

SEGMENTASI PELANGGAN MENGGUNAKAN ALGORITMA K-MEANS BERDASARKAN MODEL *RECENCY FREQUENCY MONETARY*

Apip Pramudiansyah¹, Hamonangan Munte²

^{1,2}Sistem Informasi, Universitas Budiluhur, Jakarta, Indonesia
¹ezmodes@gamil.com, ²monang.munthe@gmail.com

ABSTRAK

Pelanggan merupakan sumber keuntungan bagi perusahaan seperti halnya pada bisnis di bidang ritel. Pelanggan merupakan aset bagi perusahaan. Saat ini, perusahaan berlomba-lomba dalam memenangkan hati para pelanggan. Karena masing-masing pelanggan tidak hanya memiliki kebutuhan, preferensi harapan dan perilaku yang berbeda-beda namun juga profil pendapatan dan biaya yang berbeda maka sebuah perusahaan perlu memetakan tingkat loyalitas pelanggannya. Namun tidak adanya pengolahan data pelanggan yang telah dimiliki menjadikan perusahaan tidak mengetahui *valuable* dari pelanggan yang ada. Penetapan rencana strategi pemasaran seharusnya dapat dilakukan salah satunya dengan memanfaatkan *valuable customer* yang dimiliki. Pemanfaatan *valuable customer* dilakukan dengan menggunakan *data mining*, penerapan model *Recency Frequency Monetary* pada tahapan data *preparation* dapat membantu untuk melihat *valuable* dari masing-masing pelanggan. Hasil pengujian menggunakan *Elbow Method* didapatkan titik siku terbaik yaitu sebanyak 4 klaster dan setelah dilakukan perhitungan menggunakan algoritma k-means maka dihasilkan sebanyak 38,4% atau berjumlah 557 pelanggan masuk kedalam kelompok kategori *Platinum*, 12,6% atau berjumlah 184 pelanggan kategori *Gold*, 22,7% atau berjumlah 330 pelanggan kategori *Silver*, dan 26% atau berjumlah 378 pelanggan kategori *Bronze*.

Kata Kunci— *Clustering, K-Means, metode elbow, Recency Frequency Monetary, valueable customer.*

ABSTRACT

Customers are a source of profit for the company as well as for businesses in the retail sector. Customers are assets for the company. Today, companies are competing to win the hearts of their customers. Because each customer not only has different needs, expectations and behavior preferences but also different revenue and expense profiles, a company needs to map the level of customer loyalty. However, the absence of processing data on existing customers means that the company does not know that there are valuable customers. One of the ways to establish a marketing strategy plan is to take advantage of valuable customers. The use of valuable customers is done by using data mining, the application of the Recency Frequency Monetary model at the data preparation stage can help to see the valuable from each customer. The results of testing using the Elbow Method obtained the best elbow points, namely 4 clusters and after calculating using the k-means algorithm, 38,4% or 557 customers were included in the Platinum category group, 12,6% or 184 customers in the Gold category. 22,7% or a total of 330 customers in the Silver category, and 26% or totaling 378 customers in the Bronze category.

Keywords— *Clustering, K-Means, elbow method, Recency Frequency Monetary, valueable customer.*

1. PENDAHULUAN

Pelanggan merupakan sumber keuntungan dan kehidupan bagi perusahaan seperti halnya pada bisnis di bidang ritel. Pelanggan merupakan aset bagi perusahaan. Saat ini, perusahaan berlomba-lomba dalam memenangkan hati para pelanggan. Karena masing-masing pelanggan memiliki kebutuhan, preferensi harapan dan perilaku yang berbeda-beda maka sebaiknya perusahaan tidak memperlakukan mereka dengan perlakuan yang sama. Pelanggan tidak hanya memiliki kebutuhan, preferensi dan harapan yang berbeda namun juga profil pendapatan dan biaya yang berbeda dan karenanya harus dikelola dengan cara yang berbeda pula. Perlakuan yang sama terhadap semua pelanggan akan menyebabkan pelanggan yang tidak begitu bernilai bagi perusahaan akan berakhir menjadi penghancur nilai daripada pencipta nilai bagi perusahaan.

Penggunaan metode manual atau tradisional untuk melakukan identifikasi pelanggan dari data membutuhkan kemampuan manusia untuk menganalisis dan menginterpretasikan data. Dengan volume data yang berkembang, baik dari jumlah *record* dan jumlah *field*. Hal ini dapat mengakibatkan perusahaan kehilangan pelanggan potensial dan merugikan perusahaan. Salah satu cara untuk mengelompokkan data penjualan berdasarkan karakteristik pelanggan dapat menggunakan klasterisasi terhadap berbagai jenisnya, klasterisasi pelanggan diperlukan untuk mengelompokkan pelanggan yang memiliki kesamaan karakteristik. Hal ini diperlukan untuk mengetahui perilaku (*behavior*) pelanggan sehingga akan membantu dalam penerapan strategi pemasaran (*marketing*) yang tepat untuk meningkatkan pendapatan perusahaan.

Terdapat beberapa penelitian yang sudah dilakukan diantara nya adalah (Ruly, Purbandini dan Wuryanto, 2017) tentang penerapan clustering k-means pada customer segmentation berbasis *recency frequency monetary* (RFM). Penelitian ini bertujuan untuk penerapan algoritma k-Means untuk membentuk segmentasi pelanggan, sehingga dapat mengetahui tingkat potensial pelanggan. Analisis data dilakukan dengan dua cara yaitu pembobotan RFM untuk menghasilkan bobot RFM dimana recency merupakan transaksi terakhir, frequency merupakan jumlah transaksi yang dilakukan dan monetary yang merupakan total biaya transaksi yang telah dilakukan. Setelah itu barulah dikelompokkan dengan metode k-Means. Dari hasil evaluasi sistem didapatkan bahwa pengelompokan pelanggan pada kategori BZ memiliki nilai presentase 54,3%, kategori MVC sebesar 21,8% dan MGC sebesar 23,9%. Untuk kesimpulan penggunaan k-Means didapatkan bahwa setiap penentuan nilai centroid secara acak yang berbeda, akan menghasilkan pengelompokan dan hasil iterasi yang berbeda.

Kemudian penelitian oleh (Savitri, Bachtiar dan Setiawan, 2018) tentang Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM, permasalahannya adalah belum adanya pemetaan pelanggan yang menyebabkan penerapan perlakuan yang sama pada seluruh pelanggan misalnya dalam hal penentuan promo, pemberian hadiah, maupun

pelayanan. Melalui penelitian ini dilakukan Segmentasi proses untuk mengetahui karakteristik pelanggan dengan kesamaan tertentu, sehingga mempermudah penggalian informasi terkait pelanggan yang *profitable*, dengan analisis RFM (*Recency, Frequency, Monetary*). Salah satu metode pengelompokan yaitu K-Means yang dapat digunakan untuk melakukan segmentasi tersebut. Analisis berdasarkan nilai RFM menunjukkan bahwa peringkat pertama merupakan kelompok pelanggan yang *profitable* karena memiliki nilai RFM yang lebih besar dibanding segmen lainnya.

Berdasarkan latar belakang masalah yang telah dikemukakan maka masalah yang dapat diidentifikasi adalah Tidak diketahuinya klasterisasi pelanggan dan Belum adanya segmentasi pelanggan sehingga *valueable customer* tidak diketahui perusahaan dengan baik.

2. TINJAUAN PUSTAKA

2.1. Pelanggan

Pelanggan atau customer merupakan salah satu aset yang dimiliki oleh perusahaan. Perusahaan harus menjaga pelanggan yang dimiliki untuk dapat terus berjalan dan berkembang. Semakin banyak dan loyal pelanggan yang dimiliki oleh perusahaan, maka semakin baik untuk perusahaan. Namun sifat pelanggan yang selalu pilih-pilih membuat perusahaan kesulitan dalam membedakan pelanggan yang dapat memberikan keuntungan besar bagi perusahaan dan yang tidak atau kurang menguntungkan bagi perusahaan [1].

2.2. Data mining

Data mining merupakan ekstraksi informasi yang tersembunyi dari database yang besar. Data mining adalah teknologi baru yang kuat dengan potensi besar untuk membantu perusahaan agar fokus pada informasi yang paling penting dalam gudang data mereka [2].

