

BAB 4 PERANCANGAN

Bab ini menjelaskan tentang rancangan penelitian yang terdiri dari pengumpulan data, *pre processing data*, dan desain visualisasi *dashboard*.

4.1 Pengumpulan Data

Pengumpulan data dilakukan dengan mengekstraksi data mentah pada Klinik Kecantikan Belle Crown Malang. Periode data transaksi yang digunakan yakni pada tanggal 2 Mei 2017 hingga 30 Oktober 2017 berupa soft copy. Data tersebut berupa *file excel* dengan variabel seperti pada Tabel 4.1.

Tabel 4.1 Variabel Data Mentah Tabel Transaksi Pelanggan

Variabel	Tipe Data	Keterangan
NoFakt	Text	ID transaksi pelanggan
Tanggal	Date	Tanggal transaksi pelanggan
Customer	Text	ID pelanggan yang melakukan transaksi
Total	Number	Total yang dibayar pelanggan dalam satu kali transaksi
Divisi	Date	Jenis layanan yang digunakan dalam transaksi

Berdasarkan data transaksi yang ada, terhitung jumlah transaksi sebanyak 21.513 transaksi dari bulan Mei hingga Oktober. Sedangkan jumlah pelanggan yang terdaftar pada tabel ini sebanyak 4716 orang.

4.2 Data Selection

Dilakukan pergantian nama variabel untuk mempermudah proses *clustering*. Berikut adalah nama variabel yang diubah pada Tabel 4.2.

Tabel 4.2 Perubahan Variabel Data Mentah Tabel Transaksi Pelanggan

Variabel lama	Variabel Baru
NoFakt	id_transaksi
Tanggal	Tanggal

Customer	id_customer
Total	Pembayaran
Divisi	Layanan

Pada Tabel 4.2, variabel yang dipilih adalah tanggal, id_customer, dan total. Sehingga tabel hasil pemilihan variabel ditampilkan dalam Tabel 4.3. Sedangkan potongan data mentah ditampilkan pada Lampiran A.

Tabel 4.3 Variabel Terpilih dari Tabel Transaksi Pelanggan

Tanggal	Id_customer	pembayaran
02-May-17	R000567	142800
02-May-17	007.2016.001105	43000
02-May-17	007.2016.003234	97185
02-May-17	001.2014.000926	54000
02-May-17		105000

4.3 Ekstraksi RFM

Tahap ini berfokus pada analisis RFM dengan memetakan data ke dalam variabel RFM kemudian dilakukan normalisasi. Proses pengelompokkan menggunakan *tools phpmyadmin* menggunakan SQL *query*. Proses ekstraksi ke dalam RFM dilakukan pada 07 November 2017.

4.3.1 Ekstraksi Variabel *Recency*

Recency adalah nilai rentang waktu antara tanggal penelitian yakni 07 November 2017 dengan transaksi terakhir masing-masing pelanggan. Variabel yang dibutuhkan yakni tanggal dan id_customer untuk mengidentifikasi pelanggan.

4.3.2 Ekstraksi Variabel *Frequency*

Frequency adalah jumlah transaksi yang dilakukan pelanggan dalam satu periode yakni periode 02 Mei 2017 sampai dengan 30 Oktober 2017. Variabel yang dibutuhkan sama dengan variabel *recency* yakni tanggal dan id_customer untuk mengidentifikasi pelanggan.

4.3.3 Ekstraksi Variabel *Monetary*

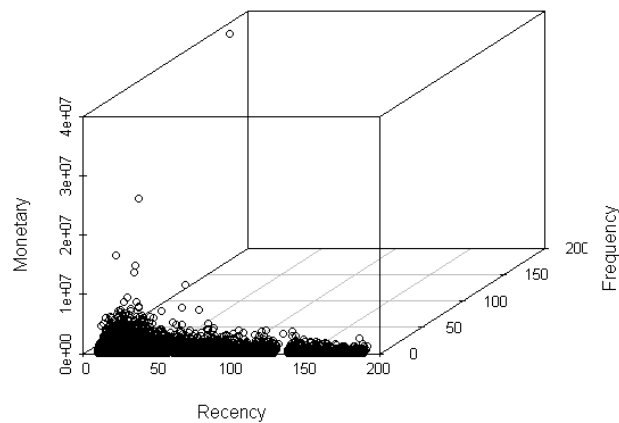
Monetary adalah besaran nilai rupiah yang dikeluarkan pelanggan selama transaksi. Variabel yang dibutuhkan sama dengan variabel *recency* yakni tanggal dan *id_customer* untuk mengidentifikasi pelanggan.

