

ANALISIS SEGMENTASI PELANGGAN MENGGUNAKAN K-MEANS CLUSTERING STUDI KASUS APLIKASI ALFAGIFT

Satria Ardi Perdana¹⁾, Sara Famayla Florentin²⁾, dan Agus Santoso³⁾

^{1,2,3}Magister Ilmu Komputer, Universitas Budi Luhur

^{1,2,3}Jalan Ciledug Raya RT 10/RW 02 Petukangan Utara, Jakarta Selatan, 12260

E-mail : satria.perdana.2022@gmail.com¹⁾, sarafamayla@gmail.com²⁾, agussantoso552@gmail.com³⁾

ABSTRAK

Segmentasi pelanggan adalah strategi yang diterapkan dengan membagi pelanggan ke dalam kelompok yang berbeda dengan karakteristik, perilaku, atau kebutuhan yang berbeda. Segmentasi pelanggan sangat penting karena dapat digunakan sebagai data pendukung untuk mengetahui loyalitas pelanggan dan menentukan strategi pemasaran yang efektif dan efisien bagi perusahaan. Untuk menentukan segmentasi pelanggan, dilakukan dengan menggunakan algoritma *K-Means clustering*. Proses *clustering* dilakukan dengan cara mengelompokkan data histori transaksi pelanggan pengguna aplikasi Alfagift selama bulan Juni 2021 berdasarkan lima kategori yaitu umur, jenis kelamin, frekuensi pembelian, tipe pembayaran dan kota. Proses analisis segmentasi pelanggan menggunakan metode CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang meliputi enam tahapan yaitu pemahaman bisnis (*Business Understanding*), pemahaman data (*Data Understanding*), persiapan data (*Data Preparation*), pemodelan (*Modeling*), evaluasi (*Evaluation*), dan penerapan (*Deployment*). Algoritma *K-Means* digunakan untuk membentuk kluster dan metode *Elbow* digunakan untuk mengevaluasi kluster-kluster yang terbentuk. Algoritma *K-means* digunakan untuk pembentukan kluster dan metode *Elbow* digunakan untuk mengevaluasi kluster-kluster yang terbentuk. Hasil *k* optimum yang terbentuk dari hasil perhitungan SSE (*Sum Squared Error*) dengan metode *Elbow* adalah tiga *cluster* dengan selisih nilai SSE terbesar yaitu 1119.28644. Hasil pengelompokan pelanggan sebanyak tiga *cluster*; *cluster* pertama berjumlah 7.219 pelanggan, *cluster* kedua sebanyak 6.902 pelanggan, dan *cluster* ketiga sebanyak 5.371 pelanggan. Hasil *K-Means clustering* tersebut selanjutnya diinterpretasikan sebagai data yang akan digunakan dalam menentukan strategi pemasaran yang tepat.

Kata Kunci: *Segmentasi Pelanggan, K-Means Clustering, Elbow Method, CRISP-DM, Data Mining*

1. PENDAHULUAN

Pada masa pandemi Covid-19, tren belanja *online* di kalangan masyarakat meningkat secara signifikan. Belanja *online* telah menjadi bagian gaya hidup masyarakat. Hal tersebut dikarenakan banyaknya kemudahan yang didapat untuk memenuhi kebutuhan primer sampai dengan tersier. Terdapat beberapa alasan masyarakat yang lebih memilih melakukan pembelian secara *online* yaitu karena adanya kemudahan untuk berbelanja yang dapat dilakukan dimana saja, lebih cepat dan praktis hanya dengan memanfaatkan *smartphone* sebagai media dalam melakukan transaksi pembelian. Pelanggan tidak perlu lagi datang langsung ke toko, namun semua dapat dilakukan dimanapun dengan cara yang sangat mudah yaitu mengunduh aplikasi belanja *online*. Salah satu *platform* belanja *online* yang sedang berkembang saat ini adalah aplikasi Alfagift, yang dimiliki oleh PT *Global Loyalty* Indonesia.

Alfagift merupakan sebuah aplikasi belanja *online* yang dimiliki oleh PT *Global Loyalty* Indonesia, yang dikategorikan sebagai perusahaan besar dengan konsumen yang berasal dari berbagai kalangan di seluruh Indonesia. Sebagai perusahaan yang baru berjalan selama 3 tahun, mempertahankan pelanggan yang sudah ada menjadi hal yang sangat penting bagi keberlangsungan perusahaan. PT *Global Loyalty* Indonesia sampai saat ini belum pernah menganalisis segmentasi pelanggan dalam melakukan

pembelian, sehingga belum ada pengelompokan pelanggan berdasarkan kesamaan karakteristik seperti usia, jenis kelamin, frekuensi pembelian, demografi, dan aspek lainnya. Kondisi saat ini, data riwayat transaksi pembelian pelanggan belum dimanfaatkan dengan maksimal sehingga kurang berdaya guna. Hal tersebut membuat perusahaan dirasa belum maksimal dalam memenuhi harapan pelanggan dalam peningkatan loyalitas pelanggan. Selain itu, strategi pemasaran saat ini juga dirasa belum efektif karena penawaran yang dilakukan oleh perusahaan kepada setiap pelanggan masih bersifat umum, perusahaan belum menawarkan produk ataupun promosi yang benar-benar dibutuhkan oleh pelanggan.

Melalui penelitian ini, penulis berusaha memberikan solusi kepada perusahaan untuk meningkatkan efisiensi strategi pemasaran yang sangat berpengaruh terhadap peningkatan loyalitas pelanggan dengan cara melakukan analisis segmentasi pelanggan. Analisis segmentasi ini mampu memberikan gambaran tentang karakteristik pelanggan di beberapa *cluster* yang berbeda. Pelanggan dikelompokkan menjadi beberapa *cluster* yang dibedakan berdasarkan perilakunya dalam melakukan transaksi pembelian.

Adapun penelitian sebelumnya yang berjudul “Analisis segmentasi konsumen menggunakan algoritma *K-Means*” dilakukan oleh Sulistyowati, Basma Eno Ketherin, Amalia Anjani Arifiyanti dan Anwar Sodik pada tahun 2018. Proses *clustering* yang dilakukan dengan *K-Means* menghasilkan gambaran karakteristik konsumen yang dikelompokkan. Proses pendataan penjualan PT Calista Alba menghasilkan tiga *cluster*/kelompok konsumen berdasarkan tiga variabel yaitu jenis sepeda motor, jenis pembelian dan pekerjaan. Hasil pengelompokan data konsumen dapat digunakan sebagai informasi pendukung untuk bagian pemasaran dalam menentukan strategi pemasaran. Pengujian perbandingan menggunakan *software* SPSS menghasilkan jumlah *cluster* yang sama dengan penelitian yang dilakukan. Namun demikian masih terdapat perbedaan distribusi persentase jumlah anggota *cluster*, perbedaan rata-rata perbandingan kedua aplikasi adalah 7%. (Sulistyowati dkk., 2018).

Penelitian berjudul “Penerapan *Clustering Algorithm* Untuk Mendukung Promosi Server Pulsa *Reload*” dilakukan oleh Firman Nurdiyansyah, Samsul Arifin, dan Fitra Marisa pada tahun 2018. Dengan pendekatan *clustering algorithm* yang menggunakan *K-Means*, penelitian ini menghasilkan tiga *cluster*/kelompok konsumen berdasarkan jumlah transaksi dan domisili yaitu: kelompok 1 merupakan pelanggan yang memiliki transaksi banyak dan berdomisili di Malang Kota, kelompok 2 adalah pelanggan dengan jumlah transaksi sedang dan berdomisili di Malang Kota dan Malang Kabupaten, dan kelompok 3 merupakan kelompok pelanggan dengan jumlah transaksi sedikit, berdomisili di luar Malang dan pulau Jawa (Nurdiyansyah dkk., 2018).