2.3. Clustering

Clustering merupakan pengelompokan objek berdasarkan informasi yang diperoleh dari suatu data yang menjelaskan hubungan antar objek satu dengan objek lainnya. Tujuannya untuk mengelompokan objek yang memiliki kesamaan karakteristik dengan objek lainnya dalam satu kelompok dan memiliki perbedaan karakteristik dengan objek kelompok lain yang bersifat tanpa arahan (*unsupervised*). Ada dua jenis data *clustering* yang sering dipergunakan dalam proses pengelompokan data yaitu *hierarchical* (hirarki) data *clustering* dan *non-hierarchical* (non hirarki) data *clustering*. Dalam *cluster* hirarki dimulai dengan membuat K *cluster* dimana setiap *cluster* beranggotakan satu objek dan berakhir dengan satu *cluster* dimana anggotanya K objek, pada setiap tahap prosedurnya, satu *cluster* digabung dengan satu *cluster* lain, lalu dapat dipilih klaster yang diinginkan dengan menentukan *cut-off* pada tingkat tertentu. Dalam *non-hierarchical*

pengelompokan objek dimasukan ke dalam K *cluster*, dapat dilakukan dengan menentukan pusat klaster awal lalu dilakukan realokasi objek berdasarkan kriteria tertentu sampai dicapai pengelompokan yang optimum. [3]

2.4. RFM (Recency, Frequency, Monetary)

Analisa RFM adalah proses menganalisis perilaku pelanggan. Hal ini umumnya digunakan dalam pemasaran *database* dan pemasaran langsung. Analisis RFM merupakan suatu Model perhitungan yang terdiri dari tiga Atribut domain yang memperhatikan transaksi pelanggan berdasarkan transaksi terakhir/keterkinian (*Recency*), jumlah transaksi (*Frequency*), serta nominal dari transaksi (*Monetary*). Tujuan dari RFM adalah untuk meramalkan perilaku Pelanggan di masa depan agar dapat mengarahkan keputusan segmentasi yang lebih baik. Model RFM ini merupakan metode yang sudah lama dan populer untuk mengukur hubungan dengan pelanggan. Dimensi model RFM menurut [4] meliputi :

3.2.1 Recency

Recency adalah mengukur nilai pelanggan dengan melihat prilaku Pelanggan yang berkenaan dengan pembelian yang dilakukannya paling akhir. Informasi terpenting yang tidak boleh dilewatkan berkenaan dengan resensi adalah tanggal pembelian terakhir, karena proses ini bergantung pada tanggal pembelian terakhir, semakin dekat proses pembelian terakhir maka pelanggan tersebut semakin loyal.

3.2.2 Frequency

Frequency adalah mengukur nilai pelanggan dengan melihat prilaku Pelanggan yang berkenaan dengan aktifitas transaksi yang dilakukan oleh Pelanggan dalam satu periode.

3.2.3 Monetary

Monetary adalah mengukur nilai pelanggan dengan melihat perilaku Pelanggan yang berkenaan dengan nilai pembelian yang dilakukan oleh pelanggan dalam periode tertentu. Semakin tinggi nilai pembelian maka semakin tinggi tingkat loyalitas pelanggan

2.5. K-means

K-means adalah salah satu algoritma terkenal dalam *clustering*, awalnya dikenal sebagai metode *Forgy's* dan telah digunakan secara luas di berbagai bidang termasuk *data mining*, analisis statistik data dan aplikasi bisnis lainnya [5].

Dari beberapa teknik klastering yang paling sederhana dan umum dikenal adalah klastering *K-means*. Teknik ini mengelompokkan objek ke dalam K kelompok atau klaster. Algoritma *K-means* membagi kelompok berdasarkan nilai rata-rata dari objek dalam *cluster*. Adapun tujuan pengelompokan data ini adalah untuk meminimalkan variasi didalam suatu kelompok dan memaksimalkan variasi antar. Konsep dasar *k means* menurut [6] adalah

pencarian pusat kelompok, langkah-langkah pengelompokan *k-means* adalah sebagai berikut:

2.5.1 Menentukan pusat kelompok secara acak

2.5.2 Kemudian dihitung jarak antara setiap objek dengan setiap pusat kelompok. Untuk melakukan penghitungan jarak objek ke- *i* pada pusat kelompok ke-*k* dapat digunakan rumus jarak euclidean, dapat dilihat pada persamaan 1.

$$d(i, k) = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \quad (1)$$

Keterangan:

dik: jarak objek ke-*i* pada pusat kelompok ke-*k*

x_{ij}: nilai objek ke-*i* pada variabel *j*

c_{kj}: pusat kelompok ke-*k* pada variabel *j*

m: jumlah variabel yang digunakan

i menyatakan objek, *k* menyatakan pusat kelompok

J menyatakan keanggotaan kelompok

2.5.3 Suatu objek akan menjadi anggota dari kelompok ke-*J* apabila jarak objek tersebut ke pusat kelompok ke-*J* bernilai paling kecil jika dibandingkan dengan jarak ke pusat kelompok lainnya.

2.5.4 Selanjutnya, kelompok objek-objek yang menjadi anggota pada setiap kelompok.

2.5.5 Menentukan nilai pusat kelompok yang baru dapat dihitung dengan cara mencari nilai rata-rata dari objek yang menjadi anggota pada kelompok tersebut, dapat dilihat pada persamaan 2.

$$c_{kj} = \frac{\sum_{h=1}^p y_{hj}}{p}; y_{hj} = x_{ij} \in \text{cluster ke- } k \quad (2)$$

keterangan :

c_{kj} : pusat kelompok ke-*k* pada variabel *j*

y_{hj} : nilai objek ke-*h* pada variabel *j*

p : jumlah kelompok terbentuk

2.5.6 Ulangi langkah 2 sampai 5 hingga sudah tidak ada lagi objek yang berpindah ke kelompok yang lain.

2.6. Davies Bouldin Index (DBI)

Dalam klasifikasi, evaluasi *system* sudah menjadi bagian penting dalam proses pembangunan model klasifikasi, ukuran dan metode untuk mengevaluasi, seperti akurasi, *precision* dan *recall*, validasi silang, dan sebagainya. Model dibangun menggunakan set data latih dengan sejumlah parameter yang diminta oleh metodenya. Selanjutnya set data latih diberikan untuk menguji kinerja *system*, ada label kelas asli yang dibandingkan dengan label kelas yang didapatkan dari proses prediksi. Terakhir,

dilakukan evaluasi terhadap sistem klasifikator yang dibuat [7].

metode *Davies Bouldin Index* (DBI). DBI diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979 adalah metric untuk mengevaluasi hasil algoritma *clustering*. Pendekatan pengukuran ini untuk memaksimalkan jarak *inter-cluster* diantara *Cluster* Ci dan Cj dan pada waktu yang sama mencoba untuk meminimalkan jarak antara titik dalam sebuah *cluster*. Jarak *intra-cluster* S_c (Q_k) dalam cluster Q_k dapat dilihat pada Persamaan 3.

$$S_c(Q_k) = \frac{\sum_i ||x_i - c_k||}{N_k} \quad (3)$$

Dengan N_k adalah banyak titik yang termasuk dalam cluster Q_k dan c_k adalah centroid dari Cluster Q_k . Rumus jarak *Inter-cluster* dapat dilihat pada Persamaan 4.

$$d_k = ||c_k - c_l|| \quad (4)$$

Dengan c_k dan c_l ialah *centroid Cluster k* dan *cluster l*. Dilain pihak, indeks Davies Bouldin dapat dilihat pada Persamaan 5.

$$DB(nc) = \frac{1}{n} \sum_{k=1}^{nc} \max \left\{ \frac{S_c(Q_k) + S_c(Q_l)}{d_{kl}(Q_k, Q_l)} \right\} \quad (5)$$

Dengan nc ialah banyak cluster.

2.7. Normalisasi Min-Max

Metode Min-Max merupakan metode normalisasi sederhana dengan melakukan transformasi terhadap data asli. Min-Max akan menyesuaikan batas yang ditentukan dengan menghubungkan pada data asli. Teknik normalisasi ini melakukan transformasi sebuah atribut numerik dalam range atau skala yang lebih kecil seperti 0.0 sampai 1.0 [8]. Angka tersebut menunjukkan bahwa batas terendah 0.0 sedangkan batas tertinggi 1.0. Normalisasi Min-Max dapat dihitung dengan persamaan 6.

$$x' = \frac{x - \text{mina}}{\text{maxa} - \text{mina}} (\text{newmax} - \text{newmin}) + \text{newmin} \quad (6)$$

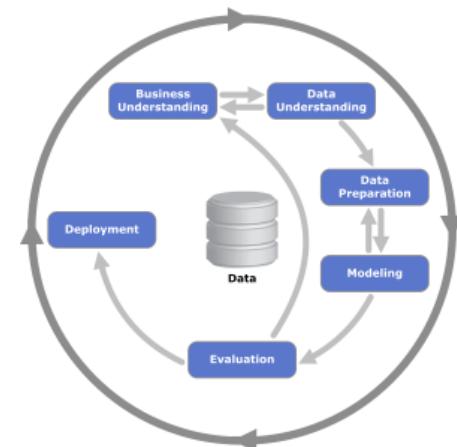
Dimana:

- x' = nilai yang telah dinormalisasi
- x = nilai mentah yang akan dinormalisasikan
- mina = nilai terendah pada setiap variabel
- maxa = nilai tertinggi pada setiap variabel
- newmax = rentang maksimal X yang bernilai 1
- newmin = rentang manimal X yang bernilai 0

Kelebihan dari Min-Max yaitu nilai perbandingan antar data sebelum normalisasi seimbang dengan data sesudah normalisasi dan kelebihan lainnya tidak ada data bias yang dihasilkan. Kelemahan dari Min-Max, jika ada data baru, akan memungkinkan terjebak pada "out of bound" error.

2.8. CRISP-DM

Cross Industry Standard Process Data Mining (CRISP-DM) yang telah dikembangkan tahun 1996 oleh analis dari beberapa industri menyediakan proses data mining yang digunakan sebagai strategi dalam pemecahan masalah secara umum untuk penelitian [9]. Langkah-langkah Crisp-DM bisa dilihat pada gambar 1.



