dikenal sebagai knowledge discovery, knowledge extraction, data/pattern analysis, information harvesting, dan lainnya [2].

Data mining berdasarkan fungsionalitasnya dapat dikelompokkan menjadi enam bagian yaitu [3] :

a. Klasifikasi (*classification*)

Diterapkan pada data baru untuk mengelompokkan jenis objek. Klasifikasi termasuk pada model *supervised*. Pada persoalan klasifikasi kita memiliki sampel data dan memprediksi beberapa class yang ada berdasarkan sampel yang ada. Hanya satu atribut di antara banyak atribut yang disebut dengan atribut predator. Klasifikasi ini juga umum, digunakan untuk pemodelan bisnis dan lainnya. Misal klasifikasi menentukan penyakit tertentu atau menentukan customer berdasarkan model pembayarannya.

b. Klastering (*Clustering*)

Berbeda dengan klasifikasi, klastering termasuk model *unsupervised*. Klastering mengelompokkan data yang tidak diketahui labelnya. Klastering yang diorganisasi ke dalam struktur hierarkikal akan mendefinisikan taksonomi dari data. Penerapan metode klastering yang tepat akan menghasilkan *clustering* yang berkualitas. Suatu cluster dikarakterisasi oleh *centroid*, atribut *histogram* dan *clustering* model *hierarchial tree*.

c. Regresi (*Regression*)

Merupakan suatu fungsi yang digunakan untuk memodelkan data untuk meminimalkan hasil kesalahan prediksi. Umumnya regresi dilakukan dengan data yang bersifat *time series*.

d. Asosiasi (*Association*)

Merupakan pemodelan kebergantungan. Fungsi asosiasi ini biasanya kita kenal dengan istilah "market basket analysis" yang merupakan fungsi untuk menemukan relasi atau korelasi antara himpunan item-item. Aturan asosiasi diartikan pada basket data yang digunakan untuk keperluan promosi, desain katalog untuk meningkatkan penjualan. Contoh penerapan asosiasi adalah ketika customer membeli pampers maka ada kemungkinan membeli bir.

e. Deteksi Anomali (*Anomaly Detection*)

Mengidentifikasi data yang tidak umum. Bisa berupa outlier, perubahan deviasi/bias yang penting dan perlu investigasi lebih lanjut.

f. Peringkasan (*Summarization*)

Menyediakan representasi data yang lebih sederhana meliputi pelaporan, visualisasi data yang dipergunakan untuk menunjang informasi dan penguatan keputusan.

Pengertian *Clustering*

Clustering adalah suatu metode yang berupaya mengelompokkan satu objek sedemikian rupa sehingga objek dalam kelompok yang sama (*cluster*). Ada dua metode dalam analisis klaster, metode hirarki dan non hirarki, metode hirarki atau *Hierarchical Clustering* adalah suatu metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan paling dekat. Kemudian proses dilanjutkan ke objek lain yang memiliki kedekatan pada keduanya, demikian seterusnya sehingga klaster akan membentuk semacam pohon dimana ada tingkatan yang jelas antara objek dari yang paling mirip hingga yang tidak memiliki kemiripan, bentuk visualisasi yang seringkali digunakan untuk menggambarkan data klaster hirarki ini adalah dengan visualisasi dendogram [4]. Pada metode hirarki terdiri dari beberapa algoritma, seperti single linkage, complete linkage, average linkage, centroid, ward dan median cluster. Sedangkan metode non hirarki digunakan untuk mengelompokkan objek berdasarkan jumlah cluster yang sebelumnya telah ditentukan sebelumnya [5], setelah proses penentuan jumlah klaster ditentukan kemudian proses klasterisasi dijalankan tanpa mengikuti proses hirarki. Dalam metode non hirarki metode yang paling umum digunakan adalah metode K-Means, K-Medoids dan lainnya.