Penelitian berjudul “*K-Means Clustering* dengan Metode *Elbow* untuk Pengelompokan Kabupaten dan Kota di Jawa Timur Berdasarkan Indikator Kemiskinan”, dilakukan oleh Anita Fitria Febrianti, Antonito Hornay Cabral, dan Gangga Anuraga pada tahun 2018. Tujuan penelitian ini yaitu untuk mengelompokkan kota kabupaten dan kota di Jawa Timur berdasarkan indikator kemiskinan agar mudah dalam menangani permasalahan dan mengetahui kabupaten dan kota yang membutuhkan bantuan pemerintah. Metode yang digunakan yaitu *K-Means Clustering* dengan metode *Elbow*. Adapun hasil dari penelitian ini yaitu terdapat 4 *cluster* dimana *cluster* 1 terdiri dari 10 kabupaten, *cluster* 2 terdiri dari 5 kabupaten, *cluster* 3 terdiri dari 19 kabupaten, dan *cluster* 4 terdiri dari 4 kabupaten. Hasil karakteristik *cluster* pertama memiliki persentase penduduk 15 tahun ke atas pendidikan SLTA yang tinggi, *cluster* 2 merupakan rumah tangga yang menerima raskin yang cukup tinggi. *Cluster* 3 memiliki angka partisipasi sekolah yang rendah, dan *cluster* 4 merupakan penduduk 15 tahun keatas yang bekerja pada sektor pertanian yang cukup tinggi (Febrianti dkk., 2018).

Penelitian berjudul “Metode *Elbow* dan *K-Means* Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional” dilakukan oleh Ninik Tri Hartanti pada tahun 2020. Penerapan algoritma *K-Means* pada penelitian ini adalah membentuk kelompok siswa berdasarkan nilai rata-rata tiga mata pelajaran UN (matematika, bahasa Indonesia dan bahasa Inggris). Metode *Elbow* digunakan untuk

menentukan jumlah *cluster* yang optimal, menunjukkan bahwa jumlah *cluster* yang optimal adalah 3, sehingga diperoleh 3 *cluster* yang terdiri dari “Siap”, “Cukup Siap” dan “Tidak Siap”. Dengan masing-masing kategori terdiri dari 7 siswa, 30 siswa, dan 29 siswa. Tujuan dari penelitian ini adalah mengelompokkan skor nilai siswa SMK Syubbanul Wathon terhadap tiga mata pelajaran UN (Matematika, Bahasa Indonesia, Bahasa Inggris), sehingga memudahkan pihak sekolah mengambil tindakan lebih lanjut demi keberhasilan siswa dan predikat sekolah. (Hartanti, 2020).

Metode yang digunakan penelitian ini adalah *K-Means clustering*, yang merupakan salah satu teknik data *mining* yang dapat digunakan untuk mengelompokkan berdasarkan kemiripan karakteristik tertentu dimana data-data yang memiliki kemiripan akan berada pada *cluster* yang sama.

Hasil dari analisis segmentasi pelanggan ini dapat mempermudah perusahaan dalam mengetahui karakteristik pelanggan. Selanjutnya, perusahaan dapat dengan mudah melakukan perencanaan strategi pemasaran yang lebih efektif untuk meningkatkan loyalitas pelanggan sehingga tingkat kesetiaan pelanggan dalam berbelanja pada aplikasi Alfagift dapat meningkat.

2. RUANG LINGKUP

Dalam penelitian ini permasalahan mencakup:

1. Penelitian ini menggunakan data riwayat transaksi pembelian pelanggan pada aplikasi belanja *online* Alfagift.
2. Data yang digunakan dalam penelitian ini yaitu riwayat data transaksi penjualan pada aplikasi Alfagift pada bulan Juni 2021 di Provinsi DKI Jakarta.
3. Penelitian hanya dilakukan di divisi IT untuk mengolah data dan tidak sampai melakukan pemasaran oleh divisi *marketing*.

3. BAHAN DAN METODE

Pada segmen ini merupakan penjelasan teori yang dipakai untuk penelitian yang mencakup rumus dalam solusi untuk kasus.

3.1 Konsep Dasar Pemasaran

Terdapat beberapa hal penting yang perlu diperhatikan untuk mencapai keberhasilan sebuah bisnis. Salah satu hal yang paling esensial yakni strategi pemasaran. Tanpa adanya strategi pemasaran yang tepat, akan sulit untuk sebuah perusahaan dapat berkembang. Dalam strategi pemasaran ada tiga elemen, yaitu:

1. Segmentasi

Segmentasi merupakan proses membagi pasar menjadi kelompok-kelompok konsumen yang lebih homogen, dimana pada setiap kelompok konsumen dapat dipilih sebagai target pasar oleh perusahaan dengan strategi yang dimiliki perusahaan.

2. Target

Target merupakan cara suatu perusahaan memilih satu atau lebih segmen pasar dan dalam penentuan target pasar perusahaan harus menggunakan konsep prioritas, variabilitas dan fleksibilitas. Prioritas digunakan karena perusahaan tidak bisa melayani setiap pelanggan yang ada di pasar. Variabilitas digunakan ketika perusahaan menghadapi suatu situasi persaingan yang sudah meningkat, karena perusahaan tidak dapat memberikan pelayanan yang sama kepada setiap konsumen yang akan diprioritaskan. Konsep fleksibilitas berhubungan dengan variasi, semakin sama variasinya maka semakin tidak optimal, perusahaan dapat optimal ketika melayani pelanggan dengan memberikan banyak variasi, karena Semakin fleksibel suatu perusahaan dapat memberikan variasi tanpa perlu mengeluarkan biaya tambahan.

3. Posisi

Posisi merupakan suatu tindakan mendesain penawaran perusahaan sehingga menciptakan tempat dan nilai tersendiri dalam pandangan konsumen. PT Global *Loyalty* Indonesia saat ini masih perlu mengembangkan aplikasi Alfagift untuk meningkatkan jumlah pelanggan serta loyalitasnya. Berdasarkan pemaparan yang berkaitan dengan strategi pemasaran, segmentasi menjadi satu poin penting yang perlu diterapkan dalam sebuah bisnis. Oleh karena itu, penulis mengambil strategi segmentasi dalam penelitian yang dilakukan.

3.2 Segmentasi Pelanggan

Dalam penyusunan sebuah strategi bisnis, identifikasi terhadap konsumen sangatlah dibutuhkan. Hal tersebut bertujuan untuk menggali informasi tentang kebutuhan pelanggan, perilaku, dan karakteristiknya. Salah satu komponen pendukung dalam meningkatkan efektivitas strategi pemasaran yang telah direncanakan yaitu segmentasi pelanggan. Segmentasi pelanggan adalah proses membagi pelanggan menjadi sub kelompok yang berbeda, bermakna, dan homogen berdasarkan atribut dan karakteristiknya (Malim dkk., 2020). Hal tersebut memungkinkan suatu perusahaan dapat memahami atau mempelajari pelanggan mereka dan membangun strategi untuk menjalin hubungan dengan pelanggan sesuai dengan karakteristik pelanggan tersebut

Segmentasi merupakan salah satu cara untuk memiliki komunikasi yang lebih dengan pelanggan (Putri dkk., 2019). Tujuan dari segmentasi adalah untuk menyesuaikan produk, jasa, dan pesan pemasaran untuk setiap segmen. Dari pernyataan tersebut, segmentasi pelanggan memiliki peran yang sangat penting dalam peningkatan pemasaran dan mempertahankan konsumen lama. Dengan menggunakan

segmentasi pelanggan, perusahaan juga akan mampu memahami setiap kebutuhan konsumen dengan lebih baik.