4.4 Pre Processing Data

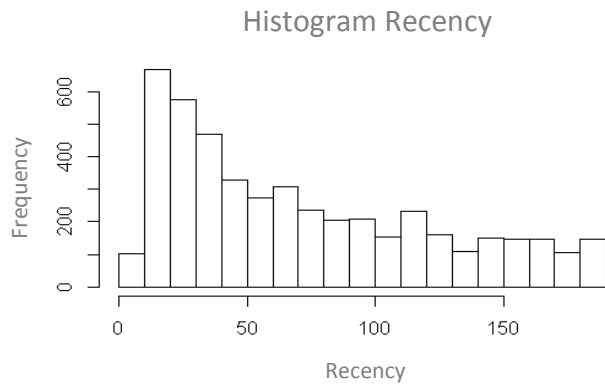
Terdapat dua metode normalisasi yang digunakan yakni normalisasi *Logarithmic* yang berfungsi untuk menyeragamkan data yang rentangnya cukup jauh antar variabel dan normalisasi *Min-Max* guna memperkecil persebaran nilai dari masing-masing variabel ke dalam angka 0 hingga 1. Normalisasi akan menggunakan metode *Min-Max*.

4.4.1 Data Transformation

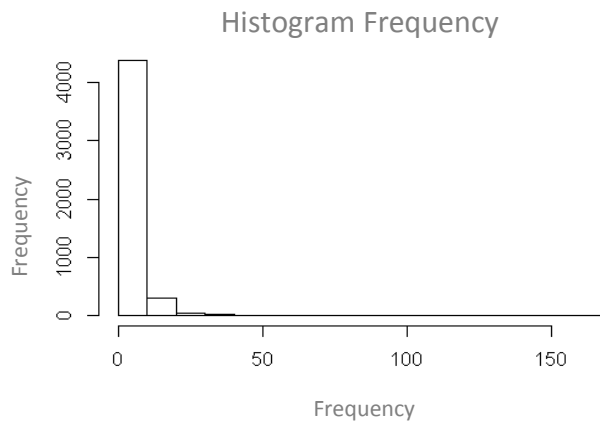
Berdasarkan hasil 3D *scstter plot* yang menggambarkan masing-masing variabel RFM seperti Gambar 4.1 menunjukkan bahwa kebanyakan data terkumpul dalam satu area. Selain itu, jika dipetakan ke dalam grafik histogram seperti pada Gambar 4.2 sampaidengan Gambar 4.3 yang menunjukkan bahwa data tidak normal. Oleh karena itu perlu dilakukan transformasi data. Sehingga diperlukan metode *Logarithmic*.



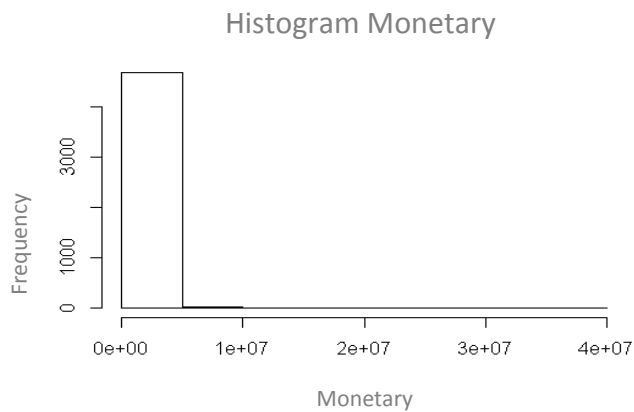
Gambar 4.1 Scatterplot Sebelum Cleansing



Gambar 4.2 Histogram Data Variabel *Recency* Sebelum *Cleansing*



Gambar 4.3 Histogram Data Variabel *Frequency* Sebelum *Cleansing*



Gambar 4.4 Histogram Data Variabel *Monetary* Sebelum *Cleansing*

4.4.2 Data Cleansing

Pada data mentah terdapat *id_customer* yang bernilai kosong, sehingga tidak dapat digunakan untuk mengidentifikasi pelanggan, sehingga perlu dilakukan *data cleansing* yakni dengan menghapus baris pada kolom yang kosong sejumlah 338 baris transaksi. Selain itu juga terdapat nilai pembayaran samadengan 1, sehingga perlu dilakukan penghapusan baris pada kolom *variable* pembayaran sejumlah 1470 baris transaksi. Setelah data dengan kolom kosong dan nilai pembayaran=1, maka dilanjutkan dengan proses ekstraksi RFM.

