

SEGMENTASI PELANGGAN MENGGUNAKAN METODE *FUZZY C-MEANS CLUSTERING* BERDASARKAN LRFM MODEL PADA TOKO SEPATU (STUDI KASUS: RIDE INC KOTA MALANG)

SKRIPSI

Untuk memenuhi sebagian persyaratan
Memperoleh gelar Sarjana Komputer

Disusun oleh:

Muhammad Taufik Dharmawan
145150400111012



PROGRAM STUDI SISTEM INFORMASI
JURUSAN SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BRAWIJAYA
MALANG
2018

PENGESAHAN

SEGMENTASI PELANGGAN MENGGUNAKAN METODE FUZZY C-MEANS CLUSTERING BERDASARKAN LRFM MODEL PADA TOKO SEPATU (STUDI KASUS: RIDE INC KOTA MALANG)

SKRIPSI

Diajukan untuk memenuhi sebagian persyaratan memperoleh gelar Sarjana Komputer

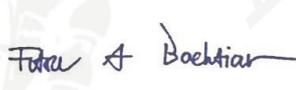
Disusun Oleh :
Muhammad Taufik Dharmawan
NIM: 145150400111012

Skripsi ini telah diuji dan dinyatakan lulus pada 11 Desember 2018
Telah diperiksa dan disetujui oleh:

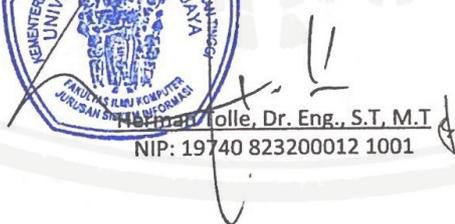
Dosen Pembimbing I

Dosen Pembimbing II


Nanang Yudi Setiawan, S.T., M.Kom.
NIP: 19760619 200604 1 001


Dr. Eng. Fitra Abdurrachman Bachtiar,
S.T, M.Eng
NIK: 201201 840628 1 001

Mengetahui
Ketua Jurusan Sistem Informasi


Herman Jolle, Dr. Eng., S.T, M.T
NIP: 19740 823200012 1001



PERNYATAAN ORISINALITAS

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah skripsi ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini dan disebutkan dalam daftar pustaka.

Apabila ternyata didalam naskah skripsi ini dapat dibuktikan terdapat unsur-unsur plagiasi, saya bersedia skripsi ini digugurkan dan gelar akademik yang telah saya peroleh (sarjana) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).

Malang, 11 Desember 2018



Muhammad Taufik Dharmawan

NIM: 145150400111012

KATA PENGANTAR

Puji Syukur ke hadirat Allah SWT atas segala rahmat yang telah diberikan sehingga memberikan kemudahan dan kelancaran dalam menyelesaikan skripsi yang berjudul “Segmentasi Pelanggan Menggunakan Metode *Fuzzy C-Means Clustering* Berdasarkan LRFM Model Pada Toko Sepatu (Studi Kasus: Ride Inc Kota Malang).”

Penyusunan dan penulisan skripsi ini tidak terlepas dari bantuan, bimbingan, dukungan, serta motivasi dari berbagai pihak. Namun, penulis dapat melewati hal tersebut karena anugerah dari Allah SWT. Oleh karena itu, dalam kesempatan ini penulis ingin menyampaikan terimakasih kepada:

1. Bapak Nanang Yusi Setiawan, S.T., M.Kom. selaku dosen pembimbing satu yang senantiasa membimbing dan memberikan masukan kepada penulis.
2. Bapak Dr.Eng. Fitra Abdurrachman Bachtiar, S.T, M.Eng. selaku dosen pembimbing dua yang senantiasa membimbing dan memberikan masukan kepada penulis.
3. Bapak Wayan Firdaus Mahmudy, S.Si., M.TI., Ph.D. selaku Dekan Fakultas Ilmu Komputer Universitas Brawijaya.
4. Bapak Dr. Eng., Herman Tolle, S.T., M.T. selaku Ketua Jurusan Sistem Informasi Universitas Brawijaya.
5. Bapak Yusi Tyroni Mursityo, S.Kom., M.AB. selaku Ketua Program Studi Jurusan Sistem Informasi.
6. Kepada Ibu Ir. Ermi Sulastri selaku orang tua dan Annisa Meilinda Handayani selaku kakak yang telah memberikan kepercayaan, motivasi, dan doa.
7. Mas Zacharia Lukkyta selaku pendiri Ride Inc. yang telah memberikan kesempatan kepada penulis untuk melakukan penelitian.
8. Lembaga KBMSI terkhusus pada Badan Pengurus Harian KBMSI dan Departemen Sosial Masyarakat periode 2017-2018 yang selalu menghibur dan menyemangati penulis.
9. Satrio Dwiartono, Wiratama Ahsani Taqwim, Claudio Canigia Guntara, Jebi Hayi Tamami, dan Fadzri Ahdi Anshori yang tidak kenal lelah menemani dan menyemangati penulis.
10. Danny Esfarditya, Fandy Adityo, Rahmi Maulidia, Annisa Sesyazhade, Intan Camila, Iffa Aulia Ulwani, dan Deafinansia Andiyani yang selalu siap mendorong penulis di saat penulis kehilangan semangat.
11. Aldino Fachrizky, Rizki Mardhotillah, Kevin Ginola, dan Eko Hartanto yang telah menemani dan menghibur penulis dalam keluarga kontrakan beriman.
12. Remaja Anjay yang telah mendengarkan serta menghibur dalam sebagai teman satu perantauan penulis.

13. Fathir Ibnu, Syafiq Al-Jufrie, dan Prisnita Naulia yang telah meluangkan waktu untuk menghibur penulis.

Malang, 11 Desember 2018

Penulis

mtaufikdhrmwn@gmail.com



ABSTRAK

Ride Inc. merupakan toko sepatu kulit yang berdiri pada tahun 2012 di Kota Malang. Data pelanggan yang dimiliki belum dimanfaatkan untuk mendapatkan nilai yang mampu membantu Ride Inc. dalam menyusun strategi bisnis untuk mendapatkan dan mempertahankan pelanggan. Selain itu, penentuan jumlah kulit yang dibeli menjadi penyebab dalam kerugian. Segmentasi pelanggan merupakan pembagian kelompok terhadap pelanggan yang memiliki kemiripan karakteristik dan mampu memberikan informasi dari kelompok yang menguntungkan dan tidak menguntungkan. Karakteristik pelanggan dapat dilihat dengan menerapkan LRFM model atau *Length* (rentang waktu transaksi terakhir dan pertama), *Recency* (rentang waktu penelitian dengan transaksi terakhir), *Frequency* (jumlah transaksi yang dilakukan), dan *Monetary* (total uang yang telah dikeluarkan). Metode yang digunakan dalam pengelompokan pelanggan adalah *Fuzzy C-Means clustering*. Dalam penerapan metode *clustering*, *elbow method* digunakan untuk membantu proses penentuan jumlah *cluster* yang baik secara iteratif dengan menampilkan hasil berupa grafik. Data yang digunakan dari Ride Inc. adalah data transaksi pelanggan sejumlah 668 data transaksi dan 522 pelanggan pada periode Juli 2017 hingga Maret 2018. Hasil jumlah *cluster* berdasarkan *elbow method* adalah dua *cluster* dan tiga *cluster* yang kemudian diimplementasikan ke dalam *Fuzzy C-Means*. Analisis segmen pelanggan berdasarkan LRFM model berupa peringkat dari kelompok pelanggan yang menguntungkan dengan melihat nilai L, F, dan M tertinggi dan R terendah. Visualisasi *dashboard* yang berisi tiga halaman dengan bentuk grafik dan diagram menjadi *output* penelitian berdasarkan nilai LRFM yang diberikan ke Ride Inc. Hasil rata-rata dari uji *usability* terhadap 2 responden adalah 65 yang berarti pihak Ride Inc. menerima visualisasi *dashboard* tersebut.

Kata kunci: segmentasi pelanggan, LRFM model, fuzzy c-means, elbow method, dashboard.

ABSTRACT

Ride Inc. is a leather shoe store established in 2012 in the city of Malang. Customer data owned has not been utilized to obtain values that are able to help Ride Inc. in creating business strategies to gain and retain customers. In addition, the determination of the amount of the purchased leather be the cause in the loss. Customer segmentation is groups of customers who have a similarity of characteristics that able to provide information from a group of profitable and unprofitable. The characteristics of customers can be seen by applying LRFM model or Length (with the span time of the first and last transaction), Recency (with the time span of research with the last transaction), Frequency (the number of transactions carried out) and Monetary (total the money has been issued). Methods used in grouping customers are Fuzzy C-Means clustering. In applying the method of clustering, elbow method used to help the process of determining the best number of clusters iteratively and to display the results in the form of graphs. The data used from the Ride Inc. are 668 customers transaction data with 522 customers in the period of July 2017 until March 2018. The results shows that there are two clusters and three clusters that formed based on elbow method that are then implemented into the Fuzzy C-Means. Customer segment analysis based on the LRFM model formed in the rank of profitable customer group based on the highest value of L, F, and also M and the lowest R. Dashboard visualization that contains three pages with graphs and diagrams become the output of the research based on the value of the given LRFM to Ride Inc. The average score of usability testing from 2 respondents is 65. This means that Ride Inc. receive the dashboard visualization.

Keywords: customer segmentation, LRFM model, fuzzy c-means, elbow method, dashboard.

DAFTAR ISI

| | |
|---|-------------------------------------|
| PENGESAHAN | Error! Bookmark not defined. |
| PERNYATAAN ORISINALITAS | Error! Bookmark not defined. |
| KATA PENGANTAR..... | iv |
| ABSTRAK..... | vi |
| ABSTRACT | vii |
| DAFTAR ISI..... | viii |
| DAFTAR TABEL..... | xii |
| DAFTAR GAMBAR..... | xiii |
| DAFTAR SCRIPT..... | xv |
| DAFTAR LAMPIRAN | xvi |
| BAB 1 PENDAHULUAN..... | 1 |
| 1.1 Latar belakang..... | 1 |
| 1.2 Rumusan masalah..... | 3 |
| 1.3 Tujuan | 3 |
| 1.4 Manfaat..... | 3 |
| 1.5 Batasan masalah | 4 |
| 1.6 Sistematika pembahasan | 4 |
| BAB 2 LANDASAN KEPUSTAKAAN | 6 |
| 2.1 Kajian Pustaka | 6 |
| 2.2 Profil Toko Sepatu Ride Inc Kota Malang..... | 7 |
| 2.3 Customer Relationship Management (CRM)..... | 9 |
| 2.4 Data Mining..... | 9 |
| 2.5 Model LRFM (<i>Length, Recency, Frequency, Monetary</i>)..... | 11 |
| 2.6 Normalisasi Min-Max..... | 11 |
| 2.7 Metode <i>Elbow</i> | 11 |
| 2.8 <i>Clustering</i> | 12 |
| 2.8.1 Definisi <i>Clustering</i> | 12 |
| 2.8.2 Metode <i>Clustering</i> | 13 |
| 2.9 <i>Fuzzy C-Means Clustering</i> | 14 |
| 2.10 <i>Modified Partition Coefficient (MPC)</i> | 15 |

| | |
|--|-----------|
| 2.11 <i>Davies Bouldin's Index</i> | 15 |
| 2.12 <i>Connectivity</i> | 16 |
| 2.13 <i>Dunn Index</i> | 16 |
| 2.14 <i>Silhouette Width</i> | 16 |
| 2.15 <i>Euclidean Distance</i> | 16 |
| 2.16 <i>Dashboard</i> | 17 |
| 2.17 <i>System Usability Scale (SUS)</i> | 17 |
| BAB 3 METODOLOGI | 18 |
| 3.1 <i>Metodologi Penelitian</i> | 18 |
| 3.1.1 <i>Pemahaman Bisnis dan Identifikasi Masalah</i> | 19 |
| 3.1.2 <i>Studi Literatur</i> | 19 |
| 3.1.3 <i>Pengumpulan Data</i> | 19 |
| 3.1.4 <i>Preprocessing Data</i> | 19 |
| 3.1.5 <i>Clustering</i> | 19 |
| 3.1.6 <i>Uji Validitas Hasil Clustering</i> | 20 |
| 3.1.7 <i>Uji Verifikasi Hasil Clustering</i> | 20 |
| 3.1.8 <i>Pembuatan Visualisasi Hasil Clustering</i> | 21 |
| 3.1.9 <i>Analisis Hasil Usability Testing</i> | 21 |
| 3.1.10 <i>Kesimpulan dan Saran</i> | 21 |
| BAB 4 PERANCANGAN | 22 |
| 4.1 <i>Pengumpulan Data</i> | 22 |
| 4.2 <i>Ekstraksi LRFM Model</i> | 22 |
| 4.2.1 <i>Ekstraksi Variabel Length</i> | 22 |
| 4.2.2 <i>Ekstraksi Variabel Recency</i> | 23 |
| 4.2.3 <i>Ekstraksi Variabel Frequency</i> | 23 |
| 4.2.4 <i>Ekstraksi Variabel Monetary</i> | 23 |
| 4.3 <i>Data Preprocessing</i> | 23 |
| 4.3.1 <i>Data Selection</i> | 23 |
| 4.3.2 <i>Data Transformation</i> | 24 |
| 4.3.3 <i>Data Cleaning</i> | 24 |
| 4.3.4 <i>Data Normalization</i> | 24 |
| 4.4 <i>Penentuan Jumlah Cluster</i> | 25 |

| | |
|--|-----------|
| 4.5 Clustering Menggunakan Fuzzy C-Means | 25 |
| 4.6 Pengujian Cluster | 25 |
| 4.6.1 Uji Validitas Cluster | 25 |
| 4.6.2 Uji Verifikasi Cluster | 25 |
| 4.7 Desain Arsitektur Proses Clustering..... | 26 |
| 4.8 Desain Visualisasi Dashboard | 26 |
| 4.9 Pengujian System Usability Scale (SUS) | 28 |
| BAB 5 IMPLEMENTASI | 29 |
| 5.1 Data Preprocessing | 29 |
| 5.1.1 Data Selection | 29 |
| 5.1.2 Pemetaan LRFM | 29 |
| 5.1.3 Data Transformation..... | 30 |
| 5.1.4 Data Cleaning..... | 31 |
| 5.1.5 Normalisasi Min-Max..... | 32 |
| 5.2 Clustering menggunakan Fuzzy C-Means | 32 |
| 5.2.1 Elbow Method | 33 |
| 5.2.2 Fuzzy C-Means..... | 33 |
| 5.3 Uji Validitas Cluster | 33 |
| 5.3.1 Modified Partition Coefficient (MPC)..... | 33 |
| 5.3.2 Davies Bouldin's Index..... | 34 |
| 5.3.3 CValid..... | 34 |
| 5.4 Uji Verifikasi Cluster | 34 |
| 5.5 Visualisasi Dashboard | 35 |
| BAB 6 ANALISIS DAN PEMBAHASAN | 38 |
| 6.1 Pemetaan LRFM..... | 38 |
| 6.2 Pre-processing Data | 39 |
| 6.2.1 Data Transformation..... | 39 |
| 6.2.2 Data Cleaning..... | 41 |
| 6.2.3 Normalisasi Min-Max | 42 |
| 6.3 Clustering | 44 |
| 6.3.1 Elbow Method | 44 |
| 6.3.2 Clustering Menggunakan Fuzzy C-Means | 45 |



| | |
|--|----|
| 6.4 Uji Validitas <i>Cluster</i> | 48 |
| 6.4.1 <i>Modified Partition Coefficient</i> (MPC) | 48 |
| 6.4.2 <i>Davies Bouldin's Index</i> | 48 |
| 6.4.3 CIVValid..... | 49 |
| 6.5 Uji Verifikasi | 50 |
| 6.6 Analisis Hasil <i>Clustering</i> | 51 |
| 6.6.1 Analisis 2 <i>Cluster</i> | 52 |
| 6.6.2 Analisis 3 <i>Cluster</i> | 53 |
| 6.7 Visualisasi Hasil <i>Clustering</i> | 56 |
| 6.7.1 Halaman Utama | 56 |
| 6.7.2 Halaman 2 <i>Cluster</i> | 58 |
| 6.7.3 Halaman Data 2 <i>Cluster</i> | 59 |
| 6.8 Analisis Hasil <i>System Usability Scale</i> (SUS) | 61 |
| BAB 7 Penutup | 63 |
| 7.1 Kesimpulan..... | 63 |
| 7.2 Saran | 64 |
| DAFTAR PUSTAKA..... | 65 |
| LAMPIRAN A DATA PENDUKUNG..... | 68 |