Kunci dari analisis *cluster* adalah menentukan tingkat kemiripan (*similarity*) antara data, semakin tinggi tingkat kemiripan pada cluster yang sama maka cluster yang terbentuk akan semakin baik, begitupun dengan tingkat kemiripan antara cluster, semakin tingkat kemiripan antara cluster rendah maka cluster yang terbentuk akan semakin baik. Dengan menggunakan Distance Measurement jarak antara data

dalam satu cluster dapat diukur tingkat kemiripannya [6], berikut ini adalah beberapa cara perhitungan distance measure yang umum digunakan :

Euclidean Distance (Pythagoras)

$$D_{Ls}(X_2, X_1) = ||X_2 - X_1||_2 = \sqrt{\sum_{j=1}^p (X_{2j} - X_{1j})^2}$$

Keterangan :

p = Dimensi Data

X_1 = Posisi Titik Data 1

X_2 = Posisi Titik Data 2

Kmeans

Algoritma K-Means salah satu metode dan clustering non hierarki yang berusaha mengelompokkan data ke dalam suatu cluster sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama [7]. Metode ini membagi data menjadi beberapa kelompok sehingga dalam setiap kelompok yang terbentuk akan memiliki karakteristik yang sama untuk setiap anggota kelompok tersebut, sebaliknya untuk setiap anggota dari kelompok yang berbeda akan memiliki karakteristik yang berbeda. Sehingga tujuan utama dalam proses klasterisasi ini adalah meminimalkan variasi didalam suatu klaster / kelompok (homogen) dan meningkatkan variasi antar klaster / kelompok (*heterogen*).

K-Means yaitu algoritma yang dimulai dengan menetapkan nilai pusat terlebih dahulu untuk menjadi pusat sementara dari centroid atau cluster, kemudian dengan menggunakan rumus menghitung jarak setiap data ke pusat sehingga data yang lebih dekat ke pusat menjadi satu kelompok dan data yang jauh menjadi kelompok lainnya. Berikut merupakan prosedur Algoritma pengelompokan *K-means* menurut dengan beberapa langkah berikut [8]:

1. Pengambilan data
2. Tentukan data yang akan diproses
3. Menentukan jumlah klaster

4. Tentukan titik awal klaster (centroid) untuk setiap klaster
5. Hitung jarak centroid klaster dengan seluruh data
6. Kelompokkan seluruh data ke klaster terdekat
7. Hitung titik centroid yang baru berdasarkan nilai rata-rata posisi data setiap klaster
8. Bandingkan klaster baru dengan klaster awal, bila masih berubah maka proses terus berlanjut hingga klaster baru sama dengan klaster sebelumnya.

2.4 Dataset

Dataset pembelian pelanggan yang telah melalui tahap preprocessing terdiri dari 4 atribut yang dapat dilihat pada tabel 1. Sedangkan isi dari dataset pembelian pelanggan dapat dilihat pada tabel 2.

Table 1. Struktur dataset pembelian pelanggan

| Nama Atribut | Tipe | Keterangan |
|---------------------|---------|-------------------------------------|
| nama_customer | varchar | Nama pelanggan |
| jumlah_pembelian | double | Jumlah pembelian pelanggan |
| frekuensi_pembelian | int | Frekuensi pembelian pelanggan |
| avg_pembelian | double | Nilai rata-rata pembelian pelanggan |

Decision Tree

Decision tree adalah algoritma *machine learning* yang menggunakan seperangkat aturan untuk membuat keputusan dengan struktur seperti pohon yang memodelkan kemungkinan hasil, biaya sumber daya, utilitas dan kemungkinan konsekuensi atau resiko. Konsepnya adalah dengan cara menyajikan algoritma dengan pernyataan bersyarat, yang meliputi cabang untuk mewakili langkah-langkah pengambilan keputusan yang dapat mengarah pada hasil yang menguntungkan. Decision Tree digunakan juga sebagai teknik pembelajaran dalam kecerdasan buatan dengan istilah *Decision Tree Learning* (DTL). DTL adalah “teknik pembelajaran

mesin yang membangun representasi aturan klasifikasi berstruktur sekuensial hirarki dengan cara mempartisi himpunan data latih secara rekursif” [9].

Penelitian Terkait

Terdapat penelitian-penelitian yang sebelumnya telah meneliti penggunaan data mining dengan algoritma K-Means *Clustering* dalam pengelompokan data yang dirangkum oleh penulis.

Penelitian pertama [10] menggunakan algoritma K-Means untuk melakukan pengelompokan data sepeda motor menjadi 3 klaster, yaitu murah, standar dan mahal. Hasil analisis performansi k-means dari 15 pengujian dari setiap uji coba yang dilakukan, diperoleh nilai rata-rata *Precision* sebesar 76%, nilai *Recall* sebesar 76% dan *Accuracy* sebesar 81%.