Data yang mengandung *outlier* dapat merusak hasil *clustering*, sehingga perlu dilakukan *data cleansing* terlebih dahulu. Proses *data cleansing* menggunakan penghapusan *outlier* oleh Klodian Dhana (2016). Hasil dari *data cleansing* dapat dibandingkan dengan keadaan data sebelum dilakukan penghapusan *outlier*.

4.4.3 Normalisasi Min-Max

Nilai *range* yang berbeda pada variabel RFM membuat hasil *clustering* menjadi tidak valid sehingga perlu dilakukan normalisasi. Pada penelitian ini, normalisasi dilakukan menggunakan metode *Min-Max* dengan menyederhanakan nilai menjadi 0 hingga 1. Berdasarkan penelitian yang telah dilakukan Angelie, Annisa Veronica (2017), dikarenakan nilai *Recency* berbanding terbalik dengan nilai *Frequency* dan *Monetary*, maka perlu dibalik dengan proses pengurangan 1 terhadap hasil normalisasi pada variabel *Rerency*. Nilai *Recency* semakin baik bila memiliki nilai yang kecil. Akan tetapi, nilai *Recency* bertentangan dengan nilai yang lain sehingga sebelumnya nilai *Recency* akan dibalik dengan melakukan pengurangan 1 dengan hasil normalisasi *Recency*.

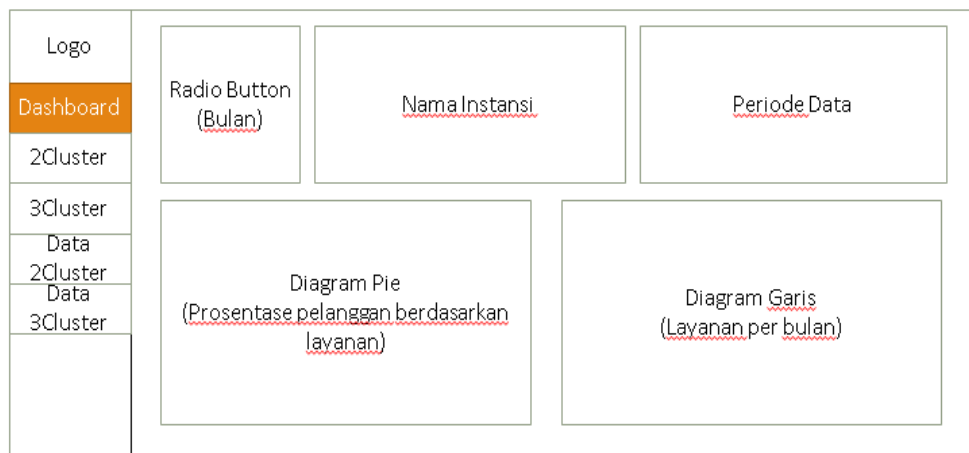
4.5 Penentuan Jumlah Cluster

Untuk menentukan jumlah *cluster* digunakan metode *Elbow* yakni dengan melihat penurunan grafik nilai *k* terhadap nilai *SSE* secara *ntuitive*. Jika suatu titik mengalami penurunan yang dramatis, kemudian titik berikutnya berada pada nilai yang stabil, maka titik tersebut dipilih menjadi nilai *k*. *Tools* yang digunakan dalam proses penentuan jumlah *cluster* dan proses *clustering* adalah R studio. Berdasarkan hasil grafik, maka jumlah *cluster* yang dipilih yakni 3 *cluster*.