Karakteristik pelanggan dapat direpresentasikan oleh beberapa kategori variabel yang terkait dengan pengelompokan, yaitu (Yustina dkk., 2021):

1. *Demographics*

Pengelompokan berdasarkan umur, jenis kelamin, besarnya keluarga, besarnya kediaman, siklus kehidupan keluarga, pemasukan, pekerjaan atau profesi, pendidikan, kepemilikan rumah, status sosial ekonomi, agama, kewarganegaraan.

2. *Psychographics*

Dalam kategori ini pengelompokan dikategorikan berdasarkan kepribadian, gaya hidup, nilai-nilai, dan sikap.

3. *Behavior*

Dari segi *behaviour*, karakteristik pelanggan dikelompokkan berdasarkan manfaat yang dicari, status pembelian, tingkat penggunaan produk, frekuensi pembelian.

4. *Geographic*

Merupakan pengelompokan pelanggan sesuai dengan negara, provinsi, kota, kode pos, dan iklim.

3.3 Konsep Clustering

Clustering atau pengelompokan data merupakan kegiatan untuk mencari kesamaan dalam data dan menempatkan data yang sama ke dalam kelompok-kelompok (Kamila dkk., 2019). Pengelompokan data dilakukan menggunakan algoritma yang ditentukan dan kemudian *dataset* akan diproses oleh algoritma yang akan berjalan untuk dikelompokkan berdasarkan karakteristiknya. Data yang mirip karakteristiknya dengan data yang lain akan dikelompokkan ke dalam satu *cluster*, sedangkan data yang jauh atau berbeda karakteristiknya dari data yang lain akan dikelompokkan ke dalam kluster yang berbeda. *Clustering* merupakan salah satu algoritma data *mining* yang bersifat *unsupervised*, yaitu algoritma bekerja tanpa arahan atau tanpa memerlukan data training (Selvida, 2019).

Ada dua jenis *clustering* yang sering digunakan dalam pengelompokan data yaitu *hierarchical clustering* dan *non-hierarchical clustering* (Selvida, 2019). *Hierarchical clustering* adalah suatu metode yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan paling dekat, kemudian pengelompokan diteruskan ke objek lain yang memiliki kedekatan kedua dan seterusnya sehingga *cluster* akan membentuk pohon hierarki (tingkatan) antar objek dari yang paling mirip sampai ke yang paling tidak mirip. Metode *non-hierarchical clustering* dimulai dengan menentukan jumlah cluster terlebih dahulu, kemudian proses *clustering*

dilakukan tanpa mengikuti proses hierarki dan biasa disebut dengan metode *K-Means Clustering*.

Berdasarkan landasan teori, dapat disimpulkan bahwa *clustering* merupakan proses mengelompokkan sekelompok data yang memiliki kemiripan karakteristik antara satu dan lainnya. Pada umumnya, metode *clustering* digunakan untuk proses analisa masalah bisnis seperti segmentasi pasar, pemetaan wilayah, dan lain sebagainya. Dalam penelitian ini, penulis menggunakan *non-hierarchical clustering* dimana penentuan nilai k ditentukan diawal, sedangkan algoritma yang digunakan yaitu K-Means.

3.4 Data Mining

Data mining merupakan disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data yang besar (Atika & Priatna, 2020). Data mining adalah istilah yang digunakan untuk menggambarkan penemuan pengetahuan di dalam *database*, serta merupakan proses yang menggunakan statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstrak, mengidentifikasi informasi yang berguna dan pengetahuan di dalam *database* yang besar (Indriyani & Putera, 2020). Data mining adalah subdomain lain dari AI dan dapat didefinisikan sebagai proses yang bertujuan untuk menghasilkan pengetahuan dari data dan menyajikan temuan secara komprehensif kepada pengguna (Schuh dkk., 2019). Teknik data mining yang digunakan dalam penelitian ini yaitu proses pengambilan data histori transaksi pembelian pelanggan pengguna Alfagift selama bulan Juni 2021 yang bersumber dari *database e-commerce* untuk mengetahui pola karakteristik pelanggan. Metode dalam data mining yaitu (Atika & Priatna, 2020):

1. Estimasi

Untuk menerka sebuah nilai yang belum diketahui, misal menerka penghasilan seseorang ketika informasi mengenai orang tersebut diketahui. Contoh metodenya adalah *Linear Regression (LR)*, *Neural Network (NN)*, *Deep Learning (DL)*, *Support Vector Machine (SVM)*, *Generalized Linear Model (GLM)*.

2. Prediksi

Untuk memperkirakan nilai masa mendatang, misal memprediksi stok barang satu tahun ke depan. Contoh metodenya adalah *Linear Regression (LR)*, *Neural Network (NN)*, *Deep Learning (DL)*, *Support Vector Machine (SVM)*, *Generalized Linear Model (GLM)*.

3. Klasifikasi

Merupakan proses penemuan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Contoh metodenya adalah *Decision Tree (CART, ID3, C4.5, Credal DT, Credal C4.5, Adaptative Credal C4.5)*, *Naive Bayes (NB)*, *K-Nearest Neighbor (KNN)*, *Linear Discriminant Analysis (LDA)*, *Logistic Regression (LogR)*.

4. Pengelompokan(*clustering*)

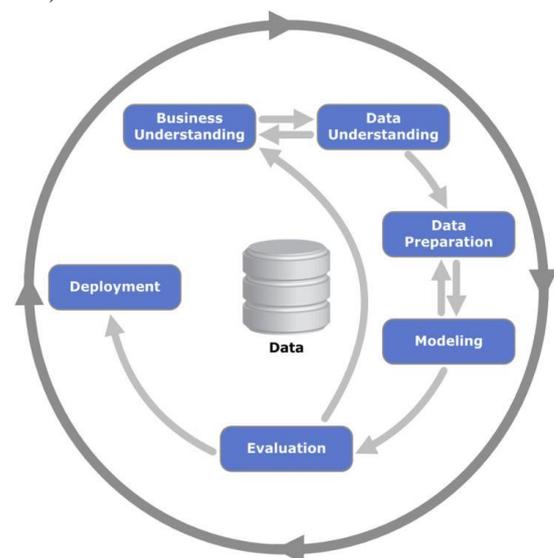
Yaitu pengelompokan mengidentifikasi data yang memiliki karakteristik tertentu. Contoh metodenya adalah *K-Means*, *K-Medoids*, *Self-Organizing Map (SOM)*, *Fuzzy C-Means (FCM)*.

5. Asosiasi

Dinamakan juga analisis keranjang pasar dimana fungsi ini mengidentifikasi item-item produk yang kemungkinan dibeli konsumen bersamaan dengan produk lain. Contoh metodenya adalah *FP-Growth*, *A Priori*, *Coefficient of Correlation*, *Chi Square*.

Dalam penelitian yang sedang dilakukan, metode *data mining* yang digunakan yaitu *clustering* metode *K-Means*, dengan tujuan untuk mengelompokkan seluruh pelanggan ke dalam beberapa *cluster* berdasarkan beberapa kesamaan perilakunya. Selain itu metode *K-Means* juga digunakan untuk menentukan jumlah kelompok pelanggan atau *cluster* dengan tujuan agar strategi pemasaran dapat tepat sasaran.