Penelitian kedua dilakukan oleh [11] melakukan penelitian tentang segmentasi pelanggan dengan menggunakan algoritma K-Means pada restoran XYZ. Data pelanggan dikelompokkan menjadi dua cluster (potensial dan tidak potensial). Kemudian setiap cluster diklasifikasikan berdasarkan kriteria mana yang lebih diprioritaskan yaitu salah satu dari kriteria jumlah berapa kali transaksi dan jumlah nominal transaksi. Cluster dengan nilai terbesar pada centroid akhir merupakan cluster yang direkomendasikan menerima reward, sedangkan cluster dengan nilai terkecil merupakan cluster yang tidak berhak menerima reward.

Penelitian ketiga [12] melakukan penelitian tentang pengelompokan data obat-obatan di RSUD Pekan Baru berdasarkan pemakaian obat yang rendah, sedang, dan tinggi. Dari hasil kesimpulan didapatkan kelompok obat yang termasuk pemakaian sedikit rata-rata permintaan obat setiap tahunnya kurang dari 18000 buah, dan obat yang termasuk pemakaian sedang rata-rata permintaan obat setiap tahunnya diantara 18000–70000 buah, sedangkan obat yang masuk kedalam kelompok obat yang pemakaian tinggi rata-rata permintaan obat setiap tahunnya diatas 70000 buah.

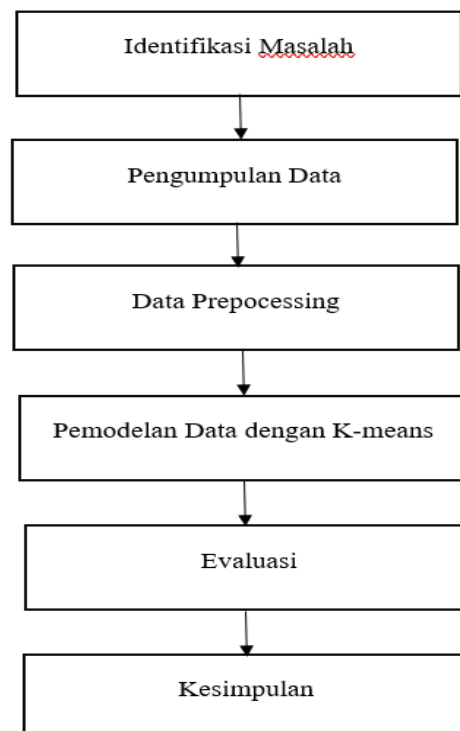
Penelitian keempat [13] melakukan penelitian tentang pengelompokan bahan

bangunan pada UD. Toko Bangunan YD Indarung menjadi beberapa kategori yaitu sangat laris, laris dan kurang laris. Hasil kesimpulan mendapatkan 3 kategori penjualan bahan bangunan, yaitu kurang laris, sangat laris dan laris. Hal ini dibuktikan berdasarkan perhitungan jarak terdekat berdasarkan penentuan nilai centroid secara random dengan menggunakan rumus Euclidean Distance, pada jumlah stok barang diambil dari jumlah barang terjual dan jumlah barang tidak terjual.

Penelitian kelima [14] melakukan penelitian tentang pengelompokan pelanggan yang berpotensi untuk menjadi pelanggan tetap atau pelanggan potensial pembelian barang dengan menggunakan atribut yaitu pembelian barang, lama kerjasama dan barang return. Hasil penelitian mendapatkan 3 kategori pelanggan yaitu Tidak Potensial, Potensial, Sangat Potensial.

METODOLOGI PENELITIAN

Adapun tahapan penelitian untuk menerapkan metode K-means pada penelitian Klasterisasi Pelanggan di PT. TAB dapat dilihat pada Gambar 1.



Gambar 1. Tahapan penelitian

Berikut penjelasan tiap tahapan untuk penerapan metode K-means pada penelitian ini.

Identifikasi Masalah

Masalah pada PT TAB adalah belum adanya segmentasi pelanggan yang mengelompokkan pelanggan atas dasar kemiripan transaksinya. Sehingga sulit bagi perusahaan untuk menentukan mana saja pelanggan yang masuk kategori sangat potensial untuk program-program loyalitas dan pelayanan khusus.

Pengumpulan Data

Sumber data merupakan data transaksional yang tercatat secara harian. Adapun sumber data yang digunakan untuk pembuatan data mining berasal dari laporan penjualan pelanggan dari aplikasi yang berjalan di PT TAB. Data tersebut terkompulasi dari tahun 2019 sampai dengan tahun 2022.