4.6 Desain Antarmuka Visualisasi

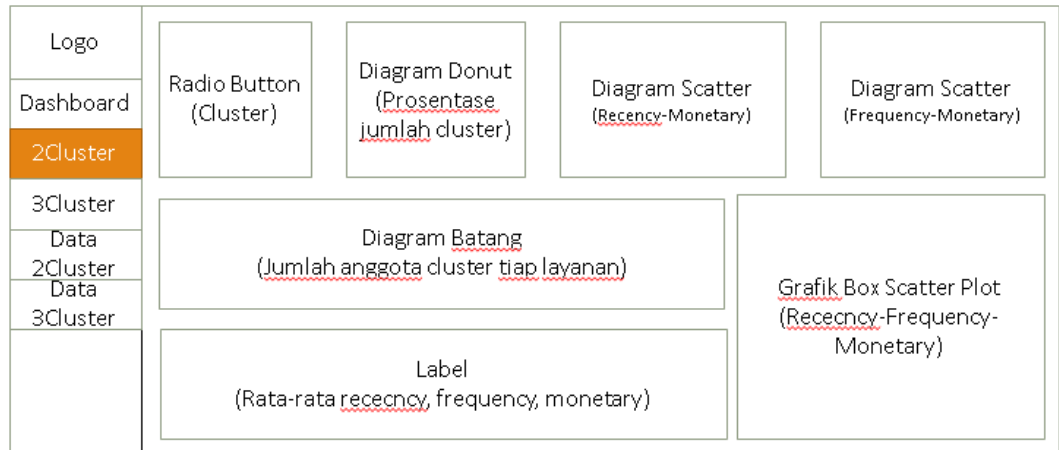
Rancangan rancangan visualisasi terdiri dari tiga halaman utama yaitu halaman *dashboard*, halaman *cluster* dan halaman data. Dalam penerapannya menggunakan *tools Power BI* yang diintegrasikan dengan R.

Gambar 4.5 menggambarkan halaman *dashboard*. Pada halaman ini dimuat data prosentase pelanggan berdasarkan layanan untuk melihat jumlah pelanggan berdasarkan layanan dalam 6 bulan terakhir dan grafik jumlah pelanggan berdasarkan layanannya per bulan.



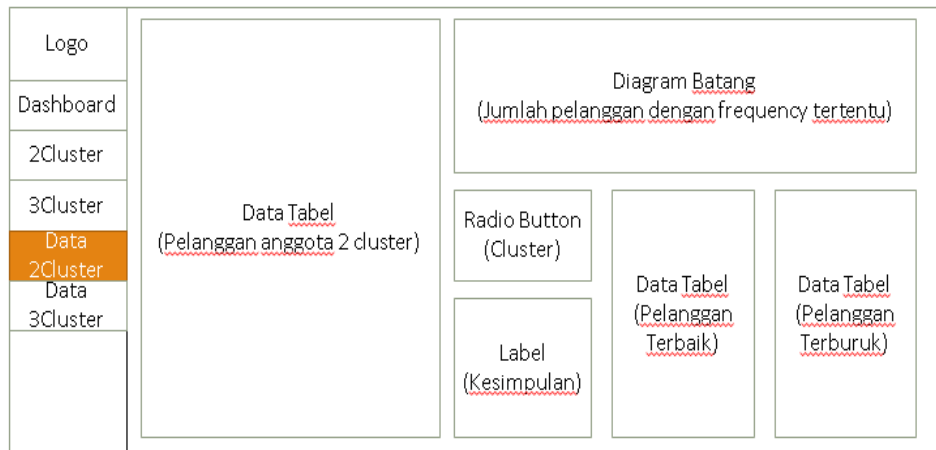
Gambar 4.5 Rancangan Halaman *Dashboard*

Sedangkan rancangan halaman *cluster* ditunjukkan pada Gambar 4.6 yang memuat beberapa grafik dan diagram, yakni diagram donut untuk melihat prosentase jumlah pelanggan tiap *cluster*, diagram *scatter* untuk membandingkan nilai antar variable *recency-monetary* dan *frequency-monetary*, untuk memvisualisasikan jumlah pelanggan tiap cluster pada setiap layanan digunakan diagram batang, box label untuk menampilkan nilai rata-rata dari persebaran *cluster*, dan grafik *box scatter plot* untuk menampilkan persebaran pelanggan berdasarkan variable. Gambar 4.6 merupakan representasi halaman *cluster* untuk 2 segmen, sedangkan untuk 3 segmen tampilannya sama, hanya berbeda pada jumlah *clusternya*.



Gambar 4.6 Rancangan Halaman Cluster

Rancangan halaman data ditunjukkan pada Gambar 4.7 yang memuat tabel informasi *cluster* dengan variabel *id_customer*, nilai RFM, dan *clusternya*. Diagram batang yang merepresentasikan jumlah pelanggan dengan *frequency* tertentu. Selain itu juga memuat peringkat pelanggan terbaik dan terburuk. Gambar 4.7 merupakan representasi halaman data pelanggan untuk 2 segmen, sedangkan untuk 3 segmen tampilannya sama, hanya berbeda pada jumlah *clusternya*.



Gambar 4.7 Rancangan Halaman Data