Terdapat kerangka kerja yang dapat digunakan untuk solusi data mining, yaitu *CRISP-DM (Cross Industry Standard Process for Data Mining)*. *CRISP-DM* merupakan suatu standar yang telah dikembangkan pada tahun 1996 yang ditujukan untuk melakukan proses analisis dari suatu industri sebagai strategi pemecahan masalah dari bisnis atau unit penelitian (Schröer dkk., 2021) Dalam *CRISP-DM* terdapat tahapan-tahapan, dan terdapat tugas-tugas di dalam setiap tahapan serta hubungan antar tugas tersebut. Terdapat enam fase dalam *CRISP-DM* seperti digambarkan pada Gambar 1 (Schröer dkk., 2021).



Gambar 1. Tahapan pada CRISP-DM

Proses pengolahan data dengan teknik data mining pada Gambar 1 dapat dijelaskan:

1. *Business Understanding* (Pemahaman Bisnis)

Tahap pertama adalah memahami tujuan dan kebutuhan dari sudut pandang bisnis, kemudian menerjemahkan pengetahuan ini ke dalam pendefinisian masalah dalam data mining. Selanjutnya, akan ditentukan rencana dan strategi untuk mencapai tujuan tersebut.

2. *Data Understanding* (Pemahaman Data)

Tahap ini dimulai dengan pengumpulan data yang kemudian akan dilanjutkan dengan proses untuk mendapatkan pemahaman yang mendalam tentang data, mengidentifikasi masalah kualitas data, atau untuk mendeteksi adanya bagian yang menarik dari data yang dapat digunakan untuk hipotesis untuk informasi yang tersembunyi.

3. *Data Preparation* (Persiapan Data)

Tahap ini meliputi kegiatan untuk membangun kumpulan data akhir (data yang akan diproses pada tahap pemodelan) dari data mentah. Tahap ini dapat diulang beberapa kali. Pada tahap ini juga mencakup pemilihan tabel, *record*, dan atribut-atribut data, termasuk proses pembersihan dan transformasi data untuk kemudian dijadikan masukan dalam tahap pemodelan.

4. *Modeling* (Pemodelan)

Dalam tahap ini akan dilakukan pemilihan dan penerapan berbagai teknik pemodelan dan beberapa parameternya akan disesuaikan untuk mendapatkan nilai yang optimal. Secara khusus, ada beberapa teknik berbeda yang dapat diterapkan dalam pemecahan masalah data mining. Di pihak lain, ada teknik pemodelan yang membutuhkan format khusus sehingga pada tahap ini masih memungkinkan kembali ke tahap sebelumnya.

5. *Evaluation* (Evaluasi)

Pada tahap ini, model sudah terbentuk dan diharapkan memiliki kualitas baik jika dilihat dari sudut pandang analisis data. Pada tahap ini akan dilakukan evaluasi terhadap apakah model dapat mencapai tujuan yang ditetapkan pada fase awal (pemahaman data). Kunci dari tahap ini adalah menentukan ada atau tidak masalah bisnis yang dipertimbangkan. Pada akhir dari tahap ini harus ditentukan penggunaan hasil proses data mining.

6. *Deployment* (Penerapan)

Pada tahap terakhir, pengetahuan atau informasi yang telah diperoleh akan diatur dan dipresentasikan dalam bentuk khusus sehingga dapat digunakan oleh pengguna. Tahap pengembangan dapat berupa pembuatan laporan sederhana atau mengimplementasikan proses data mining yang berulang dalam perusahaan.

Berdasarkan penjabaran, penulis menggunakan kerangka kerja CRISP-DM sebagai proses dalam melakukan analisis segmentasi pelanggan pengguna aplikasi Alfagift. Proses awal dimulai dengan mempelajari dan menggali informasi terkait tujuan perusahaan sebagai pelaku bisnis, yang kemudian diformulasikan menjadi masalah penelitian. Proses inilah yang akan dilaksanakan untuk mendapatkan

sebuah informasi baru bagi perusahaan guna kepentingan strategi pemasaran.

3.5 Metode *Elbow*

Metode *Elbow* merupakan salah satu metode untuk menentukan jumlah *cluster* yang tepat melalui persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik (Dewi & Pramita, 2019). Penerapan metode *Elbow* digunakan untuk menentukan k optimal, yakni optimasi jumlah *cluster* yang dapat terbentuk yang selanjutnya akan digunakan dalam proses *K-Means clustering*. Metode *Elbow* digunakan untuk menghasilkan informasi dalam menentukan jumlah *cluster* (k) terbaik dengan melihat persentase perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik (Nainggolan dkk., 2019). Metode *Elbow* memberikan ide dengan memilih nilai *cluster*, kemudian menjumlahkan nilai *cluster* tersebut untuk dijadikan model data dalam menentukan *cluster* terbaik dan persentase perhitungan yang dihasilkan merupakan perbandingan antara jumlah *cluster* yang ditambahkan.

Hasil persentase yang berbeda dari masing-masing nilai *cluster* dapat ditampilkan dengan menggunakan grafik sebagai sumber informasinya. Jika nilai *cluster* pertama dengan nilai *cluster* kedua memberikan sudut pada grafik atau nilai memiliki penurunan terbesar maka itu merupakan nilai *cluster* terbaik. Untuk mendapatkan perbandingan dengan menghitung SSE (*Sum of Square Error*) dari setiap nilai *cluster*. Semakin besar jumlah *cluster* nilai SSE akan semakin kecil.

Proses tahapan *Elbow* dalam menentukan nilai k pada *K-Means clustering* :

1. Tentukan nilai awal k , misal $k=2$.
2. Tambahkan nilai k dengan 1
3. Menghitung hasil SSE dari nilai k yang mengalami penurunan drastis bentuk siku. Itulah merupakan nilai k terbaik.

Untuk menghitung SSE dapat digunakan rumus seperti pada persamaan (1) :

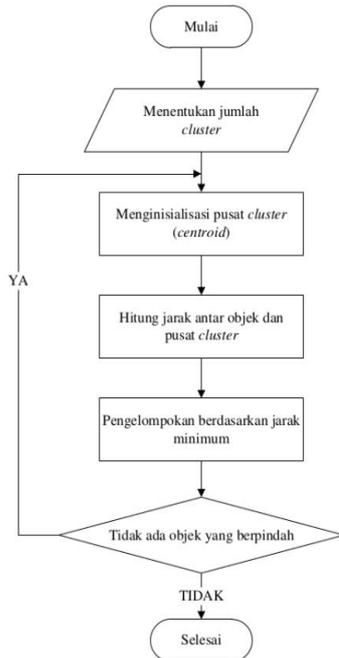
$$SSE = \sum_{i=1}^n (d)^2 \quad (1)$$

3.6 Algoritma *K-Means Clustering*

K-Means yaitu salah satu dari metode pengelompokan data non-hirarki (sekatan) yang dapat membagi data ke dalam bentuk dua kelompok ataupun lebih (Gustientiedina dkk., 2019). *Clustering* adalah metode yang digunakan dalam data mining dengan cara kerjanya mencari data dan mengelompokkan data yang mempunyai kemiripan karakteristik antara data satu dengan data lainnya yang telah diperoleh (Dinata., dkk 2020). *K-Means clustering* merupakan salah satu data *clustering* non-hirarki yang berusaha membagi data yang ada ke

dalam bentuk satu atau lebih *cluster* atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lainnya (Mardalius, 2018).