DAFTAR TABEL

| | |
|---|----|
| Tabel 4.1 Data Mentah Transaksi Pelanggan | 22 |
| Tabel 4.2 Pergantian dan Pemilihan Variabel | 24 |
| Tabel 4.3 Variabel Baru dari Transaksi Pelanggan | 24 |
| Tabel 4.4 Instrumen dari Metode SUS | 28 |
| Tabel 5.1 Variabel Terpilih untuk Proses <i>Clustering</i> | 29 |
| Tabel 6.1 Pemetaan Nilai Variabel <i>Length</i> | 38 |
| Tabel 6.2 Pemetaan Nilai Variabel <i>Recency</i> | 38 |
| Tabel 6.3 Pemetaan Nilai Variabel <i>Frequency</i> | 38 |
| Tabel 6.4 Pemetaan Nilai Variabel <i>Monetary</i> | 39 |
| Tabel 6.5 Nilai Hasil Transformasi Data | 39 |
| Tabel 6.6 Data Hasil Proses Penghapusan <i>Outliers</i> | 42 |
| Tabel 6.7 Hasil Nilai Normalisasi Data Terhadap Variabel <i>Length</i> | 42 |
| Tabel 6.8 Hasil Nilai Normalisasi Data Terhadap Variabel <i>Recency</i> | 43 |
| Tabel 6.9 Hasil Pengurangan Normalisasi <i>Recency</i> | 43 |
| Tabel 6.10 Hasil Nilai Normalisasi Data Terhadap Variabel <i>Frequency</i> | 43 |
| Tabel 6.11 Hasil Nilai Normalisasi Data Terhadap Variabel <i>Monetary</i> | 43 |
| Tabel 6.12 Tabel Penggabungan dari Normalisasi LRFM | 44 |
| Tabel 6.13 Hasil 2 <i>Cluster</i> | 46 |
| Tabel 6.14 Hasil 3 <i>Cluster</i> | 48 |
| Tabel 6.15 Nilai Hasil dari <i>DB Index</i> | 49 |
| Tabel 6.16 Nilai Hasil dari <i>Connectivity, Dunn Index, dan Silhouette Width</i> | 49 |
| Tabel 6.17 Verifikasi <i>Euclidean Distance 2 Cluster</i> | 50 |
| Tabel 6.18 Verifikasi <i>Euclidean Distance 3 Cluster</i> | 51 |
| Tabel 6.19 Analisis <i>Cluster 1</i> | 52 |
| Tabel 6.20 Analisis <i>Cluster 2</i> | 53 |
| Tabel 6.21 Analisis <i>Cluster 1</i> | 54 |
| Tabel 6.22 Analisis <i>Cluster 2</i> | 54 |
| Tabel 6.23 Analisis <i>Cluster 3</i> | 55 |
| Tabel 6.24 Perhitungan Hasil <i>System Usability Scale (SUS)</i> | 62 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 2.1 Proses Bisnis Pengadaan Produk..... | 8 |
| Gambar 2.2 Tahapan KDD..... | 10 |
| Gambar 2.3 Grafik metode <i>elbow</i> | 12 |
| Gambar 3.1 Diagram Alur Penelitian | 18 |
| Gambar 3.2 Proses <i>Clustering</i> | 20 |
| Gambar 4.1 Desain Arsitektur..... | 26 |
| Gambar 4.2 Halaman Utama <i>Dashboard</i> | 27 |
| Gambar 4.3 Halaman <i>Cluster</i> | 27 |
| Gambar 4.4 Halaman Data..... | 28 |
| Gambar 5.1 Halaman Utama <i>Dashboard</i> | 36 |
| Gambar 5.2 Halaman <i>Cluster Dashboard</i> | 36 |
| Gambar 5.3 Halaman Data <i>Dashboard</i> | 37 |
| Gambar 6.1 Histogram Nilai Variabel <i>Length</i> Sebelum dan Sesudah Transformasi Data | 40 |
| Gambar 6.2 Histogram Nilai Variabel <i>Recency</i> Sebelum dan Sesudah Transformasi Data | 40 |
| Gambar 6.3 Histogram Nilai Variabel <i>Frequency</i> Sebelum dan Sesudah Transformasi Data..... | 40 |
| Gambar 6.4 Histogram Nilai Variabel <i>Monetary</i> Sebelum dan Sesudah Transformasi Data..... | 41 |
| Gambar 6.5 Histogram Perbandingan Ada dan Tidaknya <i>Outliers</i> pada Variabel <i>Monetary</i> | 41 |
| Gambar 6.6 Hasil <i>Elbow Method</i> | 44 |
| Gambar 6.7 Hasil 2 <i>Cluster</i> | 46 |
| Gambar 6.8 Hasil 3 <i>Cluster</i> | 48 |
| Gambar 6.9 Persentase Total Setiap Tipe Sepatu..... | 57 |
| Gambar 6.10 Diagram Batang Penjualan Tipe Sepatu Setiap Bulan..... | 57 |
| Gambar 6.11 Grafik Penjualan Setiap Bulan | 58 |
| Gambar 6.12 Presentase 2 <i>Cluster</i> | 58 |
| Gambar 6.13 Diagram Batang Jumlah Transaksi Setiap Tipe Sepatu Berdasarkan <i>Cluster</i> | 59 |
| Gambar 6.14 <i>Scatterplot 3D</i> Variabel <i>Length</i> , <i>Recency</i> , dan <i>Monetary</i> | 59 |

Gambar 6.15 Diagram Batang Perbandingan 2 Cluster Berdasarkan *Frequency* . 60
Gambar 6.16 Tabel Pelanggan Terbaik Berdasarkan *Frequency* 60
Gambar 6.17 Tabel Pelanggan Terbaik Berdasarkan *Monetary* 61
Gambar 6.18 Kesimpulan 61
Gambar 6.19 Penentuan Hasil 62



DAFTAR SCRIPT

| | |
|---|----|
| Script 5.1 Perubahan Data Menjadi Nilai LRFM | 30 |
| Script 5.2 Transformasi Data dengan Metode <i>Logarithmic</i> | 30 |
| Script 5.3 Penghapusan <i>Outlier</i> dengan <i>Script</i> Klodian Dhana (2016)..... | 32 |
| Script 5.4 Normalisasi Data dengan Metode <i>Min-Max</i> | 32 |
| Script 5.5 Penentuan Jumlah <i>Cluster</i> dengan Metode <i>Elbow</i> | 33 |
| Script 5.6 <i>Clustering</i> dengan Metode <i>Fuzzy C-Means</i> | 33 |
| Script 5.7 Uji Validitas <i>Cluster</i> dengan <i>Modified Partition Coefficient</i> (MPC)..... | 34 |
| Script 5.8 Uji Validitas <i>Cluster</i> dengan <i>DB Index</i> | 34 |
| Script 5.9 Uji Validitas <i>Cluster</i> dengan <i>CValid</i> | 34 |
| Script 5.10 Verifikasi dengan <i>Euclidean Distance</i> | 35 |
| Script 6.1 Validitas MPC 2 <i>Cluster</i> dan Hasilnya..... | 48 |
| Script 6.2 Validitas MPC 3 <i>Cluster</i> dan Hasilnya..... | 48 |
| Script 6.3 Verifikasi 2 <i>Cluster</i> Hasil Pengecekan dengan <i>Fuzzy C-Means</i> | 50 |
| Script 6.4 Verifikasi 3 <i>Cluster</i> Hasil Pengecekan dengan <i>Fuzzy C-Means</i> | 51 |

DAFTAR LAMPIRAN

| | |
|---|----|
| LAMPIRAN A DATA PENDUKUNG | 68 |
| A.1 Potongan Data LRFM Pelanggan | 68 |
| A.2 Uji <i>Usability</i> dengan <i>System Usability Scale (SUS)</i> | 71 |
| A.3 <i>Scatterplot Matrices</i> Dari Data Awal LRFM | 73 |



BAB 1 PENDAHULUAN

1.1 Latar belakang

Pada zaman modern ini, perkembangan teknologi yang semakin pesat merupakan peluang dari sebuah perusahaan atau organisasi guna mendukung kebutuhan dengan menggunakan sistem informasi untuk dapat bersaing dengan kompetitornya. Pesatnya pertumbuhan ekonomi yang disertai tingginya tingkat persaingan dalam dunia bisnis membuat perusahaan semakin sulit untuk meningkatkan jumlah konsumen (Suwitho, 2014). Dengan adanya kondisi tersebut, setiap perusahaan bersaing untuk memberikan layanan terbaik kepada pelanggan. CRM atau *Customer Relationship Management* merupakan salah satu penerapan strategi guna meningkatkan nilai pelanggan yang dapat memaksimalkan pendapatan perusahaan serta mampu bersaing dengan kompetitor.

Customer Relationship Management (CRM) merupakan strategi bisnis yang berfokus pada pemahaman, antisipasi serta pengelolaan kebutuhan pelanggan. Dengan menerapkan strategi bisnis yang tepat, kegiatan akan berjalan lebih efektif dan efisien. Salah satu cara strategi yang dapat digunakan untuk mengelola pelanggan adalah dengan memahami serta memberikan perlakuan yang berbeda-beda sesuai dengan karakteristik pelanggan atau bisa disebut dengan segmentasi pelanggan. Segmentasi pelanggan dapat dilakukan untuk mengetahui perilaku dari pelanggan guna menyusun strategi pemasaran dengan tepat secara efektif dan efisien bagi perusahaan. Menurut Francis Buttler (2004:369), CRM masuk ke dalam tiga kegiatan manajemen yaitu, mendapatkan pelanggan baru, menguasai dan mempertahankan pelanggan yang ada, mengembangkan nilai-nilai pelanggan. Disisi lain, terdapat analisis model LRFM yang dapat membantu implementasi CRM yang berfokus pada karakteristik pelanggan.

Model LRFM (*Length, Recency, Frequency, Monetary*) merupakan modifikasi dari model RFM yang ditambahkan variabel *Length* untuk melihat lama atau tidaknya pelanggan dengan perusahaan. Penerapan model LRFM dapat digunakan sebagai salah satu cara untuk memetakan pelanggan dengan cara menyeleksi atribut-atribut yang ada sesuai dengan aturan model. Atribut-atribut yang digunakan dalam model LRFM adalah interval waktu pembelian dari awal hingga akhir, waktu transaksi terakhir, jumlah frekuensi transaksi, jumlah nominal transaksi. Atribut yang telah diseleksi dapat dimasukkan kedalam metode-metode data mining untuk mendapatkan kelompok data yang diinginkan, seperti menggunakan metode *clustering* yang digunakan untuk mendapatkan segmen. Dalam mengimplementasikan model LRFM, digunakan data pelanggan dari objek penelitian, yaitu Ride Inc.

Ride Inc merupakan bisnis yang bergerak pada bidang *fashion*, khususnya sepatu dengan bahan dasar kulit untuk laki-laki dan wanita. Ride Inc telah berdiri sejak tahun 2011 di Kota Malang yang dikelola oleh Zacharia Lukkyta saat duduk ditingkat akhir perkuliahan. Hasil dari wawancara dengan *owner* Ride Inc, Zacharia

Lukkyta, menyatakan bahwa terdapat masalah dalam upaya mempertahankan pelanggan yang sudah pernah melakukan transaksi. Ride Inc. berfokus pada penjualan sepatu dimana tidak semua orang akan membeli sepatu dalam kurun waktu yang cepat. Selain itu, catatan transaksi dari pembelian pelanggan terdapat dalam data transaksi. Pemanfaatan data transaksi hanya digunakan sebagai alat hitung pendapatan dari organisasi setiap bulannya. Analisis segmentasi pelanggan merupakan upaya dalam membantu menentukan pelanggan yang memiliki nilai dengan menemukan karakteristik pelanggan yang nantinya dapat membantu bidang pemasaran dalam menentukan strategi bisnis yang dapat membantu proses pemasaran.

Clustering merupakan metode data mining yang digunakan untuk mengelompokkan data dengan karakteristik yang sama ke suatu wilayah yang sama dan dengan karakteristik berbeda ke wilayah yang lain. Data yang dikelompokkan dalam *clustering* adalah data yang belum teridentifikasi klasifikasinya dari tiap-tiap kelompok. Dalam *clustering*, terdapat beberapa metode seperti, *partitioning method*, *hierarcial method*, *density-based method*, *grid-based method*, *model-based method*, *clustering high-dimensional data*, dan *constraint-based clustering*. Metode *partitioning data* merupakan metode yang dapat mengelompokkan data secara langsung dengan berbasiskan jarak dan terdapat minimal satu objek pada *cluster*. Metode *partition* bekerja secara iteratif terhadap data yang tersebar dan akan bergerak menuju *cluster* yang sesuai dengan kesamaan kriteria pada setiap data. Namun, metode ini memerlukan proses penentuan jumlah *cluster* yang akan dibuat pada tahap awal. Terdapat metode penerapan dalam *clustering* yaitu algoritma *K-Means* dan *Fuzzy C-Means*.

Fuzzy C-Means diperkenalkan oleh Bezdek (1981) yang merupakan pengembangan dari metode *Hard C-Means clustering*. *Fuzzy C-Means* atau biasa disebut FCM merupakan metode data *clustering* yang termasuk dalam *partitioning method* dengan berlandaskan teori logika fuzzy. Terdapat perhitungan dari kondisi awal hingga kondisi akhir secara iteratif yang akan memperbaiki pusat *cluster*. Metode *Fuzzy C-Means* digunakan karena memiliki kelebihan dalam penempatan data secara tepat dibandingkan dengan metode *clustering* yang lain dengan proses yang mampu menghitung derajat keanggotaan data, sehingga setiap data memiliki kejelasan nilai dalam setiap *cluster*. Hasil dari pemetaan data mentah transaksi yang terdapat dalam Lampiran A.3 menunjukkan bahwa data memiliki letak yang berdekatan sehingga penggunaan metode *Fuzzy C-Means* sesuai dengan kondisi tersebut sesuai dengan kelebihanannya. Penelitian ini dilakukan dengan tujuan dapat mendapatkan segmen pelanggan dengan karakteristik berdasarkan LRFM.

Dashboard merupakan tampilan visual dari informasi paling penting yang diperlukan untuk mencapai satu atau beberapa tujuan yang diatur pada satu layar sehingga informasi dapat ditampilkan dengan jelas (Few, 2004). Hasil data dari *Fuzzy C-Means clustering* akan digunakan sebagai metode untuk pengelompokkan pelanggan. *Dashboard* digunakan untuk memvisualisasikan hasil dari pengelompokkan pelanggan. *Dashboard* tersebut akan diserahkan dan digunakan oleh pihak Ride Inc. Untuk menentukan kelayakan dari *dashboard*, peneliti

melakukan *usability testing* terhadap pihak internal Ride Inc. menggunakan metode *System Usability Scale* (SUS). Metode tersebut diperkenalkan oleh Brooke (1986) dan memiliki beberapa kondisi dalam implementasinya.

Dari uraian diatas, peneliti akan melakukan pengelompokan pelanggan menggunakan metode *clustering* berdasarkan data transaksi dari Ride Inc. Metode penyelesaian yang digunakan dalam *clustering* adalah *Fuzzy C-Means*. Implementasi *Fuzzy C-Means clustering* menggunakan metode *Customer Relationship Management* (CRM), yaitu LRFM model sebagai variabel dalam pengelompokan data pelanggan. Hasil yang diberikan adalah visualisasi hasil *clustering* atau berupa *dashboard* serta melakukan *usability testing* kepada pihak internal Ride Inc. untuk mengetahui kelayakan dari *dashboard* tersebut. Oleh karena itu penulis mengangkat judul “**Segmentasi Pelanggan Menggunakan Metode Fuzzy C-Means Clustering Berdasarkan LRFM Model Pada Toko Sepatu (Studi Kasus: Ride Inc Kota Malang)**”.

1.2 Rumusan masalah

Dengan penjelasan latar belakang yang ada, maka dapat disusun rumusan masalah sebagai berikut:

1. Bagaimana LRFM model dapat merepresentasikan data pelanggan dengan data transaksi Ride Inc?
2. Bagaimana implementasi algoritma *Fuzzy C-Means clustering* untuk proses segmentasi pelanggan berdasarkan LRFM model pada Ride Inc?
3. Bagaimana pengujian *cluster* dapat merepresentasikan *cluster* yang tepat?
4. Bagaimana visualisasi *dashboard* berdasarkan segmentasi pelanggan dapat membantu pengambilan keputusan pada Ride Inc?

1.3 Tujuan

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Merepresentasikan data pelanggan berdasarkan LRFM model dengan data transaksi yang terdapat pada Ride Inc.
2. Melakukan proses segmentasi pelanggan menggunakan algoritma *Fuzzy C-Means clustering* berdasarkan LRFM model pada Ride Inc.
3. Melakukan pengujian performa *cluster* untuk mengukur tingkat ketepatan hasil *cluster*.
4. Menghasilkan visualisasi *dashboard* berdasarkan segmentasi pelanggan dan mampu membantu Ride Inc dalam mengambil keputusan.

1.4 Manfaat

Manfaat dari penelitian ini adalah:

1. Membantu perusahaan atau organisasi untuk mengetahui kelompok pelanggan yang ada berdasarkan hasil dari proses *clustering*.

2. Memberikan informasi terkait segmentasi pelanggan dengan hasil visualisasi *dashboard*.
3. Menyajikan informasi yang dapat digunakan perusahaan dalam pengambilan keputusan strategis guna meningkatkan kualitas layanan, seperti:
 - a. Rekomendasi promo atau layanan bagi pelanggan berdasarkan kelompok pelanggan.
 - b. Prioritas layanan terhadap pelanggan yang memiliki potensi terhadap perusahaan.
 - c. Menambah wawasan dan pengetahuan terkait masalah yang berkaitan dengan *Customer Relationship Management* (CRM) dan segmentasi pelanggan.