Data Preprocessing

Pra pemrosesan data dimulai dari pembersihan data dari data-data yang tidak valid. Kemudian menghilangkan atribut – atribut yang tidak digunakan. Setelah itu dilakukan agregasi data pada atribut jumlah pembelian, frekuensi pembelian dan nilai rata-rata pembelian. Kemudian dilakukan transformasi nilai yang bertujuan mengubah data mentah ke dalam bentuk yang lebih terukur pada atribut jumlah pembelian, rata - rata pembelian dan frekuensi pembelian sehingga menghasilkan dataset yang siap untuk diproses selanjutnya.

Pemodelan Data

Dataset yang sudah dilakukan preprocessing sebelumnya akan diolah dengan metode kmeans dengan perhitungan jarak yang digunakan adalah *Euclidean Distance*. Metode kmeans sangat bergantung dari jumlah klaster (nilai K) yang akan dihasilkan, nilai K tersebut dapat dihasilkan dari arahan pihak perusahaan atau dapat juga dihitung berdasarkan hasil evaluasi menggunakan metode Elbow.

Evaluasi

Pada tahap ini hasil klasterisasi atau segmentasi yang telah dihasilkan pada tahap modeling berikutnya akan dievaluasi jumlah segmentasi yang paling optimal. Dengan menggunakan metode Elbow Curve dilakukan evaluasi untuk mencari nilai K terbaik. Tujuan dari metode ini adalah mencari titik di kurva di mana penambahan jumlah klaster tidak memberikan penurunan signifikan dalam *within-cluster sum of squares (WCSS)*. Biasanya, titik ini memiliki bentuk seperti siku, yang memberi nama metode ini. Jumlah klaster yang sesuai dengan titik siku tersebut dianggap sebagai jumlah klaster optimal.

Kesimpulan

Pada kesimpulan dan saran berisikan dari hasil segmentasi pelanggan beserta saran dari apa yang telah didapatkan dari hasil segmentasi pelanggan untuk penelitian selanjutnya yang akan dilakukan di masa yang akan datang.

HASIL DAN PEMBAHASAN

Sumber data penelitian yang digunakan untuk proses data mining berasal dari laporan penjualan pelanggan dari aplikasi yang berjalan di PT Tab. Data tersebut terkompulasi dari tahun 2019 sampai dengan tahun 2022 dalam bentuk file excel yang kemudian diimport ke database mysql menjadi satu tabel yang diberi nama ds_penjualan . Tabel 2 merupakan contoh baris data beserta dengan nama kolomnya untuk data transaksi penjualan.

Tabel 2. Data Transaksi Penjualan PT.TAB

| Tgl invoice | customer | Total |
|-------------|---|-----------|
| 03/01/2019 | Sumber Cahaya Mulia, PT (supplier) | 9147600 |
| 03/01/2019 | Mulia Senayan Hotel (Pt. Mulia Intanlestari) | 88550 |
| 03/01/2019 | Harris Yello Manggarai (SUKSES SEJAHTERA PT) | 2028950 |
| 03/01/2019 | Shangri-La Hotel Jakarta (Pt. Swadharma Kerry Satya) | 17077504 |
| 03/01/2019 | BW Luxury Jambi (Jambi Permai Abadi, PT) | 1451587,5 |
| 03/01/2019 | Mutia Culinary Concepts , PT | 3987500 |
| 03/01/2019 | Mutia Culinary Concepts , PT | 40221500 |
| 03/01/2019 | Prime Park Hotel Bandung (PP Properti Tbk, PT) | 602600 |
| 03/01/2019 | Potato Head (Potato Bar and Bistro Jkt,PT) | 212960 |
| 03/01/2019 | Prime Park Hotel Bandung (PP Properti Tbk, PT) | 4807000 |
| 03/01/2019 | HOUSE OF CULINAIRE | 13684770 |
| 03/01/2019 | Excelso Multi Rasa, PT | 13160400 |
| 03/01/2019 | Fave hotel Manahan Solo (PT. Delta Merlin Dunia Property) | 4807000 |

Pada proses selanjutnya dataset penjualan dibersihkan dari data-data penjualan yang tidak valid yaitu data penjualan yang memiliki isi field total_pembelian berisi 0. Kemudian dataset penjualan yang masih mempunyai beberapa kolom yang tidak diperlukan akan dihapus untuk mereduksi jumlah data menjadi beberapa dimensi saja. Data yang sudah direduksi jumlah kolomnya, kemudian akan dilakukan proses agregasi menggunakan perintah SQL `sum(jumlah)` pada kolom total sebagai atribut jumlah_pembelian, `count` pada atribut no_invoice sebagai atribut frekuensi_pembelian, dan pembagian nilai atribut total_pembelian dibagi frekuensi

sebagai atribut avg_pembelian dimana tiga atribut ini akan digunakan untuk proses clustering pada proses selanjutnya.