Gambar 1 Langkah-langkah CRISP-DM
 Tahapan – tahapan dalam model CRISP-DM terdiri atas 6 tahapan [10], antara lain:

2.8.1 Pemahaman Bisnis (Business Understanding)

Pada tahapan *business understanding*, dilakukan pemahaman mengenai tujuan dan kebutuhan dari sudut pandang bisnis, yang kemudian diterjemahkan pendefinisian masalah ke dalam *data mining*. Kemudian akan digunakan dalam menentukan rencana dan strategi untuk mencapai tujuan bisnis.

2.8.2 Pemahaman Data (Data Understanding)

Data understanding merupakan proses pengumpulan data yang selanjutnya dilakukan pemahaman lebih mendalam mengenai data tersebut serta melakukan identifikasi terhadap kualitas data, untuk mendeteksi bagian data yang menarik agar dapat dijadikan hipotesa dalam informasi yang tersembunyi.

2.8.3 Data Preparation

Yang dilakukan pada tahap *data preparation* yaitu menyiapkan data mentah yang akan digunakan dalam penelitian menjadi dataset yang siap digunakan. Terdiri atas beberapa langkah dimulai dengan *data selection*, *data preprocessing* serta *data transformation*. Langkah tersebut dapat dilakukan berulang sampai mendapatkan bentuk data yang diinginkan untuk dapat digunakan dalam tahapan *modelling*.

2.8.4 Modelling

Modelling merupakan tahap pemilihan dan penerapan teknik algoritma yang sudah ditentukan sebelumnya serta biasanya menggunakan bantuan *framework* tertentu dalam melakukan penghitungan data. Untuk kemudian digunakan atau diterapkan pada tahap *deployment* menggunakan bahasa pemrograman yang juga telah ditentukan sebelumnya.

2.8.5 Evaluation

Evaluation atau kegiatan evaluasi data yang diolah pada tahap sebelumnya untuk dapat mengetahui bahwa data yang diperoleh merupakan data yang baik kualitasnya agar hasil yang diperoleh sesuai dengan tujuan perusahaan.

2.8.6 Deployment

Tahapan *deployment* dalam model CRISP-DM merupakan tahapan akhir yang berupa pengimplementasian proses data mining dalam perusahaan secara berulang dan pembuatan laporan atas hasil yang telah diperoleh pada tahapan sebelumnya.

2.9. Rapid miner

RapidMiner merupakan sebuah *framework* yang digunakan sebagai alat bantu dalam melakukan penghitungan data yang telah melalui proses data *preparation*. Metode atau algoritma *clustering data mining* yang sudah ditetapkan akan diterapkan dalam penghitungan menggunakan *framework* RapidMiner ini. Yang kemudian akan digunakan sebagai dasar dalam pembuatan prototipe maupun aplikasi *clustering data mining* [10].

2.10. WEKA

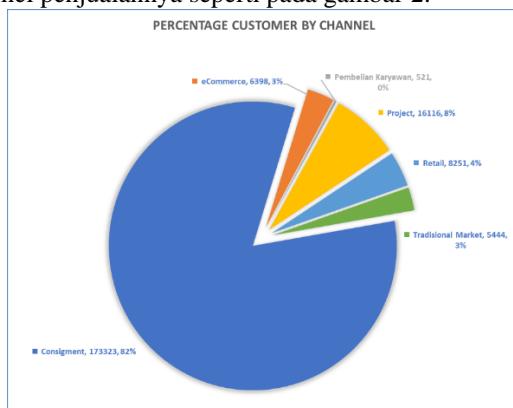
WEKA merupakan salah satu *software* yang digunakan dalam *data mining* untuk mendukung proses *data mining* mulai *preprocessing data*, pengelompokan, klasifikasi, regresi, visualisasi, dan seleksi fitur.

Kelebihan yang dimiliki WEKA adalah *user friendly*, mudah digunakan dan proses klasifikasi sebuah *database* menjadi mudah. Kekurangan WEKA dapat terjadi *overloading* data jika *database* yang dimasukkan melebihi kuota yang ditentukan di dalam WEKA. [11].

3. METODE PENELITIAN

3.1 Pemahaman Bisnis (Business Understanding)

Perusahaan swasta yang bergerak dibidang industri jual beli yang berkomitmen untuk meningkatkan pengalaman hidup di rumah orang Indonesia dan berkomitmen untuk menawarkan solusi serta layanan terbaik kepada pelanggan perhotelan dan ritel di seluruh negeri yang menjual produk atau jasa layanan kepada konsumen secara eceran atau jumlah kecil untuk memenuhi kebutuhan pribadi dan bukan untuk dijual kembali, dari data yang ada penjualan didominasi dengan *channel consignment* sebesar 82%, sedangkan penjualan secara *e-commerce*, pembelian karyawan, *project*, *retail* dan tradisional semua tidak lebih dari 10% dari total penjualan selama tahun 2019, berikut adalah gambaran berdasarkan *channel* penjualannya seperti pada gambar 2.



Gambar 2 Percentage Customer By Channel

3.2 Pemahaman Data (Data Understanding)

Data yang akan diolah menggunakan *data mining* yakni data penjualan dalam 12 bulan dimulai sejak bulan januari 2019 sampai dengan bulan Desember 2019. Pada

data penjualan produk tersebut terdapat atribut *Date*, *Order No*, *city*, *channel*, *Customer ID*, *Customer Name*, *Brand*, *Item ID*, *Item Description*, *QTY*, *Unit ID*, *Location ID*, *Location Name*, *Unit Price*, *Net Price*.

3.2.1 Data Preparation

Tiga hal mendasar yang dilakukan pada tahapan *data preparation*, antara lain:

3.2.2 Data Selection

Data sekunder yang digunakan akan melalui pemilihan data berdasarkan kolom yang ada. Kolom yang tidak relevan dengan metode yang akan digunakan akan dihilangkan.

3.2.3 Data Preprocessing

Setelah melalui tahap *data selection*, data sekunder akan melalui proses *data preprocessing* agar data yang digunakan bersih dari *noisy data* dan *missing values*. Data akan di *sort* berdasarkan kolom *amount* untuk menghilangkan data yang memiliki nilai nol pada kolom *amount*.

3.2.4 Data Transformation

Setelah data yang akan digunakan bersih dari *noisy data* maupun *missing values* maka dilakukan perubahan nilai data dengan mengelompokkan dan memberi bobot pada masing-masing kolom *recency*, *frequency* dan *monetary*. Sehingga data yang akan diolah menggunakan algoritma *data mining* pada proses *modelling* memiliki nilai yang seragam atau sama. Berikut adalah hasil dari pemilihan atribut yang telah dilakukan, dengan menyesuaikan metode RFM yang berkaitan dengan atribut, *recency*, *frequency*, *monetary*. Pemilihan atribut tersebut seperti pada tabel 1.

Tabel 1 Pemilihan atribut

Atribut	Keterangan
KodePelanggan	Nama pelanggan yang dirubah menjadi sebuah kode untuk mengamankan data perusahaan
JarakPembelianAkhir	Merupakan <i>recency</i> , yakni tanggal terakhir pembelian atau transaksi terakhir yang dilakukan oleh pelanggan
FrekuensiBeli	Merupakan <i>frequency</i> yang menunjukkan jumlah transaksi selama periode yang telah ditentukan
TotalBeli	Merupakan <i>monetary</i> , jumlah uang sesuai periode yang ditentukan

sumber: [9]

3.2.5 Modelling

Setelah data yang akan digunakan di proses pada data *preparation*, maka selanjutnya data tersebut diolah dengan cara diberikan bobot sesuai dengan domain nilai dengan tujuan mempermudah dalam proses pengolahan data sebelum dilakukan pemodelan menggunakan algoritma

clustering data mining berbasis partisi, sehingga dapat menghasilkan data derajat keanggotan dari masing-masing variabel untuk menentukan label pelanggan yaitu kelas pelanggan. Perhitungan data akan menggunakan bantuan *framework data mining*.

Berdasarkan data penjualan, akan dilakukan pengolahan data menggunakan algoritma *clustering* data mining berbasis partisi, disesuaikan dengan atribut yang akan digunakan dalam segmentasi pelanggan perusahaan. Penerapan atribut yang dilakukan pada dataset akan dimanfaatkan untuk menentukan jenis pelanggan yang terbagi menjadi empat, yakni: *Platinum, Gold, Silver, Bronze*.

3.2.6 Evaluation

Jika hasil dari evaluasi menunjukkan belum mampu membaca *cluster* dengan tepat, maka akan dilakukan pemodelan ulang atau proses *modeling* kembali untuk memperbaiki struktur yang digunakan. Tahapan evaluasi dianggap sudah selesai apabila *business understanding* sudah terjawab dengan baik.

3.2.7 Deployment

Penerapan aplikasi data mining tersebut menggunakan data penjualan dengan periode bulan Januari sampai bulan Desember 2019 yang telah diolah pada tahap sebelumnya.