1.5 Batasan masalah

Penelitian ini memiliki batasan masalah yang berfungsi untuk mengarahkan tujuan agar sesuai dengan rumusan masalah, diantaranya adalah:

1. Menggunakan data transaksi pelanggan Ride Inc dari bulan Juli 2017 hingga Maret 2018.
2. Penentuan variabel untuk proses *clustering* berdasarkan LRFM model.
3. Penggunaan metode *Fuzzy C-Means Clustering* untuk proses *clustering*.
4. Pengujian validitas terhadap hasil *clustering* menggunakan *Modified Partition Coefficient* (MPC), *Davies Bouldin's Index*, dan *packages CValid*.
5. Penggunaan bahasa pemrograman R dalam pengolahan data.
6. Penggunaan perangkat lunak Power BI untuk menghasilkan visualisasi *dashboard*.

1.6 Sistematika pembahasan

Sistematika pembahasan dalam menyusun laporan skripsi ini adalah:

BAB 1 : PENDAHULUAN

Bab 1 memuat hal yang berkaitan dengan latar belakang, rumusan masalah, tujuan, batasan masalah, manfaat, dan sistematika pembahasan.

BAB 2 : LANDASAN KEPUSTAKAAN

Bab 2 berisi uraian mengenai penelitian-penelitian terdahulu dan teori-teori mengenai pengimplementasian metode *clustering* yang digunakan sebagai referensi dalam penelitian.

BAB 3 : METODOLOGI PENELITIAN

Bab 3 memuat langkah-langkah yang digunakan saat melakukan penelitian.

BAB 4 : PERANCANGAN

Bab 4 memuat hal-hal yang berkaitan dengan perancangan, seperti.

BAB 5 : IMPLEMENTASI

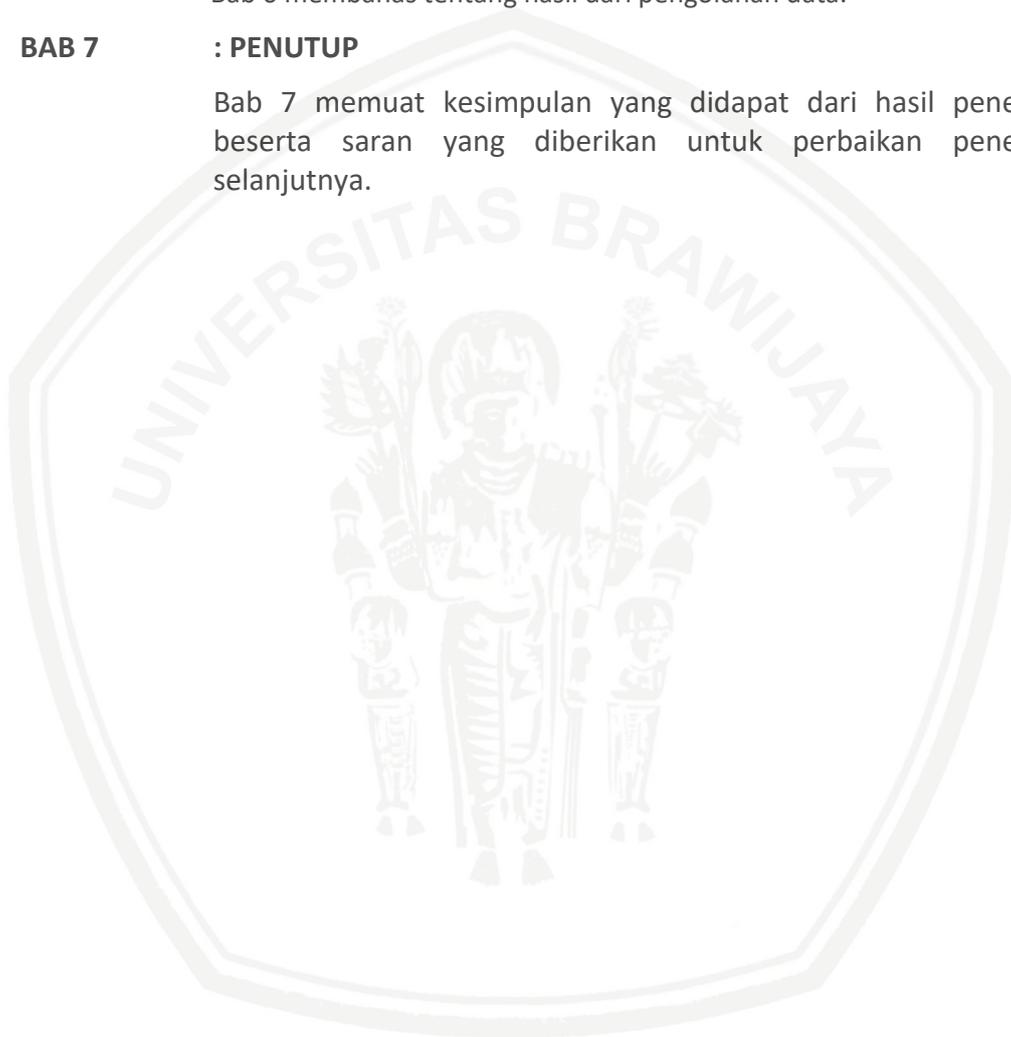
Bab 5 membahas tentang penerapan dari hasil rancangan.

BAB 6 : ANALISIS DAN PEMBAHASAN

Bab 6 membahas tentang hasil dari pengolahan data.

BAB 7 : PENUTUP

Bab 7 memuat kesimpulan yang didapat dari hasil penelitian beserta saran yang diberikan untuk perbaikan penelitian selanjutnya.



BAB 2 LANDASAN KEPUSTAKAAN

2.1 Kajian Pustaka

Penelitian pertama yang dilakukan oleh Astria dan Suprayogi (2017) bertujuan untuk mengembangkan strategi pemasaran pada CV. Mataram Jaya Bawen dengan mengelola data pelanggan. Pengelolaan data pelanggan dilakukan untuk mendapatkan kelompok pelanggan dengan menggunakan metode *Fuzzy C-Means* dan RFM model sebagai pemilihan atribut yang akan digunakan dalam pengolahan data. Data yang digunakan sebesar 709 transaksi dari 75 pelanggan. Hasil yang didapat setelah melakukan penelitian adalah tingkat validitas tiga *cluster* (*golden, silver, bronze*) menggunakan *Fuzzy C-Means* adalah 0.596277 yang berarti tingkat akurasi cukup baik.

Penelitian terdahulu yang dilakukan oleh Selviana (2016) bertujuan untuk menghasilkan strategi pembelajaran untuk meningkatkan motivasi belajar mahasiswa dengan menggunakan metode *clustering*. Jurnal tersebut menganalisis serta membandingkan hasil dari penggunaan *K-Means* dan *Fuzzy C-Means* dengan dibantu model ARCS (*attention, relevance, confidence, satisfaction*). Hasil analisa dari penelitian yang terdapat tiga *cluster* adalah tingkat validitas dari *Fuzzy C-Means* sebesar 0.5098 dan *K-Means* sebesar 0.2896. Hasil tersebut menyatakan bahwa penggunaan metode *Fuzzy C-Means* mendapatkan hasil yang valid dibandingkan *K-Means*.

Penelitian lain yang dilakukan oleh Ansari dan Riasi (2016) bertujuan untuk mendapatkan kelompok pelanggan dari industri baja dengan mengidentifikasi karakteristik dan kebutuhan pelanggan untuk meningkatkan keunggulan dalam persaingan. Penelitian tersebut menggabungkan *Fuzzy C-Means clustering* dengan *Genetic Algorithms* untuk mendapatkan keakuratan *cluster*. Jurnal tersebut melakukan pengelompokan variabel dengan model LRFM sebelum mengimplementasikan kedua algoritma tersebut. Setelah dilakukan penelitian, hasil yang didapat adalah *cluster* terbagi menjadi dua, *cluster* pertama merupakan pelanggan yang memiliki hubungan yang tinggi, hubungan perdagangan dan memiliki frekuensi transaksi yang tinggi namun nilai moneter yang lebih rendah, dan *cluster* kedua merupakan pelanggan yang memiliki nilai moneter yang tinggi. Di sisi lain, hasil dari MSE dari penggunaan *Fuzzy C-Means* dan *Genetic Algorithms* lebih rendah dengan nilai 0.0152 dibanding hanya menggunakan *Fuzzy C-Means* yang berarti penggunaan kombinasi algoritma tersebut menghasilkan keakuratan *cluster* yang lebih tinggi. Dan hasil lain membuktikan *run time* dari kombinasi algoritma lebih tinggi dengan nilai 0.8474 yang berarti waktu yang dibutuhkan untuk mengolah data lebih lama.

Penelitian selanjutnya yang dilakukan oleh Karomi (2016) memiliki tujuan untuk membangun model pelanggan dari klinik kecantikan sebagai target promosi yang potensial dan strategi pemasaran yang tepat. Penelitian tersebut didukung dengan *Fuzzy C-Means* berdasarkan model RFM. Dataset penjualan yang digunakan hanya yang melakukan transaksi dari bulan November 2013 hingga

Desember 2013 sebesar 16.595. Dari jurnal tersebut mendapatkan 8 tipe kelompok pelanggan yang memiliki satu data *outlier* dan akan diuji validitasnya menggunakan *Sum of Squared Error (SSE)* dan *Silhouette Coefficient (SC)*. Hasil dari pengujian menyatakan nilai SSE sebesar 63,339 dan SC sebesar 0,964384. Hasil tersebut meningkat setelah *outlier* dihilangkan dari dataset.

Referensi lain yang diambil dari jurnal penelitian mengenai segmentasi pelanggan oleh Yuliari (2015) bertujuan untuk menemukan karakteristik pelanggan yang memiliki potensi serta menemukan tingkat loyalitas pelanggan. Dalam penelitiannya, didapatkan empat karakteristik pelanggan menggunakan RFM, yaitu *superstar*, *golden customer*, *typical customer*, dan *occasional customer*. Setelah dilakukan implementasi ke dalam *Fuzzy C-Means*, dilakukan uji validitas menggunakan *Modified Partition Coefficient (MPC)*. Hasil yang didapat adalah terdapat dua *cluster* dominan yaitu *superstar* dan *occasional* serta terdapat dua *cluster* terbaik setelah diuji validitasnya menggunakan MPC.

Referensi terakhir dari Afrin, et. al (2015) bertujuan untuk mendapatkan segmentasi pelanggan terbaik dengan membandingkan dua metode. Metode yang digunakan adalah Principal Component Analysis (PCA) berbasis K-Means dan PCA berbasis Fuzzy C-Means. Hasil yang didapat adalah penggunaan PCA berbasis Fuzzy C-Means mendapat kinerja lebih baik dibandingkan dengan PCA berbasis K-Means karena eksekusi waktu pada Fuzzy C-Means lebih rendah pada pengujian tiga, lima dan tujuh *cluster*.

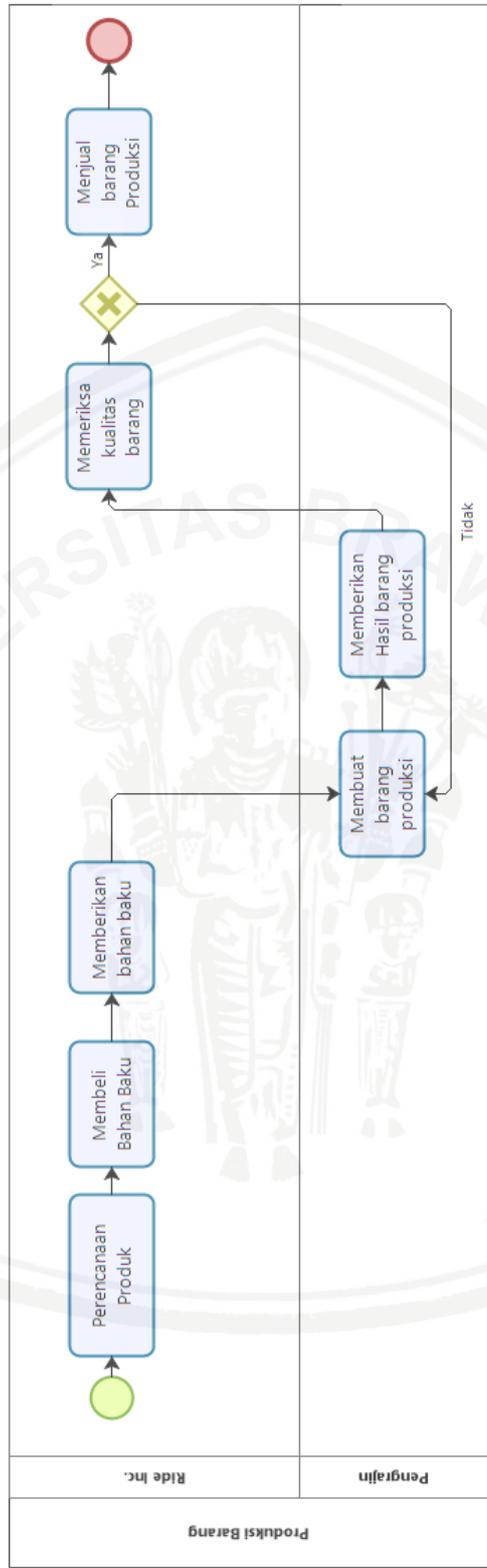
Dari beberapa referensi jurnal tersebut, penelitian akan membahas metode segmentasi pelanggan dengan metode *Fuzzy C-Means Clustering* yang dianggap dapat menghasilkan *cluster* yang tepat. Penelitian bertujuan untuk mendapatkan karakteristik pelanggan yang memiliki potensi serta dapat membantu penentuan dalam pembelian tipe kulit berdasarkan tipe sepatu. Tetapi, metode *Fuzzy C-Means Clustering* pada skripsi akan dibantu oleh LRFM model sebagai atribut yang akan digunakan dalam pengolahan data untuk membedakan dengan penelitian terdahulu. LRFM model merupakan pengembangan dari RFM model dengan penambahan atribut *Length*.

2.2 Profil Toko Sepatu Ride Inc Kota Malang

Ride Inc merupakan salah satu toko sepatu dengan bahan dasar kulit yang berdiri sejak tahun 2011 di Kota Malang. Dalam upaya menarik perhatian pelanggan, Ride Inc menawarkan berbagai macam model sepatu mulai dari model *casual* hingga formal dengan harga yang lebih terjangkau dibandingkan toko sepatu yang berbahan dasar kulit lainnya. Ride Inc menggunakan *platform* Instagram dan website dalam menawarkan produk yang dimilikinya. Selain itu, Ride Inc menawarkan sepatu dengan cara *made-to-order* atau kustom kepada pelanggan dengan berbagai macam model dan kulit yang tersedia. Namun, Ride Inc tetap menyediakan sepatu *ready-to-wear* atau stok yang bisa langsung dibeli oleh pelanggan secara *online* atau langsung mendatangi toko tanpa menunggu lama. Selain sepatu, Ride Inc juga menyediakan beberapa produk dari *brand* lain seperti *leather goods* dan kaos kaki. Adapun proses yang dilakukan dari



perencanaan produk hingga menjadi sepatu yang layak dijual seperti pada Gambar 2.1.



Powered by bizagi Modeler

Gambar 2.1 Proses Bisnis Pengadaan Produk

Dari Gambar 2.1 dapat diketahui bahwa langkah pertama adalah merencanakan produk yang akan dijual dengan tim. Setelah itu, tim akan membeli bahan baku dari pembuatan sepatu. Bahan baku tersebut diberikan kepada pengrajin yang membuat sepatu. Sepatu yang telah selesai dibuat akan diberikan kepada tim untuk diperiksa kembali kualitas dari produk. Jika produk ada yang tidak sesuai maka akan dikembalikan kepada pengrajin untuk direparasi dan jika produk telah sesuai, maka akan dimasukkan ke dalam penjualan.