Tabel 3. Potongan Data Hasil Agregasi

| Customer | JP | FP | AP |
|---|--------------|----|----------|
| 1% Bistro And Cafe (Anissa Belfa Persero, PT) | 1027 2506 | 1 | 10272506 |
| 1001 Hotel (Perkasa Tunggal Mandiri, Pt) | 1092 1770 | 4 | 2730443 |
| 8 Degree Projects (Harta Tersembunyi,Pt) | 4128 250 | 1 | 4128250 |
| 88 Hotel Fatmawati (Fatmawati Sukses Cemerlang, Pt) | 2106 7200 | 2 | 10533600 |
| Abuba, PT | 6500 000 | 3 | 2166667 |

Keterangan :

JP = Jumlah pembelian

FP = Frekuensi pembelian

AP = Jumlah rata-rata pembelian

Sebelum dilakukan normalisasi data dengan metode *minmax* terlebih dahulu dilakukan transformasi nilai yang bertujuan mengubah data mentah ke dalam bentuk yang lebih terukur pada atribut jumlah pembelian, rata - rata pembelian dan frekuensi pembelian. Kemudian tentukan kriteria atribut dengan nilai 5,4,3,2,1 yang merepresentasikan kontribusi pelanggan terhadap keuntungan perusahaan. Nilai 5 menunjukkan sangat tinggi dan 1 menunjukkan sangat rendah.

Tabel 4. Skala Nilai Atribut

| Skala | N | FP | JP | AP |
|-----------|---|------------------------------|---|--|
| Very High | 5 | > 100 | > 1000.000.00 0 | > 50.000.00 0 |
| High | 4 | >= 48 dan <= 100 | >= 500.000.001 dan <= 1000.000.00 0 | >= 5000.001 dan <= 50.000.00 0 |
| Medium | 3 | >= 15 | >= 100.000.001 | >= 1000.001 |

| | | | | |
|----------|---|----------------|----------------------------------|----------------------------|
| | | dan <= 47 | dan <= 500.000.000 | dan <= 5000.000 |
| Low | 2 | >= 2 dan <= 14 | >= 20.000.001 dan <= 100.000.000 | >= 500.001 dan <= 1000.000 |
| Very Low | 1 | <= 1 | <= 20.000.000 | <= 500.000 |

Tabel 5. Contoh Hasil Transformasi Nilai Atribut

| Customer | JP | FP | AP |
|---|----|----|----|
| 1% Bistro And Cafe (Anissa Belfa Persero, PT) | 1 | 1 | 4 |
| 1001 Hotel (Perkasa Tunggal Mandiri, Pt) | 1 | 2 | 3 |
| 8 Degree Projects (Harta Tersembunyi,Pt) | 1 | 1 | 3 |

Pada tahap selanjutnya akan dilakukan percobaan pengelompokan dengan algoritma kmeans dengan menggunakan nilai k (jumlah kluster) 3, 4 dan 5. Untuk menentukan kualitas sebuah kluster atau segmentasi maka dilakukan pengukuran dengan menggunakan metode *Elbow Curve*. Metode *Elbow Curve* adalah metode yang digunakan dalam analisis klusterisasi untuk membantu menentukan jumlah kluster yang optimal dalam suatu dataset. Berikut ini adalah tahapan pengujiannya :

Tentukan nilai K atau jumlah kluster yang akan diuji

Pada percobaan ini nilai K atau jumlah kluster yang digunakan dari 2 sampai 5.