Dalam penerapannya, *K-Means clustering* membutuhkan proses yang sistematis yang dimulai dengan penentuan jumlah *cluster* hingga ditemukan jarak terpendek antar *cluster*. Proses *clustering* menggunakan algoritma *K-Means* ditunjukkan oleh *flowchart* pada Gambar 2. (Adiana dkk., 2018).



Gambar 2. Gambaran Flowchart Algoritma K-Means Clustering

Berdasarkan Gambar 2. proses komputasi pada algoritma *K-Means clustering* dijabarkan (Adiana dkk., 2018):

1. Menentukan jumlah *cluster* (k).
2. Inisialisasi vektor *centroid cluster* acak
3. Untuk setiap data vektor, hitung jarak antar data vektor dengan masing-masing kluster *centroid* yang akan ditentukan vektor data minimum dengan kluster dan jarak yang dihitung menggunakan persamaan (2).

$$d(z_p, M_j) = \sqrt{\sum_{k=1}^d (z_p, k - M_j, k)^2} \quad (2)$$

4. Hitung ulang *cluster centroid* menggunakan persamaan (3).

$$M_j = \frac{1}{n_j} (\sum Z_p) \quad (3)$$

5. Ulangi langkah 3) dan 4) hingga berhenti memenuhi kriteria. Kriteria yang memuaskan dapat berupa jumlah iterasi atau perubahan posisi *centroid* dalam iterasi berurutan.

3.7 Normalisasi Min - Max

Pada proses persiapan data, normalisasi diperlukan agar data tidak terlalu jauh perbedaannya, sehingga digunakan *range* antara 0 sampai 1 dengan metode Min-Max. *Range* 0 sampai 1 digunakan agar nilai normalisasi pada penelitian ini memiliki nilai antara 0 dan 1, dimana 0 adalah nilai minimal dan 1 adalah nilai maksimal (Monalisa, 2018).

Tujuan dari normalisasi menurut (Yunita, 2018):

1. Untuk menghilangkan kerangkapan data
2. Untuk mengurangi kompleksitas
3. Untuk mempermudah modifikasi data

Dari penjabaran terkait tujuan normalisasi data, tidak dapat dipungkiri bahwa normalisasi *min-max* sangat dibutuhkan dalam proses persiapan agar basis data menjadi mudah diakses, data mudah dikelola, dan meminimalkan tempat penyimpanannya.

Persamaan (4) digunakan untuk melakukan proses normalisasi data:

$$X_n = \frac{x_0 - x_{min}}{x_{max} - x_{min}} \quad (4)$$

Keterangan (4):

X_n = nilai data normal

x_0 = nilai data aktual

x_{min} = nilai minimum data aktual keseluruhan

x_{max} = nilai maximum data aktual keseluruhan

4. PEMBAHASAN

Pada analisis segmentasi pelanggan dilakukan dengan menggunakan metode CRISP-DM, yaitu CRISP-DM (*Cross Industry Standard Process for Data Mining*)

4.1 Pemahaman Bisnis (*Business Understanding*)

Aplikasi Alfagift merupakan aplikasi berbelanja *online* yang terintegrasi oleh toko Alfamart di seluruh Indonesia. Aplikasi ini menyediakan kemudahan dalam metode pembayaran seperti *gopay*, *shopeepay*, *bank transfer*, *credit card* dan *virgo*. Fitur promo juga disediakan untuk menarik minat pembeli berupa potongan harga, harga spesial dan *cashback* dengan syarat tertentu. Selain itu, strategi promosi juga dikirimkan kepada seluruh pelanggan melalui pesan *email* dan *Whatsapp*. Untuk saat ini promosi tersebut dianggap kurang efektif karena tidak memperhatikan kebutuhan pelanggan. Oleh karena itu, diperlukan segmentasi pelanggan untuk melakukan strategi pemasaran terbaik dengan menggunakan algoritma *K-Means Clustering*.

4.2 Pemahaman Data (*Data Understanding*)

Data yang akan digunakan meliputi data *history* transaksi pelanggan seluruh provinsi Indonesia. Berdasarkan tujuan segmentasi pelanggan, dilakukan pemahaman data pada data *history* untuk memperoleh data yang lebih kecil. Setelah melewati proses pemahaman data, maka data yang akan digunakan adalah data *history* pada bulan Juni 2021 di provinsi DKI Jakarta yaitu, Jakarta Barat, Jakarta Selatan, Jakarta Pusat, Jakarta Timur, Jakarta Utara dan Kepulauan Seribu.

4.3 Persiapan Data (*Data Preparation*)

Proses *preparation* dalam penelitian ini mencakup empat hal utama yaitu data *selection* (pemilihan data), data *reduction* (reduksi data), data *cleansing* (pembersihan data) dan data *transformation* (transformasi data)

1. Pemilihan Data (*Data Selection*)

Hal utama yang perlu dilakukan yaitu pemilihan atribut (*attribute selection*) dengan tujuan untuk mengurangi dimensi data, menghilangkan data yang tidak relevan serta dapat meningkatkan akurasi. Data awal terdiri dari 32 atribut yang kemudian dilakukan pemilihan atribut menjadi 5 atribut.

2. Reduksi Data (*Data Reduction*)

Beberapa hal yang dilakukan dalam reduksi data adalah pemilihan data selama periode Juni 2021, mengeliminasi transaksi berdasarkan status pembayaran, serta mereduksi data yang bukan merupakan transaksi yang dilakukan di provinsi DKI Jakarta.

3. Pembersihan Data (*Data Cleansing*)

Beberapa hal yang dilakukan dalam proses pembersihan data (*data cleansing*) antara lain membersihkan data dari duplikasi dan ketidakkonsistenan data serta membuang data kosong (Null).

4. Transformasi Data

Proses transformasi data dilakukan dengan berbagai cara antara lain :

1) Perubahan Nama Atribut dan Tipe Data

Perubahan tipe data juga harus dilakukan agar data dapat diolah dalam pemodelan K- *Means*, karena salah satu syarat data agar dapat diproses harus berbentuk *numeric* (angka).

2) Pelabelan Data

Pada data pengujian ini, atribut yang membutuhkan pelabelan data adalah kota. Pemberian

label tersebut adalah : kota Jakarta Pusat menjadi 1, Jakarta Selatan menjadi 2, Jakarta Utara menjadi 3, Jakarta Barat menjadi 4, Jakarta Timur menjadi 5 dan selain itu menjadi 6.

3) Normalisasi Data

Proses normalisasi ini dilakukan menggunakan *query* dengan metode Min-Max *Normalization*. Metode *normalisasi* Min-Max mentransformasikan seluruh nilai atribut menjadi rentang nilai antara 0 hingga 1. Atribut yang dapat dinormalisasi kecuali atribut *id_pelanggan* karena atribut tersebut mengandung kode khusus untuk setiap data transaksi pelanggan, sehingga data yang telah dinormalisasi siap untuk diolah menggunakan algoritma K-Means *Clustering*.

4.4 Permodelan (*Modeling*)

Beberapa tahapan dalam proses pemodelan.