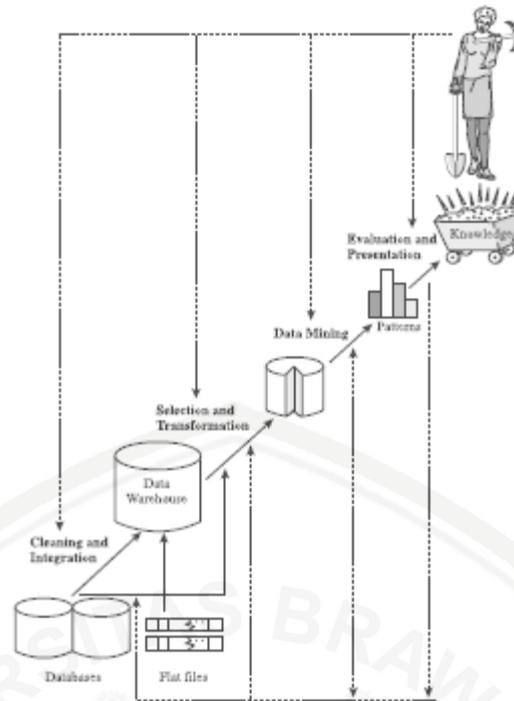
2.3 Customer Relationship Management (CRM)

Menurut Buttle (2009) perusahaan teknologi informasi cenderung menggunakan istilah CRM untuk menggambarkan *software* yang mengotomatisasi fungsi pemasaran, penjualan dan layanan bisnis. Terdapat beberapa definisi dari *Customer Relationship Management* atau Manajemen Hubungan Pelanggan dari bukunya yang berjudul "*Customer Relationship Management Concepts and Technologies*", diantaranya adalah:

1. CRM merupakan pendekatan terpadu untuk mengidentifikasi, mengakuisisi, dan mempertahankan pelanggan yang dapat membantu organisasi memaksimalkan nilai setiap interaksi pelanggan dan mendorong kinerja perusahaan.
2. CRM adalah strategi bisnis yang memaksimalkan keterampilan, pendapatan dan kepuasan pelanggan dengan mengatur dengan mengatur segmen pelanggan, mendorong perilaku yang memuaskan pelanggan dan menerapkan proses *customercentric*.

2.4 Data Mining

Data mining merupakan proses untuk menemukan pengetahuan menarik dari sejumlah data yang besar dan tersimpan didalam database, data warehouses, atau repositori lainnya menurut Han dan Kamber (2006). Definisi lain dari Hand (2001) mengatakan bahwa data mining adalah analisis dari sejumlah data yang besar untuk menemukan hubungan antar data dan meringkas data yang dapat dimengerti dan berguna bagi pemilik data. Data mining merupakan langkah dari proses *Knowledge Discovery from Data* (KDD). Langkah-langkah dalam proses KDD dapat dilihat dalam Gambar 2.2:



Gambar 2.2 Tahapan KDD

Sumber: Han (2006)

Penjelasan dari tahapan KDD diatas, antara lain:

a. *Data Cleaning dan Integration*

Proses ini berfungsi untuk menghilangkan data yang bersifat *noisy* dan tidak konsisten. Selain itu, proses ini melakukan penggabungan data yang memiliki sumber yang sama.

b. *Data Selection dan Data Transformation*

Proses ini berfungsi untuk menyeleksi data yang akan digunakan untuk kebutuhan penelitian dan mentransformasikan data ke dalam bentuk yang sesuai untuk proses data mining.

c. *Data Mining*

Proses ini berfungsi untuk menerapkan metode dari data mining yang akan digunakan untuk menghasilkan *patterns*.

d. *Evaluation dan Presentation*

Proses ini berfungsi untuk mengidentifikasi kelayakan *patterns* yang dihasilkan dan menyajikan hasil kepada pengguna.

Proses *data cleaning*, *data integration*, *data selection* dan *data transformation* termasuk dalam *data preprocessing*, dimana data disiapkan untuk proses *mining*.

2.5 Model LRFM (*Length, Recency, Frequency, Monetary*)

LRFM model merupakan model pengembangan dari RFM model yang merupakan metode untuk menganalisis nilai pelanggan untuk segmentasi pasar. RFM model memiliki tiga atribut yaitu, *Recency* yang didefinisikan sebagai cara melihat pembelian terakhir dari pelanggan dengan hari atau bulan, *Frequency* yang didefinisikan sebagai jumlah pembelian yang dilakukan dalam jangka waktu tertentu dan *Monetary* didefinisikan sebagai jumlah rata-rata uang yang dikeluarkan dalam jangka waktu tertentu. Menurut Wu dkk. (2014) RFM model tidak dapat membedakan pelanggan mana yang memiliki hubungan jangka pendek atau jangka panjang terhadap perusahaan. Oleh karena itu, terdapat pertimbangan dalam penambahan jangka panjang (*Length*), yang didefinisikan sebagai jumlah periode waktu dari pembelian pertama hingga pembelian terakhir. Nilai *Length* didapat dari pengurangan tanggal pembelian akhir dengan pembelian pertama dan hasil berupa total selisih hari. Nilai *Recency* didapat dari pengurangan tanggal penelitian dengan pembelian terakhir dan hasil berupa total selisih hari. Nilai *Frequency* didapat dengan menjumlahkan transaksi yang dilakukan setiap pelanggan. Dan nilai *Monetary* didapat dengan menjumlahkan seluruh nominal pada setiap transaksi dari pelanggan.

2.6 Normalisasi Min-Max

Menurut Han dkk. (2012) normalisasi data berfungsi mencoba menyetarakan bobot pada semua atribut dengan bobot yang sama agar terhindar dari ketergantungan pada setiap data. Pemberian bobot menggunakan metode Min-Max adalah dengan memberi rentang nilai antara 0 hingga 1. Rumus dari normalisasi Min-Max dapat dilihat pada persamaan 2.1.

$$v' = \frac{v - \min_A}{\max_A - \min_A} (\text{newmax}_A - \text{newmin}_A) + \text{newmin}_A \dots (2.1)$$

Dimana:

v' = nilai data yang sudah dinormalisasi

v = nilai data yang belum dinormalisasi

\max_A = nilai maksimum data dari atribut ke-A

\min_A = nilai minimum data dari atribut ke-A

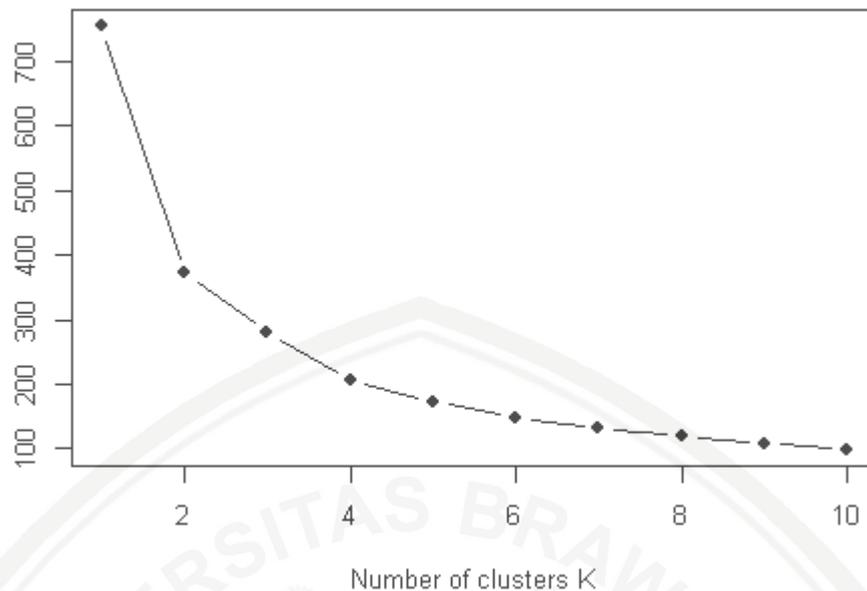
newmax_A = nilai maksimum data baru dari atribut ke-A

newmin_A = nilai minimum data baru dari atribut ke-A

2.7 Metode *Elbow*

Metode *Elbow* menurut Makwana dan Kodinariya (2013) adalah metode visual yang digunakan untuk menentukan jumlah *cluster* yang tepat pada sebuah *dataset* dengan cara kerja tahap awal memulai $k=2$ dan akan terus bertambah setiap step selagi menghitung nilai variabel. Grafik yang dihasilkan akan turun

secara drastis sebelum akhirnya menjadi stabil dan disitulah nilai k yang optimum. Berikut adalah ilustrasi dari grafik metode elbow dalam Gambar 2.3:



Gambar 2.3 Grafik metode *elbow*

Sumber: Savitri (2018)

Selain itu, untuk mendapatkan presentase perhitungan, diperlukan perhitungan SSE (*Sum of Square Error*) pada masing-masing *cluster*. Adapun algoritma dari metode *elbow* adalah sebagai berikut:

- Mulai
- Inisialisasi awal nilai K
- Menaikkan nilai K
- Menghitung hasil SSE dari setiap nilai K
- Melihat hasil SSE dari nilai K yang turun secara drastis
- Tetapkan nilai K yang berbentuk siku
- Selesai

2.8 Clustering

2.8.1 Definisi Clustering

Clustering merupakan salah satu metode dalam data mining. Menurut Han dan Kamber (2006) *clustering* merupakan cara mengidentifikasi daerah padat dan jarang dalam suatu objek yang bertujuan untuk menemukan pola distribusi secara keseluruhan dan korelasi yang menarik antar atribut. Analisis *cluster* telah banyak digunakan dalam berbagai hal, contohnya adalah dalam bisnis, *clustering* dapat membantu bidang pemasaran dalam menemukan kelompok pelanggan berdasarkan pola pembelian dan dalam bidang biologi. *Clustering* dapat digunakan

untuk mengkategorikan gen dengan fungsi serupa. Oleh karena itu, *clustering* juga disebut sebagai segmentasi data karena dapat mengelompokkan partisi dari data yang besar dan ditetapkan ke dalam kelompok sesuai kemiripannya. Sebagai salah satu metode dari data mining, analisis *cluster* dapat digunakan sebagai alat untuk mendapatkan wawasan mengenai distribusi data guna mengamati masing-masing cluster dan berfokus pada kelompok tertentu untuk diamati lebih lanjut.

2.8.2 Metode *Clustering*

Terdapat beberapa metode *clustering* menurut Han dan Kamber (2006), diantaranya adalah:

a. *Partitioning Method*

Partitioning method dapat mengklasifikasikan data ke dalam kelompok secara bersama-sama dengan memenuhi dua syarat yaitu, masing-masing kelompok mengandung setidaknya satu objek, dan setiap objek harus dimiliki oleh satu kelompok.

b. *Hierarcial Method*

Hierarcial method dapat menciptakan dekomposisi hierarkis dari kumpulan objek data. Terdapat dua pendekatan untuk meningkatkan kualitas pengelompokan hierarkis yaitu, melakukan analisis terhadap keterkaitan objek pada partisi data, dan mengintegrasikan aglomerasi hierarkis dan pendekatan lainnya dengan terlebih dahulu menggunakan algoritma aglomeratif hierarkis.

c. *Density-based Method*

Terdapat beberapa metode dalam *density-based method* diantaranya, OPTICS yang merupakan metode yang dapat menumbuhkan kelompok berdasarkan analisis konektivitas *density-based*, DENCLUE yang merupakan metode pengelompokan objek berdasarkan distribusi nilai fungsi *density*, dan DBSCAN.

d. *Grid-based Method*

Grid-based method merupakan metode yang mengkuantifikasi ruang objek menjadi sejumlah sel yang akan menghasilkan sebuah struktur *grid*. Metode ini tidak tergantung pada jumlah objek data dan hanya bergantung pada jumlah sel setiap dimensi di ruang yang terkuantisasi.

e. *Model-based Method*

Model-based method merupakan metode yang dapat menghipotesiskan masing-masing kelompok dan menemukan data sesuai dengan model serta dapat menemukan *cluster* dengan membangun fungsi kepadatan yang mencerminkan distribusi parsial dari titik data.

f. *Clustering High-dimensional Data*

Clustering High-dimensional Data dapat menghasilkan dimensi yang hasilnya mungkin tidak relevan. CLIQUE dan PROCLUS merupakan metode pengelompokan yang dapat mencari *cluster* pada subruang data dan tidak mencari dengan keseluruhan data.

g. *Constraint-based Clustering*

Constraint-based clustering merupakan pendekatan *clustering* yang menggabungkan batasan yang ditentukan oleh pengguna dan mampu mengungkapkan ekspektasi pengguna dengan menggambarkan hasil pengelompokan yang diinginkan.

2.9 Fuzzy C-Means Clustering

K-Means clustering merupakan metode *clustering* yang sederhana dan sering digunakan karena mampu mengelompokkan data dalam jumlah banyak dengan waktu yang relative cepat dan efisien (Santosa, 2007). Selain *K-Means clustering*, terdapat metode lain dalam *clustering* yaitu, *Fuzzy C-Means clustering*. *Fuzzy C-Means clustering* merupakan algoritma pengelompokan *unsupervised* yang diterapkan untuk masalah yang terkait dengan analisis, pengelompokan, dan penggolongan desain. Selain itu, *Fuzzy C-Means clustering* merupakan metode pengelompokan objek menjadi beberapa kelas yang tidak dipaksa sepenuhnya untuk dimiliki oleh satu kelas, melainkan dengan memiliki derajat keanggotaan antara 0 hingga 1 yang menunjukkan keanggotaan objek (Suganya dan Shanthi, 2012). Algoritma dari *Fuzzy C-Means clustering* antara lain:

1. Inisialisasi $U = [u_{ij}]$ matrix,
2. Menghitung nilai tengah dari vektor $C^{(k)} = [c_j]$ dengan $U^{(k)}$

$$c_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m},$$

3. Mengubah $U^{(k)}$ menjadi $U^{(k+1)}$,

$$d_{ij} = \sqrt{\sum_{i=1}^n (x_i - c_i)^2}$$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}}$$

4. *if* $||U(k+1) - U(k)|| < \epsilon$ *then STOP* dan kembali ke langkah nomor 2.

Dimana:

m merupakan nilai asli yang lebih dari 1,

u_{ij} merupakan derajat keanggotaan dari x_i dalam *cluster* j , dan

c_j merupakan nilai tengah dari *cluster*.

2.10 Modified Partition Coefficient (MPC)

Modified Partition Coefficient (MPC) merupakan fungsi validitas yang digunakan pada *Fuzzy C-Means* untuk menguji validitas serta mengevaluasi kinerja dari hasil *clustering*. MPC merupakan penyempurnaan dari *Partition Coefficient* (PC) yang dirancang oleh Bezdek (1974) untuk mengukur tingkat *overlap* antar *cluster* yang dapat dilihat pada persamaan 2.2.

$$PC(c) = \frac{1}{N} \sum_{i=1}^c \sum_j^N (\mu_{ij}^2) \dots (2.2)$$

Dimana:

μ_{ij} = nilai keanggotaan j dalam *cluster* i .

Modified Partition Coefficient (MPC) diusulkan oleh Dave (1996) untuk mengurangi kecenderungan perubahan nilai yang monoton dengan rentang nilai 0 hingga 1 dan mencari nilai maksimal. Rumus dari MPC dapat dilihat pada persamaan 2.3.

$$MPC(c) = 1 - \frac{c}{c-1} (1 - PC) \dots (2.3)$$

Dimana:

c = banyak kelompok, dan

PC = indeks *partition coefficient* (PC).

2.11 Davies Bouldin's Index

Davies Bouldin's Index (DB Index) adalah salah satu indeks validitas untuk mengevaluasi nilai hasil *clustering*. Menurut Davies (1979) DB Index dapat mengidentifikasi *cluster* yang sesuai dilihat dari jarak satu sama lain dan *cluster* terbaik memiliki nilai index yang terkecil. Rumus dari index DB dapat dilihat pada persamaan 2.4.

$$DU_k = \frac{1}{k} \sum_{i=1}^k \max_{j=1, \dots, k, i \neq j} \left\{ \frac{\text{diam}(c_i) + \text{diam}(c_j)}{\|c_i - c_j\|} \right\} \dots (2.4)$$

Variabel *diam* atau diameter dari *cluster* didefinisikan pada persamaan 2.5 sebagai:

$$\text{diam}(c_i) = \left(\frac{1}{n_i} \sum_{x \in c_i} \|x - z_i\|^2 \right)^{\frac{1}{2}} \dots (2.5)$$

Dimana:

n_i = jumlah *cluster*, dan

z_i = *centroid* dari *cluster*.

2.12 Connectivity

Menurut Handl dkk. (2005) *connectivity* berfungsi untuk mengukur seberapa jauh hubungan dari pengamatan pada kelompok dengan kelompok terdekat. Rumus dari *connectivity* dapat dilihat pada persamaan 2.6 (Brock dkk., 2011).

$$Conn(\mathcal{C}) = \sum_{i=1}^N \sum_{j=1}^L X_{i,nn_{i(j)}} \dots (2.6)$$

Dimana $X_{i,nn_{i(j)}}$ merupakan jarak dari tetangga terdekat dari observasi i dan akan bernilai 0 jika i dan $nn_{i(j)}$ berada dalam *cluster* yang sama, dan bernilai 1 jika sebaliknya serta memiliki nilai yang terkecil.

2.13 Dunn Index

Dunn Index merupakan index yang diusulkan oleh Dunn (1974) yang berfungsi untuk mengidentifikasi *cluster* yang terpisah serta memaksimalkan jarak antar *cluster* sekaligus meminimalkan jarak isi dari setiap kelompok. *Dunn Index* didefinisikan pada persamaan 2.7 (Brock dkk., 2011).

$$D(\mathcal{C}) = \frac{\min_{C_k, C_l \in \mathcal{C}, C_k \neq C_l} \left(\min_{i \in C_k, j \in C_l} dist(i, j) \right)}{\max_{C_m \in \mathcal{C}} diam(C_m)} \dots (2.7)$$

Dimana $diam(C_m)$ merupakan jarak maksimal dalam penelitian *cluster* C_m . *Dunn Index* memiliki nilai dengan range 0 hingga 1 dan menggunakan nilai index yang terbesar.