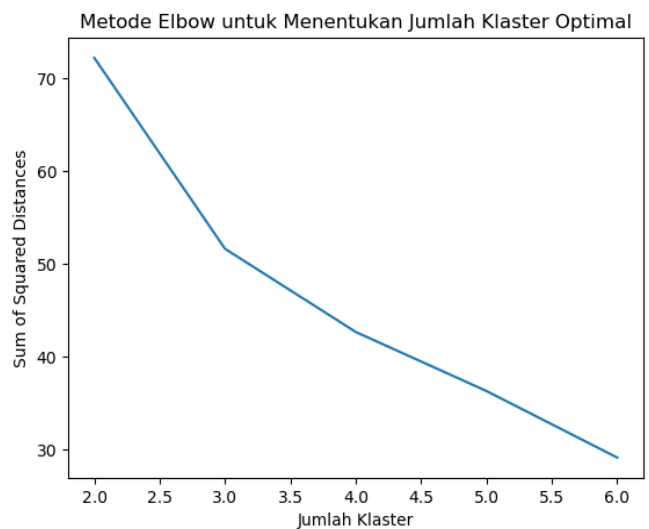
Hitung selisih dari nilai *Sum of Square Error (SSE)* dari setiap kluster atau nilai K

Nilai K atau jumlah kluster yang paling baik berdasarkan tabel 5 adalah nilai K yang memiliki selisih SSE terbesar dari nilai K sebelumnya, terlihat pada tabel 5 bahwa nilai K = 3 atau jumlah kluster 3 merupakan

nilai K yang memiliki nilai selisih SSE terbesar dengan nilai SSD pada K sebelumnya.

Tabel 6. Nilai SSE tiap kluster

| K (Kluster) | Nilai SSE | Selisih SSE Sebelumnya |
|-------------|-------------|------------------------|
| 2 | 72,17107134 | - |
| 3 | 51,61907863 | 20,55199271 |
| 4 | 42,66634444 | 8,952734191 |
| 5 | 36,30931299 | 6,357031454 |



Gambar 2. Visualisasi SSE yang menyerupai siku berada diposisi K3

Tabel 6 merupakan tabel yang berisi potongan data hasil percobaan dengan 3 kluster :

Tabel 1. Potongan data hasil percobaan dengan 3 kluster

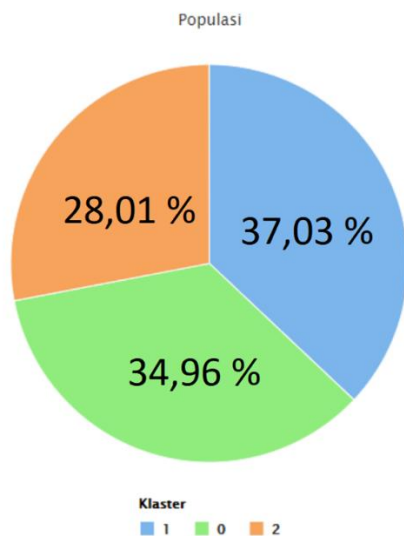
| customer | k |
|---|---|
| 1% Bistro And Cafe (Anissa Belfa Persero, PT) | 0 |
| 1001 Hotel (Perkasa Tunggal Mandiri, Pt) | 0 |
| 8 Degree Projects (Harta Tersembunyi,Pt) | 0 |
| 88 Hotel Fatmawati (Fatmawati Sukses Cemerlang, Pt) | 1 |
| 88 Hotel Grogol (Cv. Citra Indonesia) | 0 |
| 88 Hotel Mangga Besar 120 (Pt. Citrabangun Nusapermai) | 1 |
| 88 Mangga Besar 10d - (Waringin Buana Citra Realty, Pt) | 1 |
| @Hom Hotel Semarang (Pt. Welstand Park Realty) | 0 |

| | |
|---|---|
| @Hom Premiere Tambun (Pt. Metropolitan Permata Development) | 2 |
| Abuba, PT | 0 |
| Accor Vacation Club (APVC Indonesia.Pt) | 0 |
| Adele Dining (Charlies Lestari Sentosa, Pt) | 1 |
| Aden Indonesia (Pt. Andalan Duta Eka Nusantara) | 1 |
| Agricon Putra Citra Optima, Pt | 1 |
| Agung Sedayu, Pt | 1 |

Tabel 2. Hasil populasi 3 klaster

| Klaster | Jumlah Data | Populasi | Total Pendapatan |
|---------|-------------|----------|------------------|
| 0 | 372 | 34,96 % | 10,97 % |
| 1 | 394 | 37,03 % | 1,72 % |
| 2 | 298 | 28,01 % | 87,31 % |

Gambar 3 merupakan visualisasi hasil populasi 3 klaster :