1. Proses Penentuan Anggota *Cluster*

Pada proses penentuan anggota tiap *cluster*, total pelanggan yang melakukan transaksi pada bulan Juni 2021 sebanyak 19.492 dengan pengujian sebanyak empat kali dengan jumlah *cluster* uji mulai dari $k=2$, $k=3$, $k=4$ dan $k=5$.

yang mana k disini berarti sebagai :

- 1) $k=2$ berarti terdapat 2 anggota *cluster*, yaitu C1 dan C2.
- 2) $k=3$ berarti terdapat 3 anggota *cluster*, yaitu C1, C2 dan C3.
- 3) $k=4$ berarti terdapat 4 anggota *cluster*, yaitu C1, C2, C3 dan C4.
- 4) $k=5$ berarti terdapat 5 anggota *cluster*, yaitu C1, C2, C3, C4 dan C5

Setelah mengetahui nilai k atau banyaknya *cluster* yang akan diproses maka selanjutnya adalah menginputkan nilai k tersebut pada program agar dapat dilihat hasilnya :

1) Proses input $k=2$

Jika nilai $k=2$ akan menghasilkan anggota *cluster* sesuai pada Gambar 3. Pada Gambar 3 terdapat field *cluster* yang berisi 0 yang artinya *cluster* 1 (C1) dan 1 artinya *cluster* 2 (C2). Total data yang diproses adalah 19492.

id_pelangi	normalisasi frekuensi	normalisasi total	normalisasi tipe pembayaran	normalisasi kota	cluster
432	0.0909091	0.0455935	0.8125	0.75	1
685	0.013986	0.0132183	0.03125	0.75	0
1826	0.048951	0.0129515	0.03125	0.25	0
2586	0.0979021	0.0241381	0.03125	0	0
3533	0.00699301	0.00691339	0.8125	0.5	1

Gambar 3. Hasil *cluster* $k=2$

2) Proses input k=3

Jika nilai k=3 akan menghasilkan anggota *cluster* sesuai pada Gambar 4. Pada Gambar 4 terdapat field *cluster* yang berisi 0 yang artinya *cluster* 1 (C1), 1 artinya *cluster* 2 (C2) dan 2 artinya *cluster* 3 (C3). Total data yang diproses adalah 19492.

dataset - DataFrame					
id pelanggi	normalisasi frekuensi	normalisasi total	normalisasi tipe pembayaran	normalisasi kota	cluster
432	0.0909091	0.0455935	0.8125	0.75	2
685	0.013986	0.0132183	0.03125	0.75	0
1826	0.048951	0.0129515	0.03125	0.25	1
2586	0.0979021	0.0241381	0.03125	0	1
3533	0.00699301	0.00691339	0.8125	0.5	2

Gambar 4. Hasil *cluster* k=3

3) Proses input k=4

Jika nilai k=4 akan menghasilkan anggota *cluster* sesuai pada Gambar 5. Pada Gambar 5 terdapat field *cluster* yang berisi 0 yang artinya *cluster* 1 (C1), 1 artinya *cluster* 2 (C2), 2 artinya *cluster* 3 (C3) dan 3 artinya *cluster* 4 (C4). Total data yang diproses adalah 19492.

dataset - DataFrame					
id pelanggi	normalisasi frekuensi	normalisasi total	normalisasi tipe pembayaran	normalisasi kota	cluster
432	0.0909091	0.0455935	0.8125	0.75	3
685	0.013986	0.0132183	0.03125	0.75	1
1826	0.048951	0.0129515	0.03125	0.25	2
2586	0.0979021	0.0241381	0.03125	0	2
3533	0.00699301	0.00691339	0.8125	0.5	0

Gambar 5. Hasil *cluster* k=4

4) Proses input k=5

Jika nilai k=5 akan menghasilkan anggota *cluster* sesuai pada Gambar 6. Pada Gambar 6 terdapat field *cluster* yang berisi 0 yang artinya *cluster* 1 (C1), 1 artinya *cluster* 2 (C2), 2 artinya *cluster* 3 (C3), 3 artinya *cluster* 4 (C4) dan 4 artinya *cluster* 5 (C5). Total data yang diproses adalah 19492.

dataset - DataFrame					
id pelanggi	normalisasi frekuensi	normalisasi total	normalisasi tipe pembayaran	normalisasi kota	cluster
432	0.0909091	0.0455935	0.8125	0.75	4
685	0.013986	0.0132183	0.03125	0.75	0
1826	0.048951	0.0129515	0.03125	0.25	2
2586	0.0979021	0.0241381	0.03125	0	2
3533	0.00699301	0.00691339	0.8125	0.5	1

Gambar 6. Hasil *cluster* k=5

2. Penentuan Nilai *Centroid*

Dalam proses *K-Means clustering* pada penelitian ini, penentuan nilai *centroid* dilakukan secara acak. Dalam hal ini, penulis menghitung *centroid* mulai dari k=2 sampai dengan k=5. Pada k=2 melalui proses iterasi sebanyak 3 kali. Pada k=3 melalui proses iterasi sebanyak 2 kali. Pada k=4 melalui proses iterasi sebanyak 3 kali. Pada k=5 melalui proses iterasi sebanyak 3 kali. dengan hasil perhitungan dapat dilihat pada tabel 1, tabel 2, tabel 3, tabel 4. Dari hasil ini dapat diketahui nilai pusat *cluster* atau *centroid* yang diperoleh untuk masing-masing input nilai *cluster* k=2, k=3, k=4

dan k=5 yang akan digunakan untuk perhitungan jarak dengan menggunakan *euclidean distance*.

3. Perhitungan Euclidean Distance

Pada penelitian ini perhitungan *Euclidean Distance* digunakan sebagai salah satu metode perhitungan jarak yang digunakan. Dataset akan dihitung jaraknya dengan pusat *cluster* yang sudah diketahui sebelumnya pada tabel 1, tabel 2, tabel 3, tabel 4. Iterasi akan dilakukan sampai anggota *cluster* saat ini nilainya sama dengan anggota *cluster* sebelumnya.

Tabel 1. Hasil Perhitungan Centroid $k=2$

Nilai ke :	Normalisasi Frekuensi	Normalisasi Total	Normalisasi Tipe Pembayaran	Normalisasi Kota
1	0.0107727	0.00945758	0.0714912	0.527822
2	0.00802969	0.00789432	0.872399	0.524549

Tabel 2. Hasil Perhitungan Centroid $k=3$

Nilai ke :	Normalisasi Frekuensi	Normalisasi Total	Normalisasi Tipe Pembayaran	Normalisasi Kota
1	0.00803087	0.00789369	0.872449	0.524479
2	0.0110174	0.00942492	0.0704258	0.268856
3	0.0104452	0.00950154	0.0729932	0.87245

Tabel 3. Hasil Perhitungan Centroid $k=4$

Nilai ke :	Normalisasi Frekuensi	Normalisasi Total	Normalisasi Tipe Pembayaran	Normalisasi Kota
1	0.00771536	0.00784236	0.872483	0.880734
2	0.0104471	0.00950104	0.072909	0.872426
3	0.0110174	0.00942492	0.0704258	0.268856
4	0.00826873	0.00793384	0.872335	0.253682

Tabel 4. Hasil Perhitungan Centroid $k=5$

Nilai ke :	Normalisasi Frekuensi	Normalisasi Total	Normalisasi Tipe Pembayaran	Normalisasi Kota
1	0.00771536	0.00784236	0.872483	0.880734
2	0.0092895	0.00882752	0.0742983	0.64271
3	0.00826873	0.00793384	0.872335	0.253682
4	0.0116175	0.00942051	0.0703307	0.175223
5	0.0118466	0.0106794	0.0686116	1