2.14 Silhouette Width

Silhouette Width berfungsi untuk menghasilkan evaluasi dari validitas *cluster* dan bisa digunakan untuk menentukan jumlah *cluster* sesuai (Rousseeuw, 1987). Nilai *silhouette* dari *cluster* yang baik adalah yang maksimal atau mendekati 1. Rumus dari *silhouette width* menurut Brock dkk. (2011) dapat dilihat pada persamaan 2.8.

$$S(i) = \frac{b_i - a_i}{\max(b_i, a_i)} \dots (2.8)$$

Dimana:

a_i = jarak rata-rata antara i dengan data observasi dalam satu *cluster*, dan

b_i = jarak rata-rata antara i dengan kelompok terdekat.

2.15 Euclidean Distance

Euclidean Distance jarak antara dua titik yang akan diukur dan merupakan fungsi jarak yang paling umum digunakan. Menurut Loohach dan Garg (2012) jarak

euclidean merupakan rumus pitagoras dan jarak titik 1 dan 2 merupakan panjang segmen garis yang menghubungkannya. Dalam *Euclidean*, jika $a = (a_1, a_2)$ dan $b = (b_1, b_2)$ maka jarak didapat pada persamaan:

$$D(a, b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2} \dots (2.9)$$

2.16 Dashboard

Enterprise Dashboard merupakan alat untuk menampilkan informasi dalam bentuk tabel, laporan serta indikator visual dan informasi tersebut akan ditampilkan secara visual di layar dengan cara yang memudahkan pengguna dalam mengawasi dan mengelola inisiatif bisnisnya (Malik, 2005). Sedangkan Few (2006) menggunakan terminologi *Information Dashboard* yang berarti tampilan visual dari informasi yang dibuat untuk dapat mengawasi kinerja organisasi dan mampu memudahkan pengguna dalam mencapai tujuan tertentu. Selain itu, menurut Few (2004) *Dashboard* merupakan tampilan visual dari informasi paling penting yang diperlukan untuk mencapai satu atau beberapa tujuan yang diatur pada satu layar sehingga informasi dapat ditampilkan dengan jelas.

2.17 System Usability Scale (SUS)

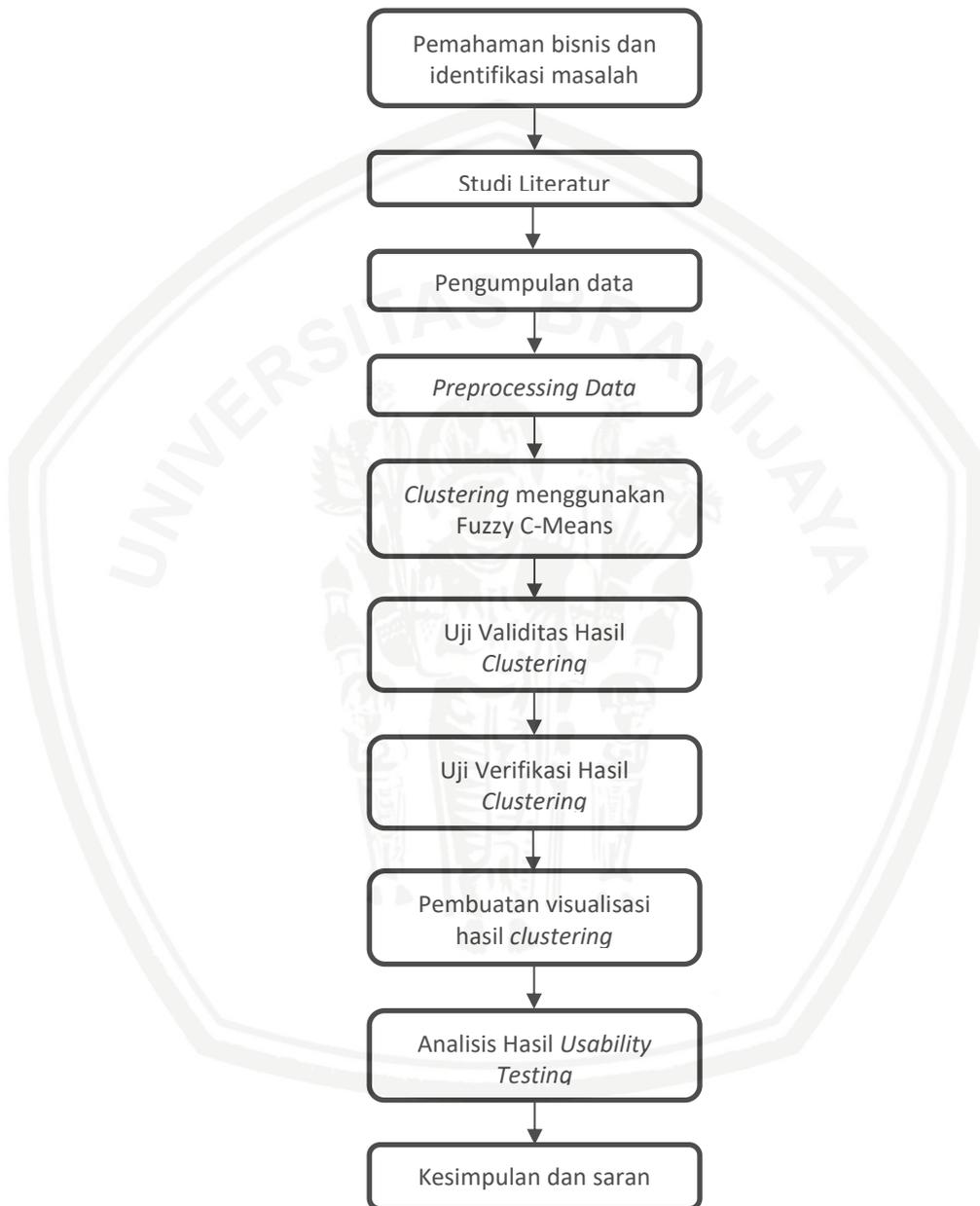
System Usability Scale (SUS) merupakan metode pengujian terhadap pengguna dari tampilan antarmuka sistem berupa kuesioner berisi 10 instrumen yang diperkenalkan oleh Brooke (1986). Penilaian terhadap kuesioner tersebut dimulai dari 1 yang berarti sangat tidak setuju hingga 5 yang berarti sangat setuju dengan kondisi seperti berikut (Sauro, 2011):

1. Pernyataan ganjil merupakan pernyataan positif yang nilainya didapat dengan cara mengurangi nilai dengan 1.
2. Pernyataan genap merupakan pernyataan negatif yang nilainya didapat dengan cara mengurangi 5 dengan nilai.
3. Hasil dari pernyataan genap dan ganjil memiliki rentang 0 hingga 4.
4. Hasil adalah total dari item genap dan item ganjil dikalikan dengan 2.5 dan memiliki hasil nilai dengan rentang 0 hingga 100.

BAB 3 METODOLOGI

3.1 Metodologi Penelitian

Terdapat alur penelitian pada segmentasi pelanggan menggunakan metode *Fuzzy C-Means clustering* berdasarkan LRFM model pada Ride Inc dalam Gambar 3.1.



Gambar 3.1 Diagram Alur Penelitian

3.1.1 Pemahaman Bisnis dan Identifikasi Masalah

Pada tahap ini terdapat dua kegiatan yaitu pemahaman bisnis dan mengidentifikasi masalah. Pemahaman bisnis dilakukan untuk mendapatkan informasi mengenai kebutuhan penelitian. Pemahaman bisnis dilakukan dengan cara wawancara langsung terhadap pihak Ride Inc. Identifikasi masalah dilakukan bersamaan dengan wawancara mengenai pemahaman bisnis. Penelitian ini dilakukan untuk memberikan solusi terhadap masalah tersebut.

3.1.2 Studi Literatur

Pada tahap ini dilakukan penggalan teori dengan mencari referensi sebelumnya. Referensi yang digunakan bersifat relevan terhadap kebutuhan penelitian. Referensi dapat berupa buku, jurnal, dan situs-situs di internet. Referensi teori yang mendukung penelitian adalah *Customer Relationship Management (CRM)*, data mining, LRFM model, metode *elbow*, *clustering*, *Fuzzy C-Means clustering*, *Modified Partition Coefficient (MPC)*, *Davies Bouldin's Index*, *Connectivity*, *Dunn Index*, *Silhouette Width*, dan *Dashboard*. Dan referensi penelitian sebelumnya berkaitan dengan metode *clustering* dan model segmentasi pelanggan.

3.1.3 Pengumpulan Data

Pada tahap ini dilakukan kegiatan berupa permohonan data terhadap pihak Ride Inc. Data yang digunakan adalah data transaksi dari pelanggan. Data yang diberikan untuk penelitian ini adalah periode Juli 2017 hingga Maret 2018. Data tersebut berisi 668 data transaksi dengan 522 pelanggan. Isi dari data tersebut adalah id dari pelanggan, nama pelanggan, tanggal transaksi, total transaksi, nomor telepon, tipe sepatu, dan ukuran sepatu yang dibeli.

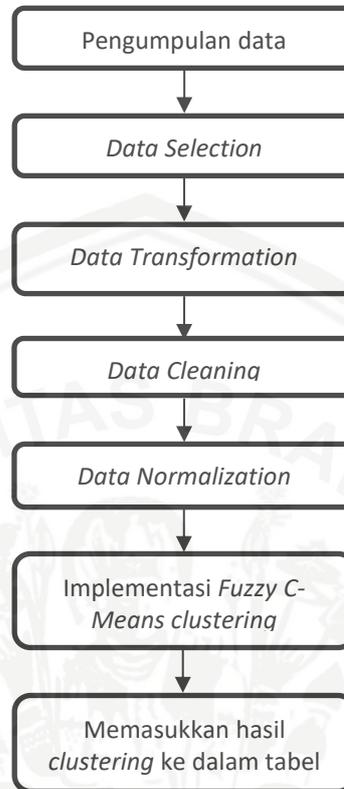
3.1.4 Preprocessing Data

Implementasi dari *preprocessing data* terdiri dari beberapa proses, diantaranya *data selection*, *data cleaning*, dan *data transformation*. Pertama, melakukan *data selection* untuk memilah data pelanggan dari Ride Inc agar siap dipakai berdasarkan aturan dari LRFM model. LRFM model terdiri dari *Length* atau jangka waktu pelanggan dari awal hingga akhir transaksi, *Recency* atau waktu transaksi terakhir pelanggan, *Frequency* atau jumlah frekuensi transaksi pelanggan dalam satu periode, dan *Monetary* atau nominal yang dikeluarkan pelanggan pada setiap transaksi. *Data cleaning* dilakukan setelah melakukan *data selection* untuk menghapus data yang bersifat *noisy* atau data yang terdeteksi memiliki jarak yang jauh dibandingkan dengan data yang lain atau disebut *outliers* pada data pelanggan Ride Inc. Setelah itu, dilakukan *data transformation* berupa normalisasi data menggunakan metode Min-Max.

3.1.5 Clustering

Setelah melakukan *preprocessing data*, langkah selanjutnya adalah melakukan *clustering* dengan metode *Fuzzy C-Means*. Langkah pertama adalah

menentukan jumlah *cluster* dengan menggunakan metode Elbow. Selanjutnya, implementasi *Fuzzy C-Means* yang akan menghasilkan derajat keanggotaan tiap-tiap data yang mencerminkan karakteristik pelanggan. Gambar 3.2 merupakan proses dari tahap *clustering*.



Gambar 3.2 Proses Clustering

3.1.6 Uji Validitas Hasil Clustering

Pada tahap ini dilakukan pengujian terhadap validitas *cluster*. Pengujian dilakukan dengan menggunakan tiga *function* dalam bahasa pemrograman R. *Function* yang digunakan adalah *Modified Partition Coefficient (MPC)*, *Davies Bouldin's Index*, dan *packages CValid*. *Packages CValid* berisi *Connectivity*, *Dunn Index*, dan *Silhouette Width*. Ketiga *function* tersebut memiliki hasil dengan rentang 0 hingga 1.

3.1.7 Uji Verifikasi Hasil Clustering

Pada tahap ini dilakukan pengujian terhadap verifikasi hasil *clustering*. Pengujian dilakukan dengan menggunakan perhitungan dari *Euclidean Distance*. Hasil dari perhitungan akan dilihat tingkat kesesuaiannya dengan hasil dari implementasi *Fuzzy C-Means clustering*. Hasil yang sesuai adalah apabila *cluster* dari pengujian *Euclidean Distance* sama dengan hasil implementasi *Fuzzy C-Means clustering*. Pengujian dilakukan dengan bahasa pemrograman R.

3.1.8 Pembuatan Visualisasi Hasil *Clustering*

Tahap selanjutnya adalah pembuatan visualisasi hasil *clustering*. Pembuatan visualisasi dilakukan setelah mendapatkan hasil uji verifikasi dan verifikasi. Terdapat tiga halaman dalam visualisasi yaitu halaman utama, halaman *cluster*, dan halaman data. Hasil dari visualisasi bertujuan untuk rekomendasi terhadap Ride Inc. dalam membantu menentukan strategi bisnis. Pembuatan visualisasi menggunakan *tools* Power BI.

3.1.9 Analisis Hasil *Usability Testing*

Pada tahap ini dilakukan uji *usability testing* sebagai dengan menggunakan metode *System Usability Scale* (SUS) berupa kuesioner. Pengujian dilakukan terhadap pihak Ride Inc. Setelah mendapatkan hasil pengujian, dilakukan analisis terhadap hasil *usability testing*. Analisis dilakukan untuk melihat kelayakan hasil visualisasi. Analisis hasil berupa perhitungan nilai dari setiap elemen dalam kuesioner dan mencari rata-rata dari jumlah responden.

3.1.10 Kesimpulan dan Saran

Pada tahap akhir dilakukan dua kegiatan yaitu penarikan kesimpulan dan saran. Penarikan kesimpulan diambil dari hasil yang menjawab rumusan masalah. Seluruh rumusan masalah harus terjawab berdasarkan penelitian. Selanjutnya, penentuan saran dilakukan untuk rekomendasi penelitian selanjutnya. Saran diambil berdasarkan kekurangan dari hasil penelitian ini.

BAB 4 PERANCANGAN

4.1 Pengumpulan Data

Pengumpulan data dilakukan pada toko sepatu Ride Inc. dengan mengekstraksi data mentah dari data transaksi. Data transaksi yang digunakan adalah dari periode 17 Juli 2017 hingga 31 Maret 2018 yang berupa *file excel* dan memiliki variabel dalam Tabel 4.1.

Tabel 4.1 Data Mentah Transaksi Pelanggan

| Variabel | Tipe Data | Keterangan |
|-------------|-----------|--|
| No | Text | ID transaksi pelanggan |
| Tanggal | Date | Tanggal transaksi pelanggan |
| NoTelp | Text | Nomor HP pelanggan |
| NamaPemesan | Text | Nama pelanggan |
| Artikel | Text | Jenis sepatu pelanggan |
| Size | Text | Ukuran sepatu pelanggan |
| Warna | Text | Warna sepatu pelanggan |
| Kulit | Text | Kulit yang akan digunakan oleh pelanggan |
| Sol | Text | Jenis sol yang akan digunakan oleh pelanggan |
| Qty | Number | Jumlah yang dibeli oleh pelanggan |
| Harga | Number | Total harga yang dibayar oleh pelanggan |

Berdasarkan data transaksi yang diberikan oleh Ride Inc, terdapat 667 transaksi serta 522 pelanggan.

4.2 Ekstraksi LRFM Model

Pada tahap ini dilakukan analisis dari nilai-nilai LRFM dengan mengelompokkan variabel yang akan digunakan sesuai dengan kebutuhannya. Dalam implementasinya, proses analisis LRFM menggunakan data berupa *excel* yang akan dimasukkan ke dalam tools *phpmyadmin* dengan basis MySQL untuk melakukan proses pengurangan jumlah hari, menghitung jumlah transaksi dan menghitung jumlah nominal transaksi dari setiap pelanggan.

4.2.1 Ekstraksi Variabel Length

Length merupakan nilai rentang waktu dari pembelian terakhir dengan pembelian pertama dari pelanggan dalam waktu tertentu. Variabel yang digunakan adalah id pelanggan dan tanggal dari transaksi pelanggan untuk menghasilkan selisih rentang waktu pembelian berupa jumlah hari dari setiap pelanggan. Proses yang dilewati berupa pengurangan waktu dari pembelian

terakhir dengan pembelian awal setiap pelanggan yang akan menghasilkan total hari. Contoh dari hasil adalah pelanggan dengan id 18 memiliki selisih 38 hari antara pembelian akhir dengan pembelian pertama.

4.2.2 Ekstraksi Variabel Recency

Recency merupakan nilai rentang waktu dari tanggal penelitian yaitu 24 April 2018 dengan pembelian terakhir setiap pelanggan. Variabel yang digunakan adalah id pelanggan dan tanggal transaksi pelanggan untuk menghasilkan selisih rentang waktu pembelian berupa jumlah hari dari setiap pelanggan. Hasil yang diperoleh berupa selisih hari dari setiap pelanggan, contohnya adalah pelanggan dengan id 18 memiliki selisih 20 hari antara tanggal penelitian dengan pembelian terakhir.