Gambar 3. Visualisasi populasi 3 klaster

Dari hasil percobaan menggunakan jumlah klaster 3 menghasilkan 3 kelompok pelanggan. Untuk memudahkan identifikasi masing - masing diberikan nama sesuai dengan tingkat profitabilitasnya. Klaster 1(K0) terdiri dari 372 pelanggan dengan populasi 34,96 % menyumbang 10,97 % dari total penjualan Perusahaan, klaster ini diberi nama *silver*. Klaster 2(K1) terdiri dari 394 pelanggan dengan populasi 37,03 % menyumbang 1,72 % dari total penjualan perusahaan,

klaster ini diberi nama *bronze*. Klaster 3(K2) terdiri dari 298 pelanggan dengan populasi 28,01 % menyumbang 87,31 % dari total penjualan Perusahaan, klaster ini diberi nama *gold*. Dengan dengan rata-rata frekuensi pembelian 28 kali klaster tipe gold ini merupakan pelanggan yang sangat loyal bagi perusahaan. Untuk dapat mengetahui karakteristik dari masing-masing klaster yang telah terbentuk diperlukan analisa yang dapat menginterpretasikan hasil klasterisasi tersebut. Untuk memudahkan pemahaman apa yang membedakan antara klaster *bronze*, *silver* dan *gold* maka dapat digunakan klasifikasi dengan metode decision tree dengan menjadikan klaster sebagai label dari dataset pembelian pelanggan seperti disajikan pada tabel 8.

Tabel 3. Potongan dataset dengan label hasil klasterisasi

| nama_customer | klaster |
|---|---------|
| 1% Bistro And Cafe (Anissa Belfa Persero, PT) | silver |
| 1001 Hotel (Perkasa Tunggal Mandiri, Pt) | silver |
| 8 Degree Projects (Harta Tersembunyi,Pt) | silver |
| 88 Hotel Fatmawati (Fatmawati Sukses Cemerlang, Pt) | bronze |
| 88 Hotel Grogol (Cv. Citra Indonesia) | silver |
| 88 Hotel Mangga Besar 120 (Pt. Citrabangun Nusapermai) | bronze |
| 88 Mangga Besar 10d - (Waringin Buana Citra Realty, Pt) | bronze |
| @Hom Hotel Semarang (Pt. Welstand Park Realty) | silver |
| @Hom Premiere Tambun (Pt. Metropolitan Permata Development) | gold |
| Abuba, PT | silver |
| Accor Vacation Club (APVC Indonesia.Pt) | silver |
| Adele Dining (Charlies Lestari Sentosa, Pt) | bronze |
| Aden Indonesia (Pt. Andalan Duta Eka Nusantara) | bronze |
| Agricon Putra Citra Optima, Pt | bronze |
| Agung Sedayu, Pt | bronze |

Dataset ini akan digunakan untuk diolah menggunakan metode klasifikasi decision tree, hasilnya disajikan pada gambar 4.



Gambar 4. Visualisasi hasil klasifikasi untuk ketiga kluster menggunakan decision tree

Dari hasil pohon keputusan yang terbentuk dapat diekstraksi beberapa aturan yang menjadi ciri-ciri dari masing-masing kluster sebagai berikut :

1. Kluster bronze (K0) :
 - a. Jumlah pembelian > 19.994.739,5 dan ≤ 99.981.634,5.
 - b. Jumlah pembelian ≤ 19.994.739,500 dengan nilai rata-rata pembelian > 5.058.775 dan frekuensi pembelian > 1.5 kali.
2. Kluster silver (K1) :
 - a. Jumlah pembelian ≤ 19.994.739,500 dan nilai rata-rata pembelian ≤ 5.058.775.
 - b. Jumlah pembelian ≤ 19.994.739,500 dan nilai rata-rata pembelian > 5.058.775 pelanggan dan frekuensi pembelian ≤ 1.5 kali.
3. Kluster gold (K2) :
 - a. Jumlah pembelian > 99.981.634,5.