3.3 Sampling/Metode Pemilihan Sample

Mengenai *sample* yang digunakan maka perlu dilakukan penyesuaian terhadap jenis penelitian yang akan dilakukan, dimana tujuan penelitian ini adalah menghasilkan sistem informasi yang dapat digunakan dalam mendukung proses segmentasi pelanggan. Sehingga sample yang digunakan merupakan data transaksi dengan periode selama 12 bulan yaitu dimulai sejak bulan januari 2019 sampai dengan bulan desember 2019.

3.4.1 Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data transaksi penjualan dengan periode satu tahun terakhir. Proses ini bertujuan untuk mengumpulkan data sebagai informasi yang akan diolah menggunakan metode yang telah ditentukan sebelumnya. Metode dalam pengumpulan data yang akan digunakan sebagai berikut:

3.4.2 Studi Pustaka

Peneliti melakukan pengumpulan data dengan mencari bahan informasi serta karya ilmiah yang berkaitan dengan penelitian yang akan dilakukan, yakni segmentasi pelanggan menggunakan RFM dan penerapan algoritma *clustering* data mining berbasis partisi, yang akan dipelajari lebih lanjut oleh peneliti.

3.4.3 Wawancara

Wawancara dilakukan penulis kepada seorang *analys data*, Wawancara ini dimaksudkan untuk mendapatkan gambaran mengenai obyek penelitian dan pengembangan aplikasi yang akan dibuat agar dapat memahami proses bisnis yang sebenarnya terjadi.

3.4.4 Instrumentasi

Instrumentasi yang dilakukan dalam penelitian ini antara lain:

3.3.4.1 *Instrumentasi Studi Pustaka*

Instrumen untuk dapat melakukan studi pustaka adalah peneliti mempelajari literatur penelitian terkait mengenai segmentasi pelanggan dengan menerapkan model RFM serta algoritma *clustering data mining* berbasis partisi.

3.3.4.2 *Instrumentasi Wawancara*

Instrumen wawancara yang dilakukan untuk mengumpulkan data menggunakan cara mewawancarai top management yang berhubungan langsung dengan data dimaksudkan untuk mendapatkan gambaran mengenai obyek penelitian dan pengembangan aplikasi yang akan dibuat agar dapat memahami proses bisnis yang sebenarnya terjadi.

3.4 Teknik Analisis, Rancangan, dan Pengujian Data/Sistem/Prototipe Model, Rencana Strategi

Teknis Analisis data menggunakan data kuantitatif, data transaksi penjualan produk yang akan diolah menggunakan model CRISP-DM serta pembobotan dengan RFM sebelum menerapkan algoritma *clustering data mining* berbasis partisi untuk *clustering*.

3.4.1 Pemahaman Bisnis (Business Understanding)

Jika melihat dari data yang didapatkan dan hasil dari observasi, selama ini tidak memberikan perlakuan khusus kepada pelanggan yang dimiliki sehingga pelanggan yang sering membeli dengan pelanggan yang jarang membeli diperlakukan sama. Hal ini disebabkan karena tidak dilakukan pemanfaatan data yang optimal terhadap data yang dimiliki untuk melihat nilai pelanggan yang ada. Oleh karena hal tersebut penelitian ini dimaksudkan untuk dapat melihat nilai pelanggan yang ada dengan menerapkan segmentasi pelanggan yang akan diolah menggunakan RFM, algoritma *clustering data mining* berbasis partisi.

3.4.2 Pemahaman Data (Data Understanding)

Data yang didapatkan untuk digunakan dalam penelitian ini diperoleh dalam periode 12 bulan yakni sejak bulan januari 2019 sampai dengan bulan desember 2019. Data tersebut terdiri dari 16 atribut, akan tetapi akan dilakukan penyesuaian dengan cara seleksi data yang diperlukan dalam proses segmentasi pelanggan seperti pada tabel 2.

Tabel 1 Atribut-atribut yang akan digunakan

No	Atribut	Judul	Keterangan
1	Tanggal transaksi	Recency	Tanggal akhir transaksi terakhir
2	Frekuensi Transaksi	Frequency	Pengulangan pembelian

3	Total transaksi	Monetary	Jumlah uang yang dibelanjakan
---	-----------------	----------	-------------------------------

3.4.3 Data Preparation

Pada *data preparation* akan diterapkan metode RFM (*Recency, Frequency dan Monetary*) sebagai dasar dalam pemilihan atribut yang akan digunakan. Hal yang dilakukan dalam *data preparation* meliputi seleksi atribut atribut data yang ada untuk dijadikan atribut penentu atau parameter dalam segmentasi pelanggan. Kemudian dilakukan pemilihan atau seleksi terhadap atribut data yang akan digunakan. Pembersihan data atau menghilangkan *noisy data* dilakukan untuk menghilangkan atribut-atribut yang tidak digunakan serta menghapus atau menghilangkan data double dan data yang kosong (*null*) atau tidak lengkap. Setelah data bersih dari *noisy data* dan *missing values* maka dilakukan transformasi data menggunakan bobot yang telah ditentukan.

3.4.4 Modelling

Proses *modelling*, menggunakan *data training* yang telah melalui tahap *data preparation* untuk dilakukan pengujian akurasi data kemuadian penerapan atau pemodelan menggunakan algoritma *clustering data mining* berbasis partisi untuk dilakukan *clustering* yang selanjutnya menghasilkan sejumlah aturan. Dalam proses modeling akan dilihat validitas dari setiap algoritma dengan membandingkan nilai *Davies-Bouldin Index* (DBI). Pada tahapan ini juga dilakukan eksperimen terhadap atribut-atribut data berupa modifikasi ataupun menghapus atribut yang tidak memiliki pengaruh yang signifikan, hal ini dilakukan untuk meningkatkan nilai akurasi. Metode dengan nilai akurasi terbaik akan diimplementasikan pada program yang akan dirancang, pengujian nilai DBI masing-masing algoritma akan dilakukan dengan bantuan *tools RapidMiner*.

3.4.5 Evaluation

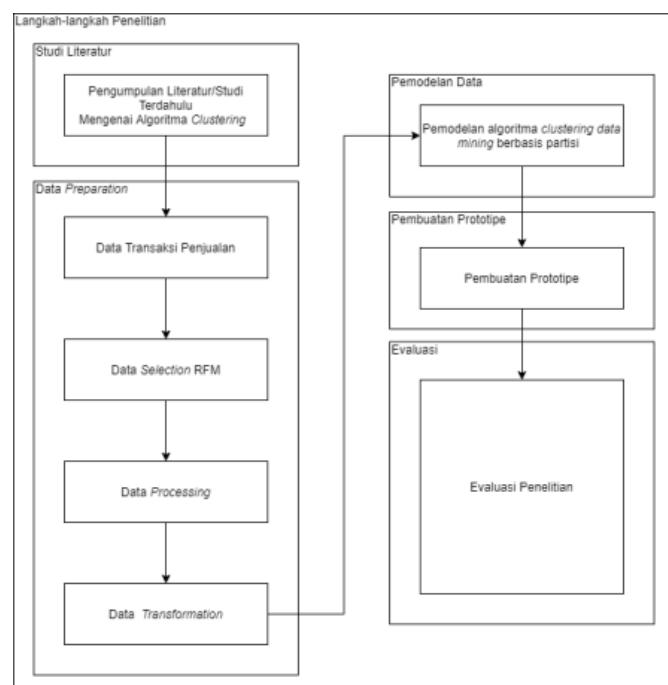
Dalam tahapan ini akan dilakukan analisa terhadap pemodelan yang sudah dilakukan, jika hasil dari evaluasi menunjukkan belum mampu membaca klaster dengan tepat, maka akan dilakukan pemodelan ulang atau proses *modeling* untuk memperbaiki struktur yang digunakan. Tahapan evaluasi dianggap sudah selesai apabila *business understanding* sudah terjawab dengan baik. Pada tahap ini digunakan dataset *evaluation* dalam satu periode yaitu data penjualan pada tahun 2019, dikarenakan data transaksi pada tahun berikutnya yaitu tahun 2020 mengalami ketidakwajaran yang disebabkan pandemi COVID19 (*Corona Virus Disease 2019*), Oleh sebab itu, dataset yang digunakan pada tahap *evaluation* ini data transaksi pada tahun 2019.

3.4.6 Deployment

Pada tahapan *deployment*, yang dilakukan pada penelitian ini adalah menerapkan model atau algoritma yang keunggulannya lebih sesuai dengan kebutuhan perusahaan menggunakan *prototype* yang dibuat menggunakan bahasa pemrograman yang akan digunakan secara berulang dengan jumlah klaster serta atribut yang sudah ditentukan.

3.4.7 Langkah-langkah penelitian

Langkah – langkah yang direncanakan pada penelitian ini digambarkan pada gambar.



Gambar 3 langkah-langkah penelitian

Langkah-langkah penelitian dimulai dengan mengumpulkan studi literatur terdahulu mengenai *data mining clustering* atau segmentasi pelanggan. Data yang akan digunakan dalam penelitian disiapkan, kemudian dilakukan proses *data preparation* dengan menerapkan metode RFM yang mencakup tahapan *data selection*, *data preprocessing* dan *data transformation*. Setelah data melalui tahap *data preparation* maka selanjutnya data melalui proses *modelling* menggunakan algoritma *data mining clustering* berbasis partisi. Setelah diperoleh data hasil dari proses *modelling* maka selanjutnya dibuat *prototype* sistem dengan menerapkan metode RFM dan algoritma *data mining* berbasis partisi. Setelah *prototype* selesai dibuat, maka dilakukan evaluasi prototipe dengan menggunakan *data evaluation*.