4.2.3 Ekstraksi Variabel Frequency

Frequency merupakan nilai dari jumlah pembelian yang dilakukan oleh setiap pelanggan dalam waktu tertentu yaitu, 17 Juli 2017 hingga 31 Maret 2018. Variabel yang digunakan adalah id pelanggan dan tanggal transaksi pelanggan untuk menghasilkan jumlah transaksi yang telah dilakukan setiap pelanggan. Proses yang dilewati adalah penjumlahan seluruh transaksi yang dilakukan setiap pelanggan, contohnya adalah pelanggan dengan id 18 memiliki jumlah pembelian pada Ride Inc. sebanyak 3 kali.

4.2.4 Ekstraksi Variabel Monetary

Monetary merupakan nilai dari total jumlah transaksi yang dikeluarkan oleh setiap pelanggan dalam waktu tertentu yaitu, 17 Juli 2017 hingga 31 Maret 2018. Variabel yang digunakan adalah id pelanggan dan nominal transaksi pelanggan untuk menghasilkan jumlah nominal yang telah dikeluarkan dari setiap pelanggan. Contoh dari hasil penjumlahan nominal adalah pelanggan dengan id 18 memiliki jumlah nominal dari transaksi sebesar Rp1.500.000.

4.3 Data Preprocessing

Sebelum masuk ke dalam tahap *clustering*, *data preprocessing* dilakukan untuk mengubah data mentah menjadi data yang bersih serta memiliki kesetaraan nilai berdasarkan komponen dari LRFM. Tahap-tahap dalam *data preprocessing* meliputi *data selection*, *data transformation*, *data cleaning* dan *data normalization*.

4.3.1 Data Selection

Pada tahap ini dilakukan pergantian nama variabel dan penyeleksian variabel untuk mempermudah proses *clustering* sesuai dengan LRFM. Berikut adalah nama variabel yang diubah dan digunakan dalam Tabel 4.2.

Tabel 4.2 Pergantian dan Pemilihan Variabel

| Variabel Lama | Variabel Baru |
|---------------|---------------|
| No | id_pelanggan |
| Tanggal | tanggal |
| Harga | total |

Tabel 4.2 merupakan variabel dari data yang siap dipakai untuk melewati tahap-tahap selanjutnya. Tabel 4.3 merupakan potongan data dari variabel yang terpilih berdasarkan data transaksi pelanggan dari Ride Inc.

Tabel 4.3 Variabel Baru dari Transaksi Pelanggan

| id_pelanggan | tanggal | total |
|--------------|------------|--------|
| 450 | 2017-07-17 | 350000 |
| 179 | 2017-07-18 | 375000 |
| 385 | 2017-07-19 | 350000 |
| 319 | 2017-07-20 | 580000 |

4.3.2 Data Transformation

Setelah melakukan data seleksi dan mendapatkan nilai dari setiap komponen LRFM, dilakukan data transformasi terhadap data LRFM. Data transformasi dilakukan untuk mengubah data yang memiliki rentang berbeda menjadi setara. Transformasi data dilakukan terhadap seluruh variabel, yaitu L, R, F, dan M. Metode yang digunakan dalam transformasi data adalah metode *Logarithmic*. Dan transformasi data dilakukan dengan bahasa pemrograman R. Contoh hasil yang didapat adalah pelanggan dengan id 18 memiliki nilai L sebesar 6.1, R sebesar 5.4, F sebesar 3.4, dan M sebesar 7.

4.3.3 Data Cleaning

Pada tahap ini dilakukan penghapusan data terhadap data transaksi Ride Inc. yang memiliki nilai kosong pada variabel total karena dapat menurunkan rentang nilai pada variabel *Monetary* dan memberikan nilai *Length*, *Recency* dan *Frequency* yang tidak sesuai. *Data cleaning* dilakukan terhadap 522 data pelanggan yang memiliki nilai kosong dan jarak dengan nilai lain terlalu jauh. Hasil yang terdeteksi sebagai *outlier* akan diubah menjadi *NA* dan pelanggan akan terhapus.

4.3.4 Data Normalization

Rentang nilai yang dihasilkan dari *data transformation* menggunakan metode *Logarithmic* belum menggambarkan kesetaraan nilai setiap variabel seutuhnya. Oleh karena itu, dilakukan normalisasi data dengan metode *Min-Max* untuk mendapatkan kesetaraan nilai setiap variabel dengan rentang 0 hingga 1.

Menurut hasil penelitian yang dilakukan oleh Angelie (2017), dilakukan pengurangan 1 dengan nilai dari hasil normalisasi *Recency* karena nilai *Recency* berbanding terbalik dengan nilai *Frequency* dan *Monetary*.

4.4 Penentuan Jumlah *Cluster*

Pada tahap ini dilakukan pengimplementasian metode *Elbow* untuk menentukan jumlah *cluster* yang akan digunakan. Hasil jumlah *cluster* ditampilkan dalam bentuk grafik. Penentuan jumlah *cluster* dilihat dari titik yang memiliki penurunan paling drastis. Implementasi metode *Elbow* dilakukan pada bahasa pemrograman R. Jumlah *cluster* yang dihasilkan adalah 3 *cluster*.

4.5 *Clustering* Menggunakan *Fuzzy C-Means*

Setelah mendapatkan jumlah *cluster*, dilakukan pengimplementasian dari *Fuzzy C-Means* untuk mendapatkan posisi data sesuai dengan kriteria *cluster*. Data yang digunakan adalah data LRFM yang telah melewati *data preprocessing*. Implementasi *Fuzzy C-Means* menggunakan *query* dari *packages* *fclust* yang disediakan oleh bahasa pemrograman R. Hasil *cluster* akan menggambarkan posisi setiap data yang memiliki kesamaan dengan data lainnya. Hasil tersebut kemudian dimasukkan ke dalam tabel LRFM sesuai id dari pelanggan untuk melakukan tahap selanjutnya.

4.6 Pengujian *Cluster*

4.6.1 Uji Validitas *Cluster*

Setelah mendapatkan hasil *clustering*, hal selanjutnya yang dilakukan adalah menguji validitas dari hasil tersebut untuk mengetahui apakah *cluster* tersebut sudah layak untuk dikatakan *cluster* terbaik atau tidak. Dalam pengujiannya metode yang digunakan adalah MPC atau *Modified Partition Coefficient* (MPC), *Davies Bouldin's Index*, dan penerapan *packages* *CValid* dengan menggunakan tools R. Hasil dari setiap metode akan dibandingkan untuk melihat *cluster* tervalid.

4.6.2 Uji Verifikasi *Cluster*

Selain uji validitas, dilakukan pengujian terhadap verifikasi *cluster* dengan menggunakan perhitungan *Euclidean Distance*. Pengujian ini dilakukan untuk memastikan *cluster* setiap data yang dihasilkan dari proses *clustering* telah sesuai dengan menghitung jarak dari posisi data ke *centroid*. Setelah mendapatkan hasil perhitungan, maka setiap data akan dimasukkan ke dalam *cluster* sesuai jarak minimum yang dihasilkan. Dan hasil *cluster* tersebut akan dicocokkan dengan hasil *clustering* metode *Fuzzy C-Means*.

4.7 Desain Arsitektur Proses Clustering

Tahap ini merupakan gambaran dari hubungan terkait data dengan tools yang digunakan. Hubungan dari data, tools R, dan PowerBI dapat dilihat dalam Gambar 4.1.

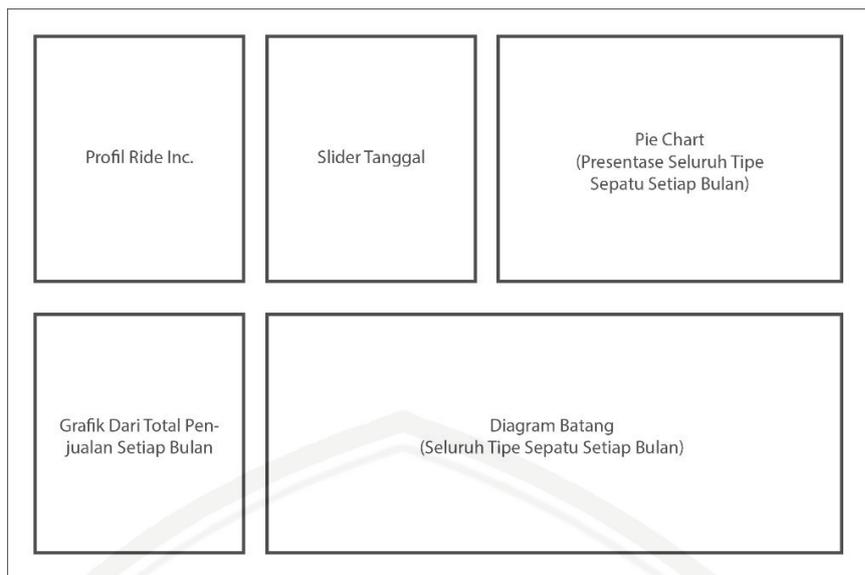


Gambar 4.1 Desain Arsitektur

Dalam implementasinya, data dalam database Ride Inc. diolah dengan tools *phpmyadmin* untuk melakukan proses seleksi dan pengambilan nilai *Length*, *Recency*, *Frequency*, dan *Monetary*. Setelah mendapatkan nilai dari LRFM, data diekspor ke dalam format *'xlsx'* untuk dimasukkan ke dalam *R*. Proses yang terjadi dalam *R* berupa *data transformation*, *data cleaning*, penentuan jumlah *cluster*, *clustering* menggunakan *Fuzzy C-Means*, dan uji validitas dengan *Modified Partition Coefficient (MPC)*. Setelah melakukan semua proses, dataset dengan nilai LRFM dan *cluster* diubah menjadi data dengan format *'xlsx'* untuk diolah menjadi *dashboard* sesuai desain yang telah dirancang.

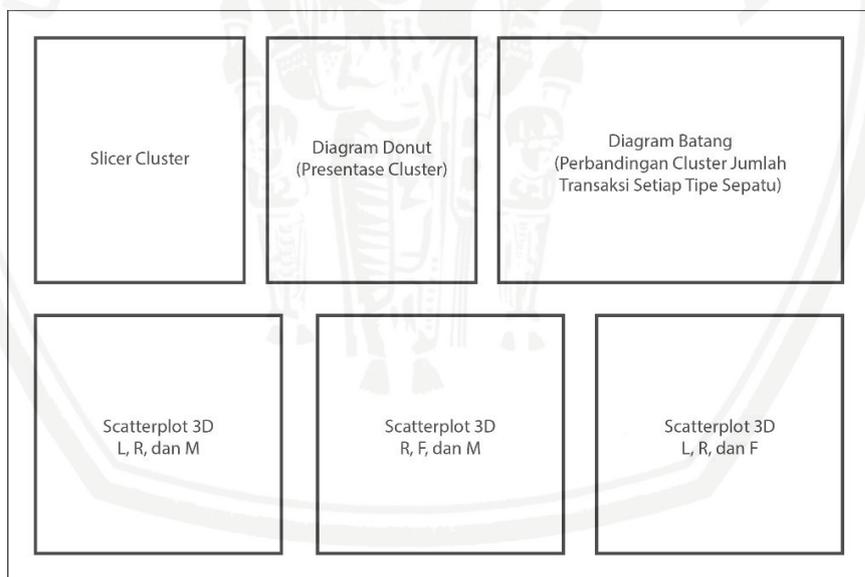
4.8 Desain Visualisasi Dashboard

Rancangan dari desain visualisasi *dashboard* terdiri dari tiga halaman yaitu *dashboard* sebagai halaman utama, *2 cluster* dan *3 cluster* sebagai halaman *cluster*, dan data *2 cluster* dan data *3 cluster* sebagai halaman data. Implementasi dari rancangan visualisasi *dashboard* menggunakan tools *PowerBI* yang diintegrasikan dengan *R*. Gambar 4.2 merupakan rancangan dari halaman utama *dashboard* yang berisi profil instansi, *slider* tanggal untuk menentukan data yang ingin ditampilkan, *pie chart* yang menunjukkan presentase penjualan seluruh tipe sepatu, grafik dari total pendapatan setiap bulan, dan diagram batang dari penjualan seluruh tipe sepatu setiap bulannya.



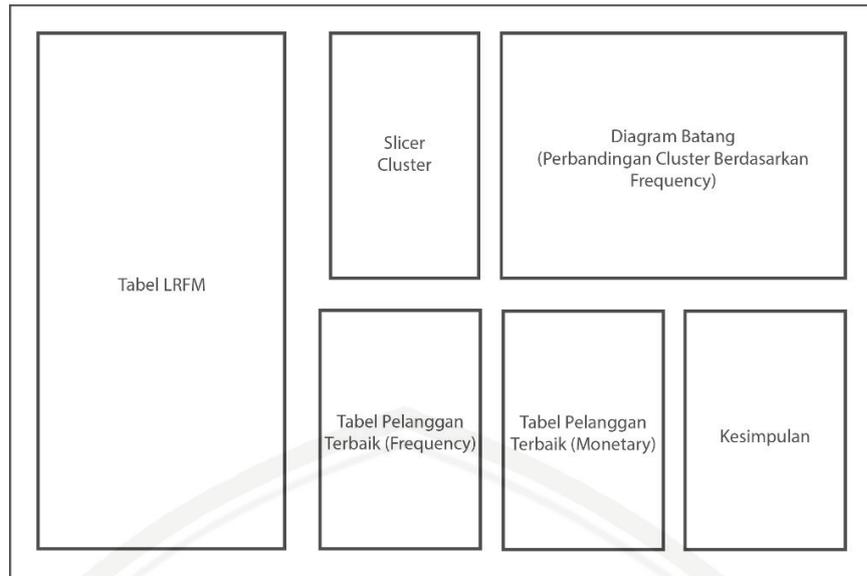
Gambar 4.2 Halaman Utama Dashboard

Selanjutnya, Gambar 4.3 merupakan rancangan dari halaman *cluster*. Pada halaman ini terdapat *slicer* dari *cluster* untuk menampilkan data berdasarkan *cluster*, diagram donut untuk menampilkan presentase dari tiap *cluster*, diagram batang untuk menampilkan jumlah transaksi untuk setiap tipe sepatu yang dijual, *scatter chart* untuk menampilkan persebaran pelanggan dari hubungan LRM, LFM, dan RFM.



Gambar 4.3 Halaman Cluster

Gambar 4.4 merupakan rancangan dari halaman data 2 *cluster* dan 3 *cluster*. Halaman ini terdapat *slicer* dari *cluster* untuk menampilkan data sesuai *cluster* yang diinginkan, memuat tabel LRFM, dan memuat tabel pelanggan terbaik berdasarkan *Frequency* dan *Monetary*, diagram batang yang menampilkan jumlah pelanggan berdasarkan variabel *Frequency* dan *fields* untuk menampilkan kesimpulan dari halaman data.



Gambar 4.4 Halaman Data

4.9 Pengujian *System Usability Scale (SUS)*

Pengujian *usability* dilakukan dengan metode *System Usability Scale (SUS)* yang dikenalkan oleh Brooke (1986). Pengujian ini memiliki 10 instrumen yang perlu diisi oleh responden dengan mengetahui bahwa instrumen ganjil merupakan instrumen positif dan instrumen genap merupakan instrument negatif. Tabel 4.4 merupakan isi dari instrument tersebut.

Tabel 4.4 Instrumen dari Metode SUS

| No. | Instrumen |
|-----|---|
| 1. | <i>I think that I would like to use this system frequently.</i> |
| 2. | <i>I found the system unnecessarily complex.</i> |
| 3. | <i>I thought the system was easy to use.</i> |
| 4. | <i>I think that I would need the support of a technical person to be able to use this system.</i> |
| 5. | <i>I found the various functions in this system were well integrated.</i> |
| 6. | <i>I thought there was too much inconsistency in this system.</i> |
| 7. | <i>I would imagine that most people would learn to use this system very quickly.</i> |
| 8. | <i>I found the system very cumbersome to use.</i> |
| 9. | <i>I felt very confident using the system.</i> |
| 10. | <i>I needed to learn a lot of things before I could get going with this system.</i> |

BAB 5 IMPLEMENTASI

5.1 Data Preprocessing

Dalam tahap *data preprocessing*, terdapat beberapa proses yang dilakukan sesuai dengan metodologi pada bab 3, yaitu *data selection*, *data transformation*, dan *data transformation*. Penjelasan setiap tahap akan dirincikan berdasarkan urutan dari pengerjaan pada penelitian.

5.1.1 Data Selection

Pada tahap *data selection* dilakukan seleksi atribut dari data mentah pada data transaksi Ride Inc. yang diperlukan dalam proses *clustering* berdasarkan LRFM model. Data yang digunakan adalah seperti dalam Tabel 5.1.