Tabel 5. Perhitungan Euclidean Distance $k=2$

Id Pelanggan	Normalisasi Frekuensi	Normalisasi Total	Normalisasi Tipe Pembayaran	Normalisasi Kota	dist0	dist1	Cluster
432	0.09090909	0.04559347	0.8125	0.75	0.77857874	0.25041256	1
685	0.01398601	0.01321834	0.03125	0.75	0.22584745	0.87087555	0
1826	0.04895105	0.01295148	0.03125	0.25	0.28332665	0.88578169	0
2586	0.0979021	0.02413814	0.03125	0	0.53667685	0.99550204	0
3533	0.00699301	0.00691339	0.8125	0.5	0.74154486	0.06475026	1
dst	dst	dst	dst	dst	dst	dst	dst

Tabel 6. Perhitungan Euclidean Distance $k=3$

Id Pelanggan	Normalisasi Frekuensi	Normalisasi Total	Normalisasi Tipe Pembayaran	Normalisasi Kota	Dist0	Dist1	Dist 2	Cluster
432	0.09090909	0.04559347	0.8125	0.75	0.25048709	0.8887431	0.75474585	0
685	0.01398601	0.01321834	0.03125	0.75	0.87094199	0.48276009	0.12947136	2
1826	0.04895105	0.01295148	0.03125	0.25	0.88580756	0.05780741	0.6250448	1
2586	0.0979021	0.02413814	0.03125	0	0.99550748	0.28562882	0.87793753	1
3533	0.00699301	0.00691339	0.8125	0.5	0.06477019	0.77725418	0.82801437	0
dst	dst	dst	dst	dst	dst	dst	dst	dst

Tabel 7. Perhitungan Euclidean Distance $k=4$

Id Pelanggan	Normalisasi Frekuensi	Normalisasi Total	Normalisasi Tipe Pembayaran	Normalisasi Kota	Dist0	Dist1	Dist 2	Dist 3	Cluster
432	0.09090909	0.04559347	0.8125	0.75	0.17039871	0.7548244	0.8887431	0.50809381	0
685	0.01398601	0.01321834	0.03125	0.75	0.85137141	0.12942201	0.48276009	0.97663526	1
1826	0.04895105	0.01295148	0.03125	0.25	1.05224783	0.62501572	0.05780741	0.84209157	2
2586	0.0979021	0.02413814	0.03125	0	1.22137839	0.87791004	0.28562882	0.883219	2
3533	0.00699301	0.00691339	0.8125	0.5	0.38543188	0.82807907	0.77725418	0.25348649	3
dst	dst	dst	dst	dst	dst	dst	dst	dst	dst



Tabel 8. Perhitungan *Euclidean Distance* k=5

id	normalisasi frekuensi	normalisasi total	normalisasi tipe pembayaran	Normalisasi kota	dist_0	dist_1	dist_2	dist_3	dist_4	cluster
432	0.09090	0.0455	0.8125	0.75	0.5080938	0.7513097	0.1703987	0.9427509	0.7895187	2
685	0.01398	0.0132	0.03125	0.75	0.9766352	0.1157824	0.8513714	0.5761212	0.2527981	1
1826	0.04895	0.0129	0.03125	0.25	0.8420915	0.3970699	1.0522478	0.0923316	0.7518495	3
2586	0.09790	0.0241	0.03125	0	0.883219	0.6503969	1.2213783	0.1997301	1.0044812	3
3533	0.00699	0.0069	0.8125	0.5	0.2534864	0.7518756	0.3854318	0.8101375	0.8963301	0
dst	dst	dst	dst	dst	dst	dst	dst	dst	dst	dst

Tabel 9. Hasil Visualisasi K-Means Clustering

Indikator/Cluster	C1	C2	C3
	<5	0	0
	6-11	0	0
	12-16	99	14
Usia (Tahun)	17-25	5.552	2.126
	26-35	7.109	3.468
	36-45	3.398	1.534
	>45	1.267	553
		476	
Jenis Kelamin	Laki-laki	4.959	2.204
	Perempuan	12.466	5.491
Frekuensi Pembelian	1-5	15.558	6.397
	6-10	849	359
	11-15	205	77
	>15	240	75
Tipe Pembayaran	Gopay	2.652	4.428
	Shopeepay	1.225	2.153
	Virtual Account BCA	5.713	467
	Cash on Delivery	7.727	621
	Lainnya	108	26
	Jakarta Utara	2.305	1.039
	Jakarta Barat	4.680	1.633
	Jakarta Timur	4.675	1.761
Kota	Jakarta Selatan	4.028	2.238
	Jakarta Pusat	1.737	1.024
	Lainnya	0	0
		41	0

Pada Tabel 5 merupakan hasil perhitungan *euclidean distance* untuk k=2, dist_0 berarti jarak perhitungan nilai dari dataset dengan nilai *centroid* 1 pada Tabel 1, sedangkan dist_1 berarti jarak perhitungan nilai dari dataset dengan nilai *centroid* 2 pada Tabel 1.

Pada Tabel 6 merupakan hasil perhitungan *euclidean distance* untuk k=3, dist_0 berarti jarak perhitungan nilai dari dataset dengan nilai *centroid* 1 pada Tabel 2, sedangkan dist_1 berarti jarak perhitungan nilai dari dataset dengan nilai *centroid* 2 pada Tabel 2 dan dist_2 berarti jarak perhitungan nilai dari dataset dengan nilai *centroid* 3 pada Tabel 2.

Pada Tabel 7 merupakan hasil perhitungan *euclidean distance* untuk k=4, dist_0 berarti jarak perhitungan nilai dari dataset dengan nilai *centroid* 1 pada Tabel 3, sedangkan dist_1 berarti jarak perhitungan nilai dari dataset dengan nilai *centroid* 2 pada Tabel 3, dist_2 berarti jarak perhitungan nilai dari dataset dengan nilai *centroid* 3 pada Tabel 3 dan dist_3 berarti jarak

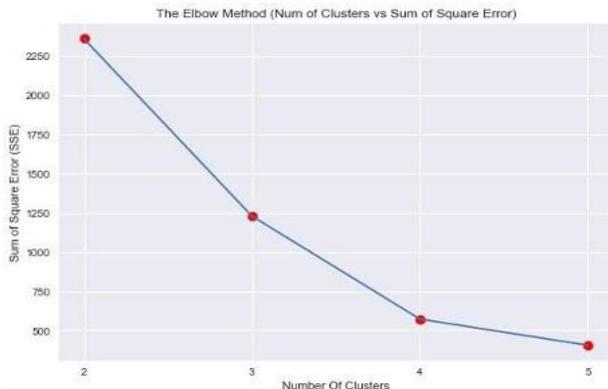
perhitungan nilai dari dataset dengan nilai *centroid* 4 pada Tabel 3.

Pada Tabel 8 merupakan hasil perhitungan *euclidean distance* untuk k=5, dist_0 berarti jarak perhitungan nilai dari dataset dengan nilai *centroid* 1 pada Tabel 4, sedangkan dist_1 berarti jarak perhitungan nilai dari dataset dengan nilai *centroid* 2 pada Tabel 4, dist_2 berarti jarak perhitungan nilai dari dataset dengan nilai *centroid* 3 pada Tabel 4, dist_3 berarti jarak perhitungan nilai dari dataset dengan nilai *centroid* 4 pada Tabel 4 dan dist_4 berarti jarak perhitungan nilai dari dataset dengan nilai *centroid* 5 pada Tabel 4.