4. HASIL PENELITIAN

pengolahan data transaksi penjualan selama 12 bulan pada tahun 2019, dikarenakan pada tahun 2020 terdapat

ketidakwajaran data disebabkan dengan adanya pandemi COVID19 yang berdampak pada penjualan pada perusahaan.

4.1 Business Understanding

Proses pengumpulan data dilakukan dengan wawancara secara langsung untuk menggali proses bisnis dari perusahaan terkait pemasaran dan kendala dalam proses pemasaran yang selama ini dihadapi. Proses berikutnya yaitu pengambilan data yang dilakukan oleh seorang data analis, Data tersebut merupakan data transaksi penjualan semua cabang periode bulan januari sampai bulan desember 2019. Data diperoleh berbentuk rekapan excel. Total atribut dari dokumen ini adalah sebanyak 13 atribut dengan jumlah sebanyak 212571 transaksi.

4.2 Data Understanding

Jenis data transaksi yang dimiliki adalah data transaksi dengan satu atau banyak *record* dalam satu nomor kontrak, jenis produk yang dijual memiliki banyak variasi dan Untuk mengetahui *valuable* pelanggan maka diterapkan model *recency, frequency dan monetary*, sehingga data yang ada dilakukan seleksi agar sesuai dengan kebutuhan model yang digunakan. Pada gambar 4 menunjukan cuplikan data history penjualan yang akan diolah pada bulan Januari 2019.

Gambar 4 contoh data penjualan bulan Januari 2019

Pada gambar 4 dapat dilihat Data sekundernya memiliki 13 kolom, data tersebut memiliki atribut-atribut seperti pada tabel 3 berikut:

Tabel 3 Atribut dan Nilai

No	Atribut	Pengertian	kategori
1	Date	Tanggal transaksi	
2	No	No Invoice Yang Diatur	1=Jakarta
		Dalam Beberapa Tergantung Perusahaan	2=Bandung
		Kode Cabang	3=Surabaya
		Yang	4=Bali
		Terdapat Transaksi	SICS=Consigment
		Contoh:	SICM=eCommerce
		1SICM19010032	SIPK=Pembelian
		(1)	Karyawan
		Code Untuk Cabang	SISP=Project

Jakarta, SICM untuk
no *Invoice* yang
ber type transaksi
eCommerce, 19
adalah tahun 2019, 01
adalah bulan januari
dan 0032 untuk no
urut *Invoice* 32.

SAPR=Project
SIOT=Project
SIPR=Project
SACS=Retail
SIRT=Retail
SITM=Traditional
Market

3	<i>Customer name</i>	Nama Customer
4	Brand	Nama Brand
5	ItemID	Kode item yang terdaftar di sistem
6	<i>Item Description</i>	nama item yang terdaftar di sistem
7	<i>QTY</i>	Jumlah quantitas yang terjual
8	<i>UnitID</i>	Keterangan jumlah item yang terjual Pcs, Box, Unit, Set, Dan PRS
9	<i>OrderNo</i>	No pesanan yang diatur dalam beberapa kode tergantung lokasi cabang perusahaan yang ada transaksi, contoh: 1 = Jakarta 2 = Bandung 3 = Surabaya 4 = Bali
10	<i>Location ID</i>	Keterangan nama gudang pada sistem untuk mengatur stock barang
11	<i>Location Name</i>	Nama customer untuk pengiriman barang
12	<i>UnitPrice</i>	Harga barang satuan
13	<i>NetPrice</i>	Total harga pembelian

Berdasarkan tabel 3 selanjutnya untuk mempermudah pengelompokan data maka dilakukan expansi terhadap atribut No dan CustomerName, untuk atribut No dibagi menjadi dua atribut yaitu atribut *City* dan *Channel*, kemudian untuk atribut CustomerName dibagi menjadi atribut *CustID* dan *CustName*. Atribut No Yang Diatur Dalam Beberapa Kode Tergantung Cabang Perusahaan Yang Terdapat Transaksi Contoh: 1SICM19010032 (1 Code Untuk Cabang Jakarta, SICM untuk no *Invoice* yang bertipe transaksi *eCommerce*, 19 adalah tahun 2019, 01 adalah bulan januari dan 0032 untuk no urut *Invoice* 32. Perubahan atribut tersebut dapat dilihat gambar 5.

No	City	Channel	Customer Name	CustID	CustName
25APR19070006	Bandung	Project	SANBE - PT. SANBE KARYAPERSADA	SANBE	PT. SANBE KARYAPERSADA
25APR19070006	Bandung	Project	SANBE - PT. SANBE KARYAPERSADA	SANBE	PT. SANBE KARYAPERSADA
25APR19070006	Bandung	Project	SANBE - PT. SANBE KARYAPERSADA	SANBE	PT. SANBE KARYAPERSADA
25APR19070006	Bandung	Project	SANBE - PT. SANBE KARYAPERSADA	SANBE	PT. SANBE KARYAPERSADA
25APR19070006	Bandung	Project	SANBE - PT. SANBE KARYAPERSADA	SANBE	PT. SANBE KARYAPERSADA
25IC519010035	Bandung	Consignment	SUBMART - Subur Mart Bandung	SUBMART	Subur Mart Bandung
25IC519010035	Bandung	Consignment	SUBMART - Subur Mart Bandung	SUBMART	Subur Mart Bandung
25IC519010035	Bandung	Consignment	SUBMART - Subur Mart Bandung	SUBMART	Subur Mart Bandung
25IC519010035	Bandung	Consignment	SUBMART - Subur Mart Bandung	SUBMART	Subur Mart Bandung
25IC519010035	Bandung	Consignment	SUBMART - Subur Mart Bandung	SUBMART	Subur Mart Bandung
25IC519010035	Bandung	Consignment	SUBMART - Subur Mart Bandung	SUBMART	Subur Mart Bandung
25IC519010035	Bandung	Consignment	SUBMART - Subur Mart Bandung	SUBMART	Subur Mart Bandung
25IC519010036	Bandung	Consignment	BTIP2 - BOUTIQUE PERENNIALS BANC BTIP2	BOUTIQUE PERENNIALS BANDUNG	BOUTIQUE PERENNIALS BANDUNG
25IC519010036	Bandung	Consignment	BTIP2 - BOUTIQUE PERENNIALS BANC BTIP2	BOUTIQUE PERENNIALS BANDUNG	BOUTIQUE PERENNIALS BANDUNG
25IC519010036	Bandung	Consignment	BTIP2 - BOUTIQUE PERENNIALS BANC BTIP2	BOUTIQUE PERENNIALS BANDUNG	BOUTIQUE PERENNIALS BANDUNG
25IC519010036	Bandung	Consignment	BTIP2 - BOUTIQUE PERENNIALS BANC BTIP2	BOUTIQUE PERENNIALS BANDUNG	BOUTIQUE PERENNIALS BANDUNG
25IC519010036	Bandung	Consignment	BTIP2 - BOUTIQUE PERENNIALS BANC BTIP2	BOUTIQUE PERENNIALS BANDUNG	BOUTIQUE PERENNIALS BANDUNG
25IC519010036	Bandung	Consignment	BTIP2 - BOUTIQUE PERENNIALS BANC BTIP2	BOUTIQUE PERENNIALS BANDUNG	BOUTIQUE PERENNIALS BANDUNG
25IC519010036	Bandung	Consignment	BTIP2 - BOUTIQUE PERENNIALS BANC BTIP2	BOUTIQUE PERENNIALS BANDUNG	BOUTIQUE PERENNIALS BANDUNG

Gambar 5 perubahan atribut no dan Customer name

Setelah dilakukan penambahan kolom Data sekunder memiliki 17 kolom kemudian dilakukan pemilihan kolom yang sesuai dengan kebutuhan sebanyak 4 kolom yaitu, *CustID*, *Date*, *CustomerName*, *NetPrice*. empat buah kolom yang telah dipilih terdiri atas satu kolom sebagai label dan tiga kolom sebagai atribut *predictor* yang akan menghasilkan satu atribut hasil.

4.3 Data preparation

Data sekunder terdiri dari 212.571 record data transaksi dengan 1.449 pelanggan yang terjadi selama bulan Januari - Desember 2019. Data sekunder diolah pada proses data preparation dengan menerapkan model *recency*, *frequency* dan *monetary*. Sehingga dilakukan pemilihan 4 buah kolom yang terdiri dari satu kolom sebagai label dan tiga kolom sebagai atribut prediktor yang akan menghasilkan satu atribut hasil. Pada tabel 4 menunjukkan Proses *Data Selection*.