Tabel 5.1 Variabel Terpilih untuk Proses Clustering

| Variabel | Keterangan |
|--------------|---|
| id_pelanggan | ID setiap pelanggan |
| tanggal | Tanggal setiap transaksi dari pelanggan |
| total | Jumlah nominal yang dikeluarkan pada setiap transaksi |

5.1.2 Pemetaan LRFM

Data yang digunakan pada tahap ini adalah data yang telah melewati tahap *data selection*, yaitu id_transaksi, tanggal dan total. Pemetaan LRFM dilakukan untuk mendapatkan nilai dari setiap atribut, yaitu *Length*, *Recency*, *Frequency* dan *Monetary* dengan menggunakan MySQL pada *phpmyadmin*.

Nilai *Length* merupakan selisih antara waktu terakhir transaksi dengan waktu pertama transaksi dari setiap pelanggan. Atribut yang digunakan untuk mendapatkan nilai *Length* adalah id_pelanggan dan tanggal. Untuk mendapatkan waktu pertama dan terakhir transaksi dari pelanggan dapat menggunakan fungsi MAX dan MIN dari tanggal transaksi berdasarkan id_pelanggan. Dan untuk mendapatkan selisih waktunya dapat menggunakan fungsi DATEDIFF.

Nilai *Recency* merupakan selisih antara waktu pengimplementasian yaitu 24 April 2018 dengan waktu terakhir transaksi dari setiap pelanggan. Atribut yang digunakan pada *Recency* adalah id_pelanggan dan tanggal. Untuk mendapatkan waktu saat ini dapat menggunakan fungsi CURDATE dan untuk mendapatkan waktu terakhir dapat menggunakan fungsi MAX dari tanggal. Sedangkan untuk mendapatkan nilai selisih waktunya dapat menggunakan fungsi DATEDIFF.

Nilai *Frequency* merupakan total dari transaksi yang dilakukan oleh pelanggan dalam satu periode. Atribut yang digunakan adalah id_pelanggan dan tanggal. Dan untuk mendapatkan total dari transaksi dari setiap pelanggan menggunakan fungsi COUNT berdasarkan id_pelanggan.

Nilai *Monetary* merupakan total nominal yang dikeluarkan setiap pelanggan. Atribut yang digunakan adalah *id_pelanggan* dan *total*. Dan untuk mendapatkan total nominal pada setiap pelanggan menggunakan fungsi *SUM* berdasarkan *id_pelanggan*.

Script 5.1 merupakan *query* yang dijalankan pada *phpmyadmin* untuk mendapatkan nilai dari *Length*, *Recency*, *Frequency*, dan *Monetary*.

```
INSERT INTO lrfm (id_pelanggan, length, recency, frequency, monetary)
SELECT id_pelanggan, DATEDIFF(MAX(tanggal), MIN(tanggal)),
DATEDIFF(CURDATE(), MAX(tanggal)), COUNT(tanggal), SUM(total) FROM
rideinc GROUP BY id_pelanggan
```

Script 5.1 Pengubahan Data Menjadi Nilai LRFM

Setelah mengeksekusi *query* diatas, terdapat 522 baris yang ditampilkan yang berarti terdapat 522 pelanggan yang melakukan transaksi pada periode 17 Juli 2017 hingga 31 Maret 2018.

5.1.3 Data Transformation

Data yang telah diubah menjadi nilai LRFM dilakukan normalisasi menggunakan metode *Logarithmic*. Normalisasi dilakukan karena nilai pada LRFM tidak normal dengan memiliki perbedaan rentang yang cukup jauh dan berbeda-beda. Menurut Elliot (1977) jika dalam sebuah dataset terdapat nilai nol, maka penambahan satu ke dalam data diperlukan sebelum melakukan transformasi *logarithmic*. Dalam variabel *Length* dan *Monetary* terdapat nilai 0, maka dalam proses transformasi variabel tersebut ditambahkan 1 agar hasil transformasi bisa teridentifikasi. Proses normalisasi *Logarithmic* dilakukan pada *tools* Rstudio dalam Script 5.2.

```
> ltransform<-log10(lrfm$length + 1)
> rtransform<-log10(lrfm$recency)
> ftransform<-log10(lrfm$frequency)
> mtransform<-log10(lrfm$monetary + 1)
> lrfmtransform<-data.frame(ltransform, rtransform, ftransform, mtransform)
```

Script 5.2 Transformasi Data dengan Metode *Logarithmic*

Data yang diolah berasal dari dataset LRFM dan akan dimasukkan kedalam variabel baru, seperti nilai *Length* dimasukkan ke dalam *ltrans*, nilai *Recency* dimasukkan ke dalam *rtrans*, nilai *Frequency* dimasukkan ke dalam *ftrans*, dan nilai *Monetary* dimasukkan kedalam *mtrans*. Setelah mendapatkan nilai pada setiap variabel, proses selanjutnya adalah memasukkan semua variabel ke dalam dataset baru yang merepresentasikan nilai transformasi dengan nama *lrfmtransform*.

5.1.4 Data Cleaning

Data *cleaning* dilakukan menggunakan script yang dibuat oleh Klodian Dhana (2016) yang berfungsi untuk menghapus *outlier* agar tidak merusak hasil dari *clustering*. Script yang digunakan adalah seperti dalam Script 5.3.

```

1 outlierKD <- function(dt, var) {
2   var_name <- eval(substitute(var),eval(dt))
3   na1 <- sum(is.na(var_name))
4   m1 <- mean(var_name, na.rm = T)
5   par(mfrow=c(2, 2), oma=c(0,0,3,0))
6
7   #Menampilkan boxlot dan histogram dengan outliers
8   boxplot(var_name, main="With outliers")
9   hist(var_name, main="With outliers", xlab=NA, ylab=NA)
10
11  #Menampilkan boxplot dan histogram tanpa outliers
12  outlier <- boxplot.stats(var_name)$out
13  mo <- mean(outlier)
14  var_name <- ifelse(var_name %in% outlier, NA, var_name)
15  boxplot(var_name, main="Without outliers")
16  hist(var_name, main="Without outliers", xlab=NA, ylab=NA)
17
18  #Pengecekan outliers
19  title("Outlier Check", outer=TRUE)
20  na2 <- sum(is.na(var_name))
21
22  #Menampilkan hasil jumlah outliers dan proporsi outliers
23  cat("Outliers identified:", na2 - na1, "n")
24  cat("Propotion (%) of outliers:", round((na2 - na1) /
25  sum(!is.na(var_name))*100, 1), "n")
26
27  #Menampilkan rata-rata outliers
28  cat("Mean of the outliers:", round(mo, 2), "n")
29  m2 <- mean(var_name, na.rm = T)
30
31  #Menampilkan rata-rata data saat belum dan setelah menghapus outliers
32  cat("Mean without removing outliers:", round(m1, 2), "n")
33  cat("Mean if we remove outliers:", round(m2, 2), "n")
34
35  #Menghapus outliers dengan respon yes
36  response <- readline(prompt="Do you want to remove outliers and to
37  replace with NA? [yes/no]: ")
38  if(response == "y" | response == "yes"){
39    dt[as.character(substitute(var))] <- invisible(var_name)
40    assign(as.character(as.list(match.call())$dt), dt, envir = .GlobalEnv)

```

```

    cat("Outliers successfully removed", "n")
    return(invisible(dt))
  } else{
    cat("Nothing changed", "n")
    return(invisible(var_name))
  }
}

```

Script 5.3 Penghapusan *Outlier* dengan *Script* Klodian Dhana (2016)

Pengimplementasian dari penghapusan *outliers* hanya dilakukan pada variabel *Recency* dan *Monetary* karena nilai pada variabel *Length* hanya terdapat 11,7% data yang tidak bernilai 0 dan pada variabel *Frequency* hanya terdapat 18,1% data yang tidak bernilai 1. Jika penghapusan *outliers* dilakukan pada variabel *Length* dan *Frequency* maka nilai pada variabel *Length* yang tidak 0 dan nilai pada variabel *Frequency* yang tidak bernilai 1 akan teridentifikasi sebagai *outliers*.

5.1.5 Normalisasi Min-Max

Selanjutnya dilakukan normalisasi menggunakan metode *Min-Max* untuk mendapatkan kesetaraan nilai pada semua atribut dengan rentang 0 hingga 1. Implementasi normalisasi *Min-Max* menggunakan *tools* Rstudio dengan script seperti dalam Script 5.4.

```

> Inormal <- (ltransform-min(ltransform))/(max(ltransform)-min(ltransform))
> rnormal <- (rtransform-min(rtransform))/(max(rtransform)-min(rtransform))
> fnormal <- (ftrans-min(ftransform))/(max(ftransform)-min(ftransform))
> mnormal <- (mtrans-min(mtransform))/(max(mtransform)-min(mtransform))

```

Script 5.4 Normalisasi Data dengan Metode *Min-Max*

Variabel yang dinormalisasi adalah *ltrans* yaitu transformasi dari *Length*, *rtrans* yaitu transformasi dari *Recency*, *ftrans* yaitu transformasi dari *Frequency*, dan *mtrans* yaitu transformasi dari *Monetary*. Hasil dari normalisasi akan disimpan dalam variabel baru, yaitu *Inorm* untuk normalisasi dari *Length*, *rnorm* untuk normalisasi dari *Recency*, *fnorm* untuk normalisasi dari *Frequency*, dan *mnorm* untuk normalisasi dari *Monetary*. Penelitian yang dilakukan oleh Angelie, Annisa Veronica (2017) mengatakan bahwa nilai *Recency* berbanding terbalik dengan *Frequency* dan *Monetary*, sehingga setelah semua variabel telah dinormalisasi, dilakukan proses pengurangan 1 dengan nilai *Recency* untuk mendapatkan hasil yang sebenarnya. Setelah mendapatkan nilai asli *Recency*, semua variabel dimasukkan ke dalam dataset baru untuk melakukan tahap *clustering*.

5.2 Clustering menggunakan Fuzzy C-Means

Pada tahap *clustering* dilakukan dua proses yaitu penentuan jumlah *cluster* dan implementasi dari metode *Fuzzy C-Means*. Jumlah *cluster* yang ditentukan menggunakan metode *elbow* dengan melihat grafik pada hasil.

5.2.1 Elbow Method

Setelah dataset telah siap digunakan, langkah selanjutnya adalah pengimplementasian metode *elbow* untuk menentukan jumlah *cluster* yang tepat berdasarkan dataset tersebut. Dataset yang digunakan adalah variabel *lrfmnorm*, yakni hasil dari normalisasi dataset yang telah melalui data seleksi dan data transformasi. Hasil dari implementasi metode *elbow* disimpan dalam variabel baru bernama *hasillrfm* dari inisiasi 1 hingga 10 percobaan jumlah *cluster*. Script yang digunakan adalah seperti dalam Script 5.5.

```
hasillrfm <- sapply(1:10, function(k){kmeans(lrfmnorm, k, nstart=10,iter.max =
10)$tot.withinss})
plot(1:10, hasillrfm, type="b", pch = 19, frame = FALSE, xlab="Number of
clusters K", ylab="Total within-clusters sum of squares")
```

Script 5.5 Penentuan Jumlah *Cluster* dengan Metode *Elbow*

Setelah berhasil menjalankan script 5.5, hasil yang didapat adalah berupa grafik *elbow* yang menampilkan penurunan drastis untuk menentukan *cluster* yang cocok untuk digunakan.

5.2.2 Fuzzy C-Means

Proses *clustering* dilakukan setelah mendapatkan jumlah *cluster* dari metode *elbow*. Proses *clustering* menggunakan metode *Fuzzy C-Means* dengan menggunakan tools *Rstudio*. Instalasi *packages* *ppclust* terlebih dahulu. Setelah menginstalasi *packages* *ppclust*, masuk ke dalam library *ppclust* untuk dapat menggunakan fungsi dari *fcm*. Pengimplementasian *fcm* menggunakan dataset *lrfmnorm* yang berisi data yang telah dinormalisasi dan jumlah *cluster* akan dimasukkan ke dalam variabel *centers* sesuai hasil dari implementasi metode *elbow* menggunakan Script 5.6.

```
install.packages("ppclust")
library(ppclust)
rcluster <- fcm(lrfmnorm, centers="n")
```

Script 5.6 *Clustering* dengan Metode *Fuzzy C-Means*

5.3 Uji Validitas *Cluster*

5.3.1 Modified Partition Coefficient (MPC)

Langkah pertama adalah masuk ke dalam library *fclust* yang telah disediakan oleh *packages* *ppclust*. Kemudian inialisasi variabel sebagai wadah untuk menyimpan hasil implementasi MPC. Terakhir, inialisasi variabel terhadap implementasi dari MPC. Script yang digunakan adalah seperti dalam Script 5.7.

```
library(fclust)
hasil <- ppclust2(cluster, "fclust")
```

```
mpc <- MPC(hasil$U)
```

Script 5.7 Uji Validitas *Cluster* dengan *Modified Partition Coefficient* (MPC)

Hasil yang didapat adalah berupa nilai dengan rentang 0 hingga 1 dan semakin mendekati angka 1, maka jumlah *cluster* yang digunakan akan semakin tepat dengan data.

5.3.2 *Davies Bouldin's Index*

Langkah pertama adalah menginstall *packages* *clusterSim* dan setelah itu masuk ke dalam library tersebut. Selanjutnya melakukan inisialisasi variabel sebagai wadah untuk menyimpan hasil dari implementasi *DB Index*. Script 5.8 adalah *script* yang digunakan dalam pengimplementasian.

```
install.packages("clusterSim")
library(clusterSim)
db <- index.DB(lrfmnormal, cluster, centrotypes="centroids", p=2)
```

Script 5.8 Uji Validitas *Cluster* dengan *DB Index*

5.3.3 *CIValid*

Langkah pertama adalah menginstall *packages* *clValid* dan setelah itu masuk ke dalam library tersebut. Selanjutnya melakukan inisialisasi variabel sebagai wadah untuk menyimpan hasil dari implementasi *clValid*. Script 5.9 adalah *script* yang digunakan dalam pengimplementasian.

```
install.packages("clValid")
library(clValid)
clusterCIValid <- clValid(dataV, nClust = 2:10, clMethods = "kmeans", validation = "internal")
```

Script 5.9 Uji Validitas *Cluster* dengan *CIValid*

5.4 Uji Verifikasi *Cluster*

Pengujian verifikasi *cluster* dilakukan dengan menggunakan *Euclidean Distance*. Variabel yang digunakan untuk menghitung jarak ke *centroid* adalah hasil normalisasi LRFM dan *centroid* yang didapat dari hasil implementasi metode *clustering*. Script 5.10 adalah *script* yang digunakan dalam perhitungan *Euclidean Distance*.

```
#mengambil centroid cluster
centroidc<-data.frame(t(r3cluster$v[1,]))

#menghitung distance cluster
```

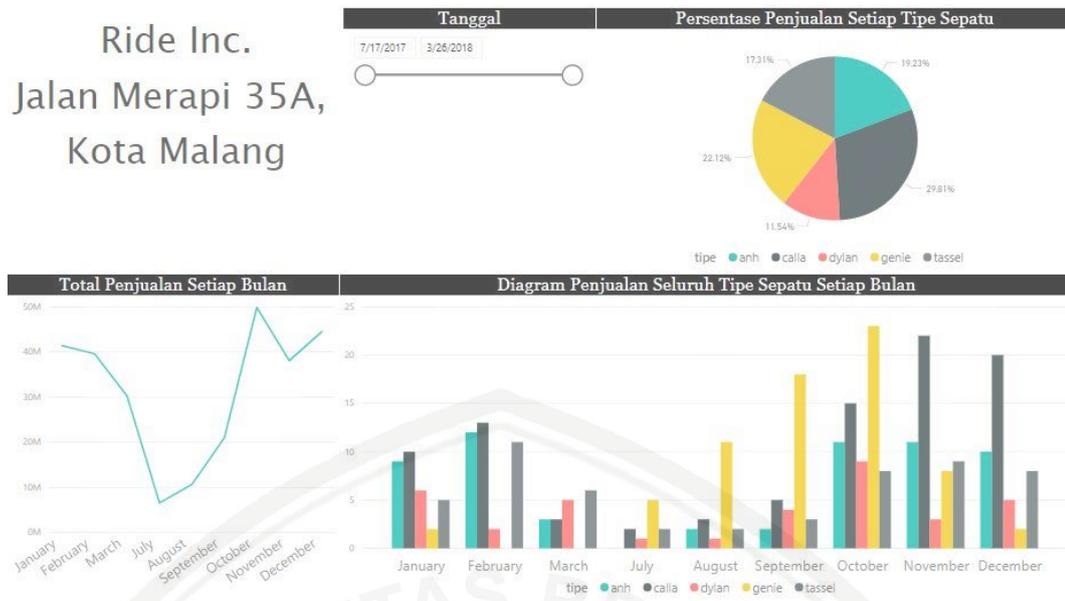
```
hasildist31<-sqrt(((lrfmnormal$lnormal-  
centroid1c$lnormal)^2)+((lrfmnormal$rbalik_baru-  
centroid1c$rbalik_baru)^2)+((lrfmnormal$fnormal-  
centroid1c$fnormal)^2)+((lrfmnormal$mnormal-centroid1c$mnormal)^2))  
  
#memasukkan hasil perhitungan kedalam tabel baru  
hasildist3<-data.frame(hasildist31)  
  
#mengambil nilai minimal dari setiap baris  
nilaimin<-apply(hasildist3, 1, min)  
  
#mengambil nilai cluster sesuai nilai minimal  
nilaimin3<-apply(hasildist3, 1, which.min)  
  
#memasukkan ke dalam tabel akhir  
hasildist3<-data.frame(id_pelanggannorm, hasildist31, hasildist32, hasildist33,  
nilaimin, nilaimin3)
```