KESIMPULAN

Penerapan metode K-Means *Clustering* digunakan untuk mengelompokkan

pelanggan di PT TAB berdasarkan kriteria jumlah pembelian, frekuensi pembelian dan nilai rata – rata pembelian. Berdasarkan hasil pengujian yang dilakukan dengan menggunakan metode elbow didapatkan jumlah kluster optimal untuk dataset ini adalah 3 kluster. Hasil klusterisasi dengan 3 (tiga) kluster menghasilkan pengelompokan pelanggan tipe bronze (K1) sebanyak 394 pelanggan, tipe silver (K0) sebanyak 372 pelanggan dan tipe gold (K2) sebanyak 298 pelanggan. Hasil penelitian juga mendapatkan bahwa kluster gold adalah tipe pelanggan paling potensial bagi perusahaan karena berkontribusi sebesar 87,31% dari total penjualan perusahaan dengan karakteristik jumlah pembelian lebih dari 99.981.634,5 dan nilai pembelian rata-rata yang tinggi mencapai 14.604.145. Kluster silver terdiri menyumbang 10,97% penjualan perusahaan dengan rata-rata frekuensi pembelian 6 dan nilai pembelian rata-rata yang mencapai 6.958.457. Kluster bronze menyumbang 1,72% pendapatan perusahaan dengan rata-rata frekuensi pembelian 2 kali atau sangat rendah dan nilai pembelian rata-rata 3.261.005. Penggunaan metode klasifikasi decision tree dapat membantu menjelaskan karakteristik dari masing-masing kluster yang terbentuk dengan menggunakan label nama kluster.

Pada penelitian selanjutnya penulis memberikan saran-saran yang perlu dipertimbangkan sebagai bentuk pengembangan penelitian yang telah penulis lakukan yaitu dapat menggunakan atribut lain, kemudian untuk proses klusterisasi juga dapat menggunakan metode klusterisasi yang lain seperti K-Medoids, Fuzzy C-Means dan sebagainya.

REFERENSI

- [1.] Rizki, B., Ginasta, N. G., Tamrin, M. A., & Rahman, A. (2020). *Customer Loyalty Segmentation on Point of Sale System Using Recency-Frequency-Monetary (RFM) and K-Means*. Jurnal Online Informatika,5(2),130. <https://doi.org/10.15575/join.v5i2.511>.
- [2.] Arhami, M. dan Nasir, M. 2020. *Data Mining: Algoritma dan Implementasi*. Penerbit Andi: Banda Aceh.
- [3.] Suyanto, 2017, *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika.
- [4.] Anggara, M., Sujiani, H., & Nasution, H. (2016). *Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokan Member Di Alvaro Fitness*. Jurnal Sistem dan Teknologi Informasi (JUSTIN), 1(1).
- [5.] Widyadhana, D., Hastuti, R. B., & Kharisudin, I. (2021). *Perbandingan Analisis Klaster K-Means dan Average Linkage untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah*. PRISMA, Prosiding Seminar Nasional Matematika, 4.
- [6.] Anggara, M., Sujiani, H., & Nasution, H. (2016). *Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokan Member Di Alvaro Fitness*. Jurnal Sistem dan Teknologi Informasi (JUSTIN), 1(1).
- [7.] Wardono, Sunarmi, & Wirawan, M. R. (2019). *Pengelompokan Kabupaten/Kota Di Provinsi Jawa Tengah Berdasarkan Indikator Kesejahteraan Dengan Metode K-Means Cluster*. Seminar Nasional Edusaintek, 599–610.
- [8.] Solichin, A., & Khairunnisa, K. (2020). *Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means*. Fountain of Informatics Journal,5(2),52. <https://doi.org/10.21111/fij.v5i2.4905>.
- [9.] Suyanto. (2018). *Machine Learning: Tingkat Dasar dan Lanjut*. Bandung: Informatika.
- [10.] Dinata, R. K., Hasdyna, N., & Azizah, N. (2020). *Analisis K-Means Clustering pada Data Sepeda Motor*. Informatics Journal, 5(1).
- [11.] Rudiarto, S. (2018). *Implementasi Algoritma K-Means Clustering Pada Aplikasi Pencari Pelanggan Potensial Pada Restoran Xyz*. Jurnal Ilmu Teknik dan Komputer,(1). <https://doi.org/10.30646/tikomsin.v7i2.435>.
- [12.] Adiya, M. H., & Desnelita, Y. (2019). *Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru*. Jurnal Nasional Teknologi dan Sistem Informasi, 05(01).
- [13.] Harahap, B. (2019). *Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung)*. pp. 394-403, 2(1).
- [14.] Prasojo, R., Utami, Y. R. W., & Vlandari, R. T. (2019). *Implementasi K-Means Clustering Pada Pengelompokan Potensi Kerjasama Pelanggan*. Jurnal Teknologi Informasi dan Komunikasi (TIKoSIN),7(2). <https://doi.org/10.30646/tikomsin.v7i2.435>.