Tabel 4 proses data selection

CustID	Date	CustName	NetPrice
SANBE	29-Jul-19	PT. SANBE	3,900,00
		KARYAPE	0.00
		RSADA	
SANBE	29-Jul-19	PT. SANBE	3,120,000.00
		KARYAPE	
		RSADA	
SANBE	29-Jul-19	PT. SANBE	2,340,000.00
		KARYAPE	
		RSADA	
SANBE	29-Jul-19	PT. SANBE	2,313,000.00
		KARYAPE	
		RSADA	
SANBE	29-Jul-19	PT. SANBE	4,626,000.00
		KARYAPE	
		RSADA	
SUBMA	31-Jan-	Subur Mart	198,000.00
RT	19	Bandung	
SUBMA	31-Jan-	Subur Mart	63,000.00
RT	19	Bandung	

Setelah dilakukan data selection maka dilakukan pembersihan data dengan melakukan sort data pada kolom *NetPrice* untuk menghilangkan missing value yang disebabkan gagalnya transaksi sehingga tidak terjadi pembayaran, maka transaksi tersebut dihilangkan dari

record. Serta dilakukan penghitungan *frequency* transaksi yang dilakukan oleh setiap pelanggan dengan melihat berapa banyak tanggal transaksi yang berbeda pada setiap pelanggan dalam satu periode data yang telah ditentukan pada kolom *Date*. Pada kolom *NetPrice* juga dilakukan penjumlahan nilai belanja yang terjadi pada setiap pelanggan dalam periode yang ditentukan untuk mengetahui nilai *monetary* setiap pelanggan. Untuk mengetahui nilai *recency* maka dilakukan pencarian tanggal terakhir transaksi setiap pelanggan pada kolom *Date* dalam satu periode yang ditentukan. Proses Data Preprocessing pada Data bisa dilihat pada tabel 5.

Tabel 5 Proses Data Preprocessing pada Data

Customer ID	Recency	Frequency	Monetary
18SVL	28-11-2019	13	20,366,000
AABH	30-12-2019	60	55,415,600
AAIK	18-06-2019	11	9,799,000
AAN	11-11-2019	2	1,079,100
ABRC	31-01-2019	3	1,752,000
ABSC	18-09-2019	22	16,033,000
ABSKT	03-06-2019	5	27,768,000
ABURK	06-11-2019	8	1,994,700
ABURM	21-10-2019	16	4,142,000
ACE	08-07-2019	10	4,307,045
ADC	19-12-2019	19	22,420,500
ADELIACF	30-11-2019	23	40,420,960
ADHL	11-02-2019	1	10,500,000

Data yang telah melalui proses *data preprocessing* kemudian dilakukan data *transformation*. Pada kolom *CustomerID* dilakukan perubahan data nama pelanggan menjadi *code* untuk menjaga kerahasiaan data. Pada kolom *recency* dilakukan penghitungan yaitu mengurangi tanggal terakhir periode dengan tanggal terakhir transaksi yang sudah tercantum pada kolom *recency* sehingga data menjadi seperti pada tabel 6.

Tabel 6 Proses Data Transformation pada Data Training

Code	Recency	Frequency	Monetary
CC1	33	13	20366000
CC2	1	60	55415600
CC3	196	11	9799000
CC4	50	2	1079100
CC5	334	3	1752000
CC6	104	22	16033000
CC7	211	5	27768000
CC8	55	8	1994700
CC9	71	16	4142000
CC10	176	10	4307045

Kemudian dilakukan pembobotan pada data *transformation* menggunakan metode Min Max agar selisih nilai antar pelanggan tidak terlalu jauh. Proses pembobotan ini dilakukan langsung di aplikasi. Metode normalisasi Min Max mentransformasikan seluruh nilai atribut menjadi rentang nilai antara 1 hingga 10. Normalisasi ini akan dilakukan pada setiap variabel *Recency*, *Frequency* dan *Monetary* dengan menggunakan nilai minimal dan maksimal setiap variabel. Nilai maksimal dan minimal dari setiap atribut sehingga menghasilkan data seperti pada tabel 7.

Tabel 7 Hasil Proses Data Preparation pada Data

Code	Recency	Frequency	Monetary
CC1	3	10	10
CC2	2	7	9
CC3	8	1	8
CC4	2	2	7
CC5	1	2	5
CC6	1	7	5
CC7	1	2	4
CC8	2	1	4
CC9	1	5	4
CC10	8	2	3

4.4 Data Modeling

Pada tahapan modelling digunakan algoritma data mining *clustering* berbasis partisi yaitu algoritma *k-means*. Percobaan modelling dilakukan dengan bantuan *framework data mining* yaitu WEKA untuk mengetahui nilai *Sum squared errors* yang terbaik. Metode ini merupakan metode visual. Idenya adalah dimulai dengan $k = 2$, dan terus meningkat dalam setiap langkah dengan ditambah 1 pada nilai k . Pada nilai $k=3$, apabila terjadi perubahan drastis yang berbanding terbalik dengan nilai sebelumnya, maka nilai sebelum terjadinya perubahan tersebut dianggap sebagai jumlah cluster yang paling tepat.

4.4.1 Pemodelan dengan dua klaster

Proses *modelling* data sekunder yang telah melalui proses *data preparation*, untuk mendapatkan klaster yang optimum berdasarkan nilai *Sum squared errors* yang dihasilkan dengan bantuan *framework data mining* WEKA maka dihasilkan:

1. Klaster pertama dengan jumlah anggota sebanyak 391 atau 27% pelanggan dan 1.058 atau sebanyak 73% pelanggan pada klaster kedua.
2. Terdapat 2 kali iterasi dengan Nilai *Sum squared errors* yang dihasilkan sebesar 38.48.

4.4.2 Pemodelan dengan tiga klaster

Dilakukan percobaan kedua dengan menggunakan tiga buah klaster, sehingga menghasilkan perubahan pada keanggotaan klaster dan nilai *Sum squared errors* yang

menggunakan bantuan *framework data mining* WEKA maka dihasilkan:

1. Klaster pertama dengan jumlah anggota sebanyak 391 atau 27% pelanggan, 20% atau 293 pelanggan pada klaster kedua dan 765 atau sebanyak 53% pelanggan pada klaster ketiga.
2. Terdapat 2 kali iterasi dengan Nilai *Sum squared errors* yang dihasilkan sebesar 22.07.

4.4.3 Pemodelan dengan empat klaster

Dilakukan percobaan ketiga dengan menggunakan empat buah klaster, sehingga menghasilkan perubahan pada keanggotaan klaster dan nilai *Sum squared errors* dengan menggunakan bantuan *framework data mining* WEKA maka dihasilkan:

1. Klaster pertama dengan jumlah anggota sebanyak 391 atau 27% pelanggan, 20% atau 293 pelanggan pada klaster kedua, 33% atau sebanyak 485 pelanggan pada klaster ketiga dan 280 atau sebanyak 19% pelanggan pada klaster keempat.
2. Terdapat 6 kali iterasi dengan Nilai *Sum squared errors* yang dihasilkan sebesar 17.67.

4.4.4 Pemodelan dengan lima klaster

Dilakukan percobaan keempat dengan menggunakan lima buah klaster, sehingga menghasilkan perubahan pada keanggotaan klaster dan nilai *Sum squared errors* yang menggunakan bantuan *framework data mining* WEKA maka dihasilkan:

1. Klaster pertama dengan jumlah anggota sebanyak 391 atau 27% pelanggan, 20% atau 293 pelanggan pada klaster kedua, 14% atau sebanyak 203 pelanggan pada klaster tiga, 19% atau sebanyak 276 pelanggan pada klaster keempat dan 286 atau sebanyak 20% pelanggan pada klaster kelima.
2. Terdapat 4 kali iterasi dengan Nilai *Sum squared errors* yang dihasilkan sebesar 16.23.

4.4.5 Pemodelan dengan enam klaster

Dilakukan percobaan lima dengan menggunakan enam buah klaster, sehingga menghasilkan perubahan pada keanggotaan klaster dan nilai *Sum squared errors* yang menggunakan bantuan *framework data mining* WEKA maka dihasilkan:

1. Klaster pertama dengan jumlah anggota sebanyak 218 atau 15% pelanggan, 20% atau 293 pelanggan pada klaster kedua, 14% atau sebanyak 203 pelanggan pada klaster tiga, 19% atau sebanyak 276 pelanggan pada klaster keempat, 286 atau sebanyak 20% pelanggan pada klaster kelima, dan 12% atau sebanyak 173 pelanggan pada klaster keenam.
2. Terdapat 4 kali iterasi dengan Nilai *Sum squared errors* yang dihasilkan sebesar 9.89.

4.4.6 Pemodelan dengan tujuh klaster

Dilakukan percobaan keenam dengan menggunakan tujuh buah klaster, sehingga menghasilkan perubahan pada keanggotaan klaster dan nilai *Sum squared errors* dengan menggunakan bantuan *framework data mining* WEKA maka dihasilkan:

1. Klaster pertama dengan jumlah anggota sebanyak 218 atau 15% pelanggan, 20% atau 293 pelanggan pada klaster kedua, 14% atau sebanyak 203 pelanggan pada klaster tiga, 19% atau sebanyak 276 pelanggan pada klaster keempat, 286 atau sebanyak 20% pelanggan pada klaster kelima, 7% atau sebanyak 100 pelanggan pada klaster keenam dan 5% atau sebanyak 73 pelanggan pada klaster ke tujuh.
2. Terdapat 4 kali iterasi dengan Nilai *Sum squared errors* yang dihasilkan sebesar 9.06.