4.5 Evaluasi Hasil Clustering

Dalam penelitian ini, metode Elbow digunakan untuk mengevaluasi hasil *clustering*. Dari hasil perhitungan metode Elbow, informasi yang dapat diketahui berdasarkan gambar 6 yaitu *cluster* dengan

nilai yang optimal berada pada *Cluster* 3, sehingga untuk kasus ini jumlah *cluster* yang ideal adalah $K=3$ dan dijadikan *default cluster* untuk menentukan karakteristik dari data - data tersebut.



Gambar 6. Hasil Perhitungan SSE

4.6 Penerapan (*Deployment*)

Setelah melalui keseluruhan proses data mining, dapat disimpulkan bahwa pengelompokan pelanggan terbentuk sesuai jumlah *cluster* terbaik yaitu sebanyak 3 *cluster* atau $k=3$. *Cluster* pertama berjumlah 7.219 pelanggan, *cluster* 2 sebanyak 6.902 pelanggan dan *cluster* 3 sebanyak 5.371 pelanggan. Dari tabel 9 dapat dilihat bahwa C1, C2, C3 memiliki persamaan karakteristik yaitu Pelanggan dalam kluster ini paling banyak berusia antara 26-35 tahun dan berjenis kelamin perempuan dengan rata-rata frekuensi pembelian sebanyak 1-5 kali dalam 1 bulan. C1 dan C2 memiliki persamaan karakteristik untuk metode pembayaran yang paling banyak dipilih adalah *COD* (*Cash on Delivery*) dengan transaksi terbanyak di wilayah Jakarta Barat, sedangkan untuk C2 metode pembayaran yang paling banyak dipilih adalah *Gopay* dengan transaksi terbanyak di wilayah Jakarta Selatan.

5. KESIMPULAN

Segmentasi pelanggan menggunakan metode *clustering K-Means* bekerja dengan baik dalam melakukan *clustering* pelanggan. Terbentuknya *clustering* pelanggan ditentukan dengan tiga kriteria yaitu umur, jenis kelamin, frekuensi pesanan, tipe pembayaran dan kota pembelian barang selama satu bulan. Setelah dilakukannya proses segmentasi data histori transaksi pembelian pelanggan, terbentuk tiga *cluster* yaitu *cluster* 1, *cluster* 2 dan *cluster* 3. Dari hasil segmentasi pelanggan yang telah divisualisasikan dan diinterpretasikan, diharapkan dapat membantu perusahaan dalam memahami kebutuhan pelanggan secara lebih akurat, serta dapat digunakan sebagai data dalam menentukan strategi pemasaran yang tepat.

6. SARAN

Melakukan perbandingan menggunakan metode lain selain *Elbow* untuk mengetahui apakah jumlah *cluster* yang terbentuk $k=3$ sudah terbaik atau belum. Dapat juga menambahkan indikator atau kriteria baru untuk menentukan hasil *cluster* dengan metode *Recency*, *Frequency*, dan

Monetary (RFM), misalnya *Recency* menganalisis pembelian terakhir pelanggan, *Frequency* menganalisis jumlah transaksi pelanggan, dan *Monetary* menganalisis total uang milik pelanggan yang dibayarkan untuk melakukan transaksi pada periode waktu tertentu. Selain itu uji-coba jumlah *cluster* pada penelitian selanjutnya diharapkan dapat lebih banyak agar hasil penelitian yang didapatkan lebih akurat dan berkolaborasi dengan bagian pemasaran untuk mengolah data riwayat transaksi setelah dilakukan proses pemasaran atau promo.

7. DAFTAR PUSTAKA

- Adiana, B. E., Soesanti, I., & Permanasari, A. E. (2018). Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model Dan Teknik Clustering. *Jurnal Terapan Teknologi Informasi*, 2(1), 23–32. <https://doi.org/10.21460/jutei.2018.21.76>
- Atika, P. D., & Priatna, W. (2020). Modul Data Mining. Dalam *Modul Perkuliahan Data Mining Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya*.
- Dewi, D. A. I. C., & Pramita, D. A. K. (2019). Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *Matrix: Jurnal Manajemen Teknologi Dan Informatika*, 9(3), 102–109. <https://doi.org/10.31940/matrix.v9i3.1662>
- Febrianti, A. F., Cabral, A. H., & Anuraga, G. (2018). K-Means Clustering Dengan Metode Elbow Untuk Pengelompokan Kabupaten dan Kota di Jawa Timur Berdasarkan Indikator Kemiskinan. *Seminar Nasional Hasil Riset Dan Pengabdian I*, 863–870.
- Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17–24. <https://doi.org/10.25077/teknosi.v5i1.2019.17-24>
- Hartanti, N. T. (2020). Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 6(2), 82–89. <https://doi.org/10.25077/teknosi.v6i2.2020.82-89>
- Indriyani, I., & Putera, M. I. A. (2020). Web Based Application for Classification Using Naïve Bayes and K-Means Clustering (Case Study: Tic-tac-toe Game). *International Journal of Engineering and Emerging Technology*, 5(1), 8–13. <https://doi.org/10.24843/ijeet.2020.v05.i01.p04>
- Kamila, I., Khairunnisa, U., & Mustakim, M. (2019). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi



- Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 5(1), 119. <https://doi.org/10.24014/rmsi.v5i1.7381>
- Mardalius, M. (2018). Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means. *Jurteks*, 4(2), 123–132. <https://doi.org/10.33330/jurteks.v4i2.36>
- Monalisa, S. (2018). Segmentasi Perilaku Pembelian Pelanggan Berdasarkan Model RFM dengan Metode K-Means. *Query: Jurnal Sistem Informasi*, 2(1), 9–15.
- Nainggolan, R., Perangin-angin, R., Simarmata, E., & Tarigan, F. A. (2019). Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using Elbow Method. *Journal of Physics: Conference Series*.
- Nurdiyansyah, Firman, Arifin, S., & Marisa, F. (2018). Penerapan Clustering Algorithm Untuk Mendukung Promosi Server Pulsa Reload. 3(2), 73–78.
- Putri, E. T., Kusnanto, G., & Thomas, C. Ju. (2019). Penerapan K-Means Clustering Untuk Segmentasi Pelanggan Pada Sistem Customer Relationship Management di PT. Unichem Candi Indonesia. *Konvergensi*, 15(2), 46–57.
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. Dalam *Procedia Computer Science* (Vol. 196, hlm. 526–534).
- Schuh, G., Reinhart, G., Prote, J. P., Sauermann, F., Horsthofer, J., Oppolzer, F., & Knoll, D. (2019). Data mining definitions and applications for the management of production complexity. *Procedia CIRP*, 81, 874–879. <https://doi.org/10.1016/j.procir.2019.03.217>
- Selvida, D. (2019). Analisis Klasifikasi Data Dengan Kombinasi Metode K-Means Dan Rapid Centroid Estimation (RCE).
- Sulistyowati, Ketherin, B. E., Arifiyanti, A. A., & Sodik, A. (2018). Analisa Segmentasi Konsumen Menggunakan Algoritma K-Means Clustering. *Seminar Nasional Sains Dan Teknologi Terapan VI 2018*, 51–58.
- Yunita, F. (2018). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Islam Indragiri). *Jurnal Sistemasi*, 7(September), 238–249.
- Yustina, I., Purwadi, D., & Khuriyati, N. (2021). Healthy Food Consumer Segmentation for Targeting and Positioning New Product Slimming Jelly. *Industria: Jurnal Teknologi Dan Manajemen Agroindustri*, 10(3), 227–238. <https://doi.org/10.21776/ub.industria.2021.010.03.4>