Script 5.10 Verifikasi dengan *Euclidean Distance*

5.5 Visualisasi *Dashboard*

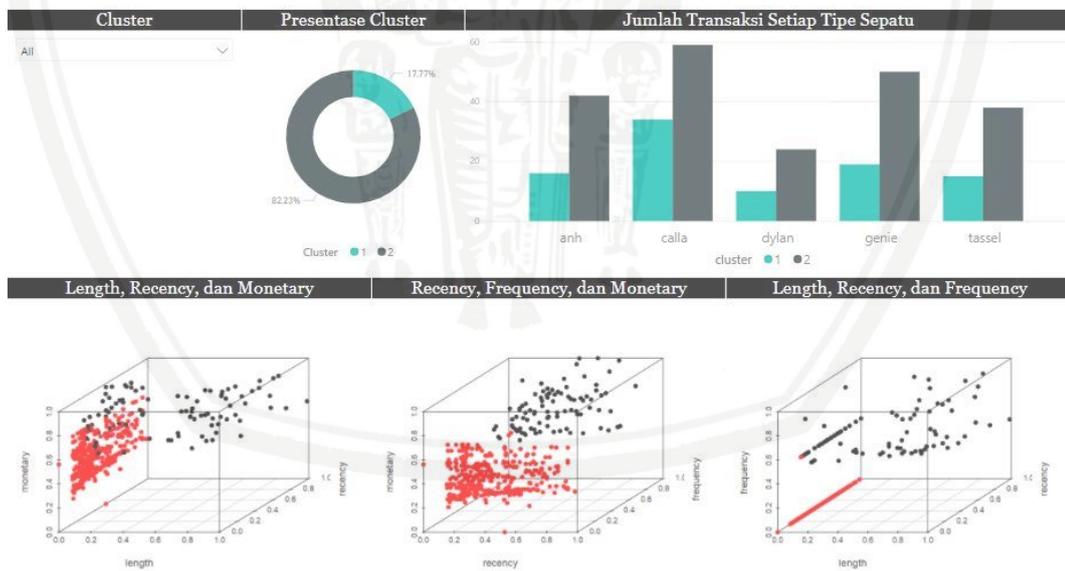
Tahap selanjutnya adalah pembuatan visualisasi dashboard dengan menggunakan Power BI. Halaman yang akan dibuat adalah halaman utama, halaman data dan halaman *cluster*. Pada halaman utama terdapat profil dari Ride Inc., tanggal untuk menentukan grafik yang akan ditampilkan pada periode tertentu, presentase penjualan setiap tipe sepatu, grafik total penjualan setiap bulannya dan diagram penjualan setiap bulan berdasarkan tipe sepatu. Gambar 5.1 merupakan tampilan dari halaman utama.

Ride Inc.
 Jalan Merapi 35A,
 Kota Malang



Gambar 5.1 Halaman Utama *Dashboard*

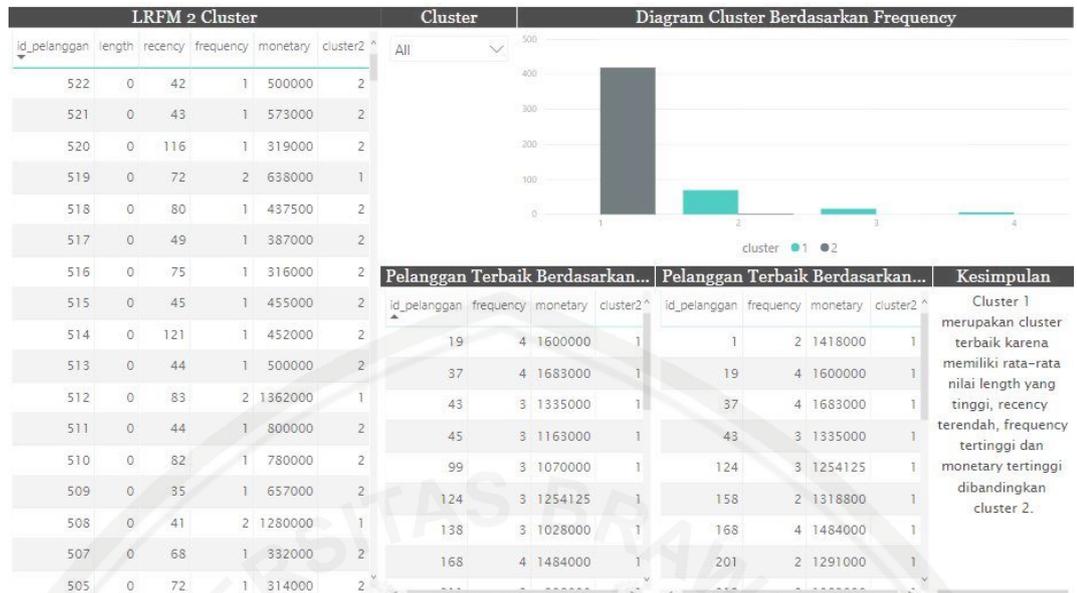
Selanjutnya adalah halaman *cluster* dan didalamnya terdapat presentase isi dari jumlah setiap *cluster*, jumlah transaksi setiap tipe sepatu pada tiap-tiap *cluster*, dan 3D Scatterplot dari LRM (*Length*, *Recency*, dan *Monetary*), RFM (*Recency*, *Frequency*, dan *Monetary*) dan LRF (*Length*, *Recency*, dan *Frequency*) untuk merepresentasikan posisi setiap data berdasarkan *cluster*. Gambar 5.2 merupakan contoh dari halaman *cluster*.



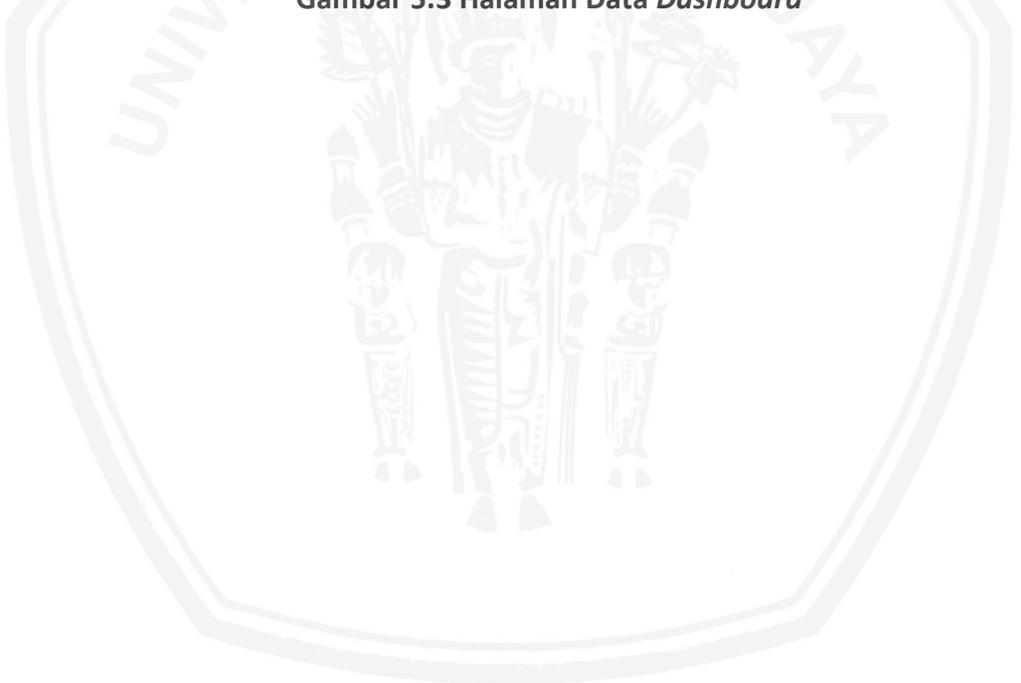
Gambar 5.2 Halaman *Cluster Dashboard*

Halaman terakhir yaitu halaman data dari *cluster* dan didalam halaman tersebut terdapat tabel LRFM (*Length*, *Recency*, *Frequency* dan *Monetary*), diagram batang *cluster* berdasarkan variabel F atau *Frequency*, tabel pelanggan

terbaik berdasarkan *Frequency*, pelanggan terbaik berdasarkan monetary dan kesimpulan. Gambar 5.3 merupakan halaman data dari *cluster*.



Gambar 5.3 Halaman Data Dashboard



BAB 6 ANALISIS DAN PEMBAHASAN

Bab ini akan menjelaskan analisis dari proses hasil *clustering* dan visualisasi hasil *clustering* pada *dashboard* dan pembahasan terkait pemetaan LRFM, proses *clustering* pada *Fuzzy C-Means*, uji validitas *cluster*, uji verifikasi *cluster*, implementasi dari perancangan *dashboard*, dan *usability testing*.

6.1 Pemetaan LRFM

Tahap ini dilakukan proses pemetaan dari LRFM berdasarkan data pelanggan yang telah diolah pada proses *data selection*. Variabel *Length*, *Recency*, dan *Frequency* menggunakan variabel tanggal dari transaksi dan variabel *Monetary* menggunakan variabel harga. Tabel 6.1 merupakan potongan dari hasil pemetaan variabel *Length*.

Tabel 6.1 Pemetaan Nilai Variabel *Length*

| Id_pelanggan | <i>Length</i> |
|--------------|---------------|
| 1 | 150 |
| 37 | 174 |
| 126 | 0 |
| 211 | 43 |

Selanjutnya pada Tabel 6.2 terdapat potongan dari hasil pemetaan variabel *Recency*.

Tabel 6.2 Pemetaan Nilai Variabel *Recency*

| Id_pelanggan | <i>Recency</i> |
|--------------|----------------|
| 1 | 79 |
| 37 | 70 |
| 126 | 193 |
| 211 | 199 |

Sedangkan potongan dari hasil pemetaan variabel *Frequency* ditunjukkan dalam Tabel 6.3.

Tabel 6.3 Pemetaan Nilai Variabel *Frequency*

| Id_pelanggan | <i>Frequency</i> |
|--------------|------------------|
| 1 | 2 |
| 37 | 4 |
| 126 | 1 |
| 211 | 3 |

Potongan dari hasil pemetaan variabel *Monetary* ditampilkan pada Tabel 6.4.

Tabel 6.4 Pemetaan Nilai Variabel *Monetary*

| Id_pelanggan | <i>Monetary</i> |
|--------------|-----------------|
| 1 | 1418000 |
| 37 | 1683000 |
| 126 | 578800 |
| 211 | 998000 |

Setelah mendapatkan nilai dari variabel *Length*, *Recency*, *Frequency*, dan *Monetary* dilakukan *pre-processing data* karena interval nilai masih beragam dan tinggi.

6.2 *Pre-processing Data*

Sub bab ini berisi pembahasan dari *data transformation*, *data cleaning*, dan normalisasi menggunakan metode Min-Max.

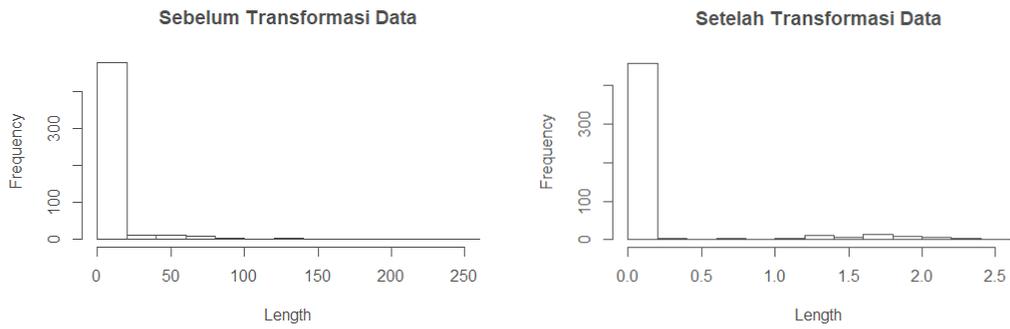
6.2.1 *Data Transformation*

Variabel LRFM diolah menggunakan metode *Logarithmic* karena nilai masih bersifat tidak konsisten dan dilakukan penambahan 1 terhadap variabel *Length* karena terdapat data yang bernilai 0. Hasil dari transformasi data ditunjukkan dalam Tabel 6.5.

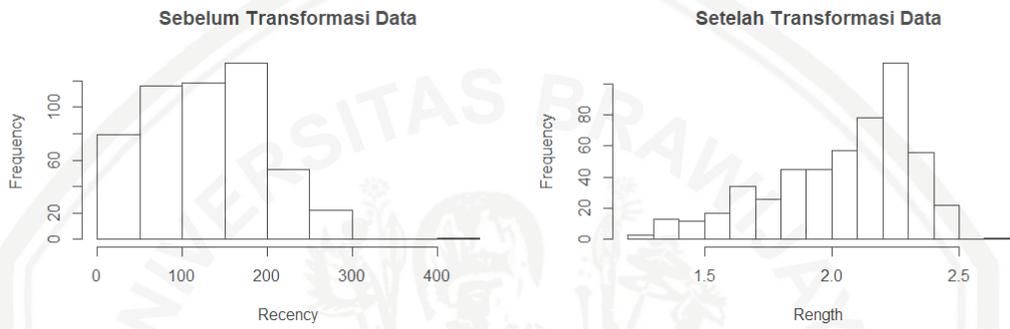
Tabel 6.5 Nilai Hasil Transformasi Data

| Id_pelanggan | <i>Length</i> | <i>Recency</i> | <i>Frequency</i> | <i>Monetary</i> |
|--------------|---------------|----------------|------------------|-----------------|
| 1 | 2.178977 | 1.90309 | 0.477121 | 6.151677 |
| 37 | 2.243038 | 1.851258 | 0.69897 | 6.226084 |
| 126 | 0 | 2.287802 | 0.30103 | 5.762529 |
| 211 | 1.643453 | 2.079181 | 0.60206 | 5.999131 |

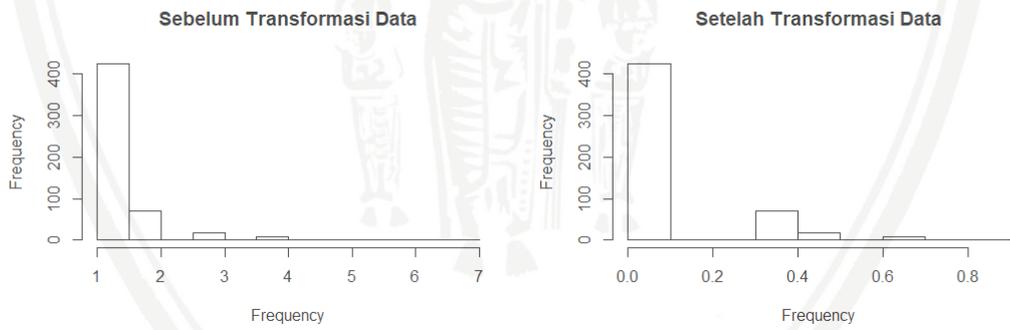
Nilai hasil dari transformasi data dapat dibandingkan dengan nilai data sebelum ditransformasikan dengan grafik histogram pada tools Rstudio yang ditunjukkan seperti dalam Gambar 6.1 hingga Gambar 6.4.



Gambar 6.1 Histogram Nilai Variabel *Length* Sebelum dan Sesudah Transformasi Data

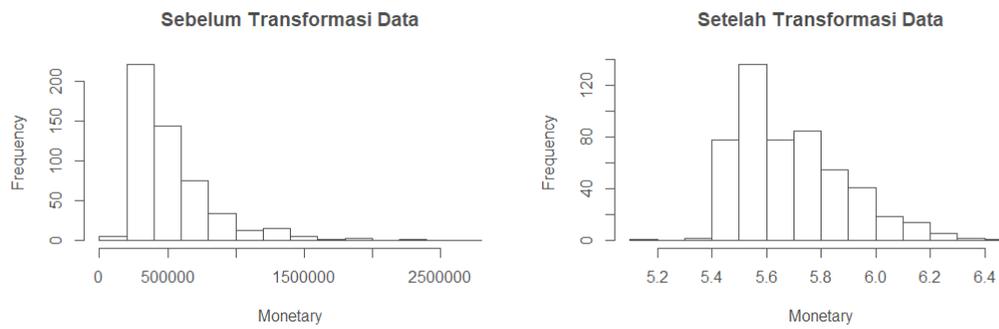


Gambar 6.2 Histogram Nilai Variabel *Recency* Sebelum dan Sesudah Transformasi Data



Gambar 6.3 Histogram Nilai Variabel *Frequency* Sebelum dan Sesudah Transformasi Data