4.4.7 Pemodelan dengan delapan klaster

Dilakukan percobaan ketujuh dengan menggunakan delapan buah klaster, sehingga menghasilkan perubahan pada keanggotaan klaster dan nilai *Sum squared errors* dengan menggunakan bantuan *framework data mining* WEKA maka dihasilkan:

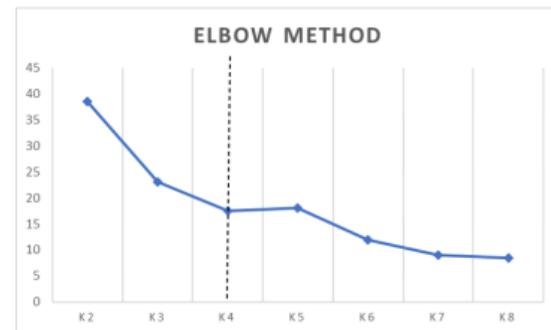
1. Klaster pertama dengan jumlah anggota sebanyak 92 atau 6% pelanggan, 20% atau 293 pelanggan pada klaster kedua, 14% atau sebanyak 203 pelanggan pada klaster tiga, 19% atau sebanyak 276 pelanggan pada klaster keempat, 286 atau sebanyak 20% pelanggan pada klaster kelima, 7% atau sebanyak 100 pelanggan pada klaster keenam, 5% atau sebanyak 73 pelanggan pada klaster ke tujuh dan yang terakhir sebanyak 126 atau 9% pelanggan.
2. Terdapat 4 kali iterasi dengan Nilai *Sum squared errors* yang dihasilkan sebesar 8.40.

Berdasarkan percobaan *modelling* diatas maka diperoleh perubahan-perubahan nilai *Sum squared errors* seperti data pada tabel 8.

Tabel 8 Perubahan -Perubahan Nilai SSE

Jumlah Klaster	Nilai SSE
K2	38.48
K3	22.07
K4	17.67
K5	16.23
K6	9.89
K7	9.06
K8	8.4

Dari tabel 8 dapat dibuatkan grafik untuk menentukan siku terbaik yang akan dijadikan



Gambar 6 Grafik nilai SSE

Berdasarkan grafik nilai *Sum squared errors* pada gambar 6 maka Penentuan nilai K menggunakan *Elbow Method* didapatkan titik siku yang terbentuk diantara K3 dan K5, sehingga dapat disimpulkan bahwa jumlah klaster berdasarkan *Elbow method* yaitu sebanyak 4 klaster.

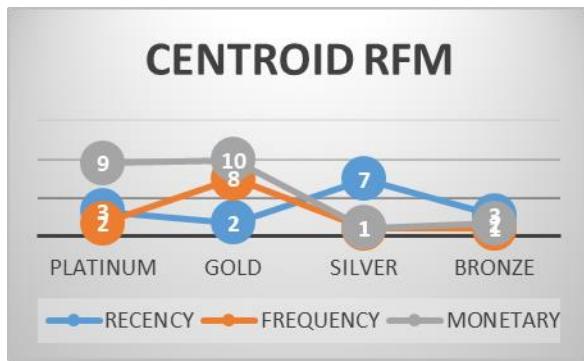
4.5 Data Evaluation

Prototipe dibuat menggunakan aplikasi python kemudian Data yang digunakan merupakan data penjualan pada tahun 2019 dengan jumlah pelanggan sebanyak 1.449 pelanggan, hasil dari pengujian menampilkan sebanyak empat buah klaster pelanggan yaitu 330 pelanggan masuk kedalam kelompok kategori *Platinum*, 557 pelanggan kategori *Gold*, 184 pelanggan kategori *Silver*, dan 378 pelanggan kategori *Bronze*. Grafik *Cluster Category* bisa dilihat pada gambar 7.



Gambar 7 Cluster Category

Pengelompokan hasil menjadi empat klaster sesuai dengan perbandingan dan analisis yang telah dilakukan menggunakan metode *elbow* yang terdapat pada tabel 8 maka diterapkan empat buah klaster dengan tiga buah atribut yaitu *recency*, *frequency* dan *monetary* menggunakan agloritma *k-means*. Hasil perhitungannya dapat dilihat pada Grafik *centroid RFM* pada gambar 8.



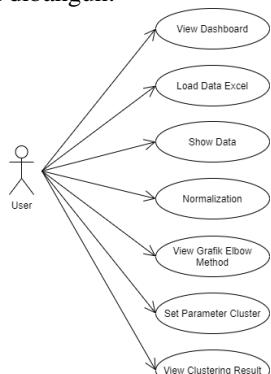
Gambar 8 Centroid RFM

4.6 Deployment

Rule yang dihasilkan dalam klasterisasi menggunakan atribut *recency*, *frequency* dan *monetary* dengan algoritma data mining berbasis partisi yaitu *k-means* akan diterapkan kedalam sebuah sistem

4.6.1 Use case

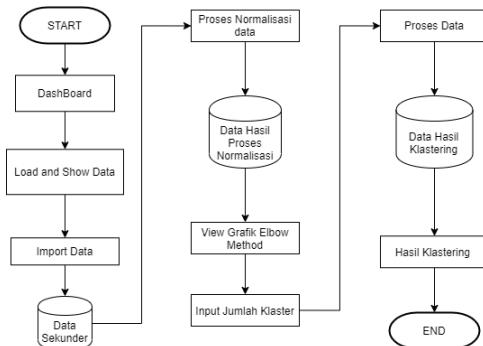
Kebutuhan sistem sesuai dengan kepentingan user yang telah dipetakantelah didapatkan pada tahapan *requirement analysis*, maka selanjutnya adalah tahapan logika perancangan. Berikut adalah rancangan *use case* user sistem yang akan dibangun.



Gambar 9 Use Case User

User dapat melakukan *Upload Data*, *view upload data*, Proses *Clustering* hingga *view clustering result*. Untuk dapat melakukan *clustering* setelah program dijalankan user harus melakukan proses load data yang telah tersedia dalam file excel, kemudian setelah proses upload data sukses maka akan ditampilkan data sekunder, user menormalisasi data dengan menekan tombol normalization setelah itu user bisa melihat grafik elbow method untuk menentukan jumlah klaster terbaik. Berdasarkan penelitian dan percobaan yang telah dilakukan, jumlah klaster yang paling optimum adalah empat klaster. Berdasarkan pertimbangan jumlah pelanggan yang mungkin dapat bertambah ataupun berkurang di tahun-tahun mendatang, sehingga mungkin dapat mempengaruhi jumlah klaster yang optimum, maka dibuat kolom untuk memasukkan jumlah klaster, setelah mengisi jumlah klaster pada kolom user menekan tombol proses K-means untuk memproses data.

4.6.2 Flowchart

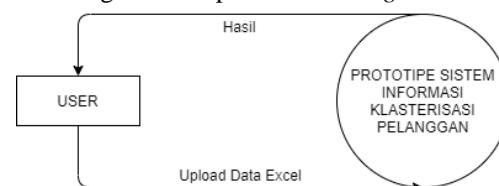


Gambar 10 Flowchart

Prototipe yang dibuat dimulai dengan proses upload atau import data berupa file excel, yang sebelumnya sudah disiapkan, kemudian dilanjutkan dengan proses normalisasi data yang didalamnya menggunakan metode Min Max, setelah itu melihat grafik yang diproses menggunakan metode Elbow untuk menentukan jumlah klasterisasi yang baik, kemudian dilanjutkan dengan proses data termasuk proses preparation data sesuai dengan metode *recency frequency monetary* dan klasterisasi data menggunakan algoritma data mining. Hasil dari data yang sudah selesai di proses dapat dilihat didalam Data tabel, *Clustering Result Data*.

4.6.3 Alur Data

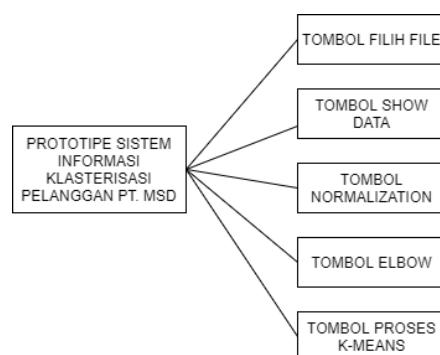
Dalam proses alur data, user hanya dapat melakukan upload data karena jika diinput satu persatu kedalam sistem akan memakan waktu kembali dan dapat menimbulkan human error, kemudian dilakukan proses normalisasi, melihat elbow grafik dan proses *clustering*.



Gambar 11 Alur Data

4.6.4 Physical Design and Integration

Perancangan *user interface* dalam penelitian ini hanya menampilkan satu halaman dengan beberapa tombol saja seperti pada gambar 12.



Gambar 12 Perancangan Antarmuka

Dalam sistem informasi ini hanya ada halaman utama yang didalamnya terdapat tombol-tombol seperti tombol pilih file yang akan diupload, tombol Show Data, tombol Normalization, tombol ELBOW dan tombol proses K-means, gambar 13 adalah tampilan sistem informasi klasterisasi.



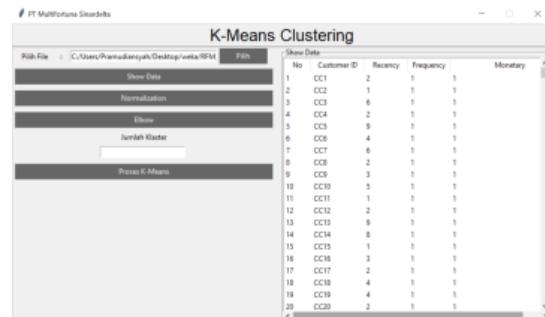
Gambar 13 Tampilan Sistem Informasi Klasterisasi

Perancangan tampilan prototipe sistem informasi ini seperti yang dijelaskan diatas yang hanya mempunyai halaman utama saja, proses penggunaan prototipe ini diawali dengan menekan tombol pilih untuk mencari file yang sebelumnya sudah disiapkan dalam bentuk excel berupa data sekunder yang akan diproses dengan cara menekan tombol show data seperti pada gambar 14 .



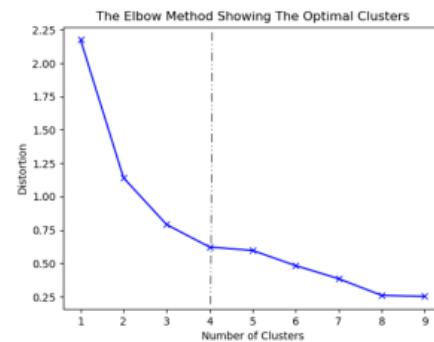
Gambar 14 Show Data upload

Setelah menekan tombol show data sistem akan menampilkan semua data yang telah diupload pada bagian tabel sebelah kanan, data yang ditampilkan adalah data sekunder yang sudah disiapkan sebelumnya, proses selanjutnya adalah menormalisasi data menggunakan metode Min Max dengan cara menekan tombol normalization, tampilan data pada kolom Recency, Frequency dan Monetary akan berubah seperti pada gambar 15.



Gambar 15 Setelah Data Dinormalisasi

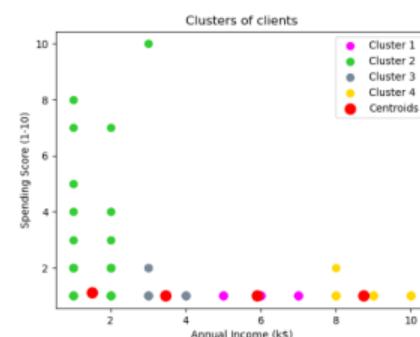
Setelah menekan tombol normalisasi prototipe akan menampilkan data recency, frequency dan monetary yang sudah dirubah menggunakan metode Min Max dengan range 1-10. Setelah dinormalisasi user bisa langsung menekan tombol elbow untuk melihat grafik, gambar 16 menunjukan grafik elbow method.



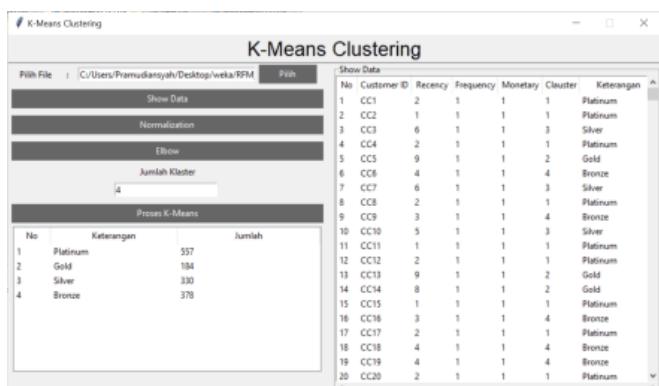
Gambar 16 Elbow Method

Berdasarkan grafik elbow method pada gambar 16 maka Penentuan nilai K menggunakan Elbow Method didapatkan titik siku yang terbentuk diantara K3 dan K5, ini sesuai dengan grafik elbow method yang menggunakan nilai SSE (Sum Square Error) menggunakan framework WEKA, sehingga dapat disimpulkan bahwa jumlah klaster terbaik adalah sebanyak 4 klaster.

Setelah mengetahui jumlah klasternya tahap selanjutnya user harus menginput nilai tersebut di field jumlah klaster, kemudian user menekan tombol proses k-means untuk memproses data yang sudah dinormalisasi, hasil proses tersebut bisa dilihat di gambar 17 plot sebaran data 4 klaster dan gambar 18 Klaster kategori.



Gambar 17 Plot Sebaran Data 4 Klaster



Gambar 18 Klaster kategori

Pada gambar 17 adalah gambar plot sebaran data dari empat klaster dinama masing-masing warna mewakili tiap klaster, dimana warna *pink* mewakili klaster 1, warna hijau mewakili klaster 2, warna abu-abu mewakili klaster 3 dan warna kuning mewakili klaster 4. Sedangkan warna merah adalah titik *centroid*.

Kemudian pada gambar 18 pada tabel sebelah kiri akan menampilkan kolom *customer category* dan jumlah, terlihat categori pertama yaitu *Platinum* dengan jumlah 557 pelanggan, kategori *gold* 184 pelanggan, kategori *silver* sebanyak 330 pelanggan dan kategori *bronze* berjumlah 378 pelanggan. Pada tampilan tabel sebelah kanan akan muncul kolom baru yaitu kolom *cluster* dan *customer category* yang isinya adalah hasil akhir dari kalstesisasi, misalkan yang terlihat pada gambar *customer id CC1* mempunyai nilai *recency* 2, *frequency* 1, dan *monetary* 1, masuk kedalam klaster ke 1 yang berarti termasuk kategori *customer platinum*.

5. KESIMPULAN

Berdasarkan hasil dari penelitian mengenai Segmentasi Pelanggan Menggunakan Algoritma *Data Mining* Berdasarkan Model *Recency Frequency Monetary* maka didapat kesimpulan bahwa hasil pengujian menggunakan *Elbow Method* didapatkan titik siku terbaik yang terbentuk diantara K3 dan K5, sehingga dapat disimpulkan bahwa jumlah klaster optimal yaitu sebanyak 4 klaster. Kemudian Algoritma yang tepat digunakan untuk model dataset yang dimiliki adalah agloritma *k-means* dengan nilai *DBI* sebesar 0.525.

Setelah dilakukan perhitungan menggunakan algoritma *k-means* dengan jumlah 4 klaster maka dihasilkan sebanyak 38,4% atau berjumlah 557 pelanggan masuk kedalam kelompok kategori *Platinum*, 12,6% atau berjumlah 184 pelanggan kategori *Gold*, 22,7% atau berjumlah 330 pelanggan kategori *Silver*, dan 26% atau berjumlah 378 pelanggan kategori *Bronze*.

6. DAFTAR PUSTAKA

- [1] D. P. Hidayatullah, R. I. Rokhmawati, and A. R.

Perdanakusuma, "Analisis Pemetaan Pelanggan Potensial Menggunakan Algoritma K-Means dan LRFM Model Untuk Mendukung Strategi Pengelolaan Pelanggan (Studi Pada Maninjau Center Kota Malang)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 8, pp. 2406–2415, 2018.

- [2] R. R. Putra and C. Wadisman, "IMPLEMENTASI DATA MINING PEMILIHAN PELANGGAN POTENSIAL MENGGUNAKAN ALGORITMA K-MEANS," *J. Inf. Technol. Comput. Sci.*, vol. 53, no. 9, pp. 1689–1699, 2019.
- [3] D. Juliawan, F. Amir, and E. Desi, "Penerapan Data Mining Metode Clustering Pada CV. Secom Infotech Menggunakan Algoritma K-Means," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. Senatkom, pp. 96–102, 2019.
- [4] B. E. Adiana, I. Soesanti, and A. E. Permanasari, "Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering," *Jutei*, vol. 2, no. 2, pp. 23–32, 2018.
- [5] R. Irmariani, "Analisis Customer Relationship Management (CRM) Pada BPR Supra Bogor," Institut Pertanian Bogor, 2013.
- [6] Kirwanto, "Pengembangan E-CRM Pada UMKM Barbershop Untuk Meningkatkan Daya Saing Terhadap Kompetitor Serta Pelayanan Terhadap Pelanggan," Budiluhur, 2017.
- [7] I. G. A. Suwartane, "Model Prediksi Penyakit Diabetes Berbasis Heterogeneous Multiple Classifiers," 2018.
- [8] H. Junaedi, H. Budianto, I. Maryati, and Y. Melani, "Data Transformation pada Data Mining," *Pros. Konf. Nas. Inov. dalam Desain dan Teknol.*, vol. 7, pp. 93–99, 2011.
- [9] D. Astria and Suparyogi, "Penerapan Algoritma Fuzzy C-Means Untuk Clustering," *Univ. Dian Nuswantoro Semarang*, vol. 6, no. Maret 2017, pp. 169–178, 2017.
- [10] G. F. Wulandari, "Segmentasi Pelanggan Menggunakan Algoritma K-Means Untuk Customer Relationship Management (CRM) Pada Hijab Miulan," *Ind. Mark. Manag.*, vol. I, no. segmentasi pelanggan, p. 7, 2014.
- [11] R. Braganca, F. Portela, A. Vale, T. Guimaraes, and M. Santos, "Data Mining Classification Models for Industrial Planning," vol. 1, no. July 2017, p. 5427, 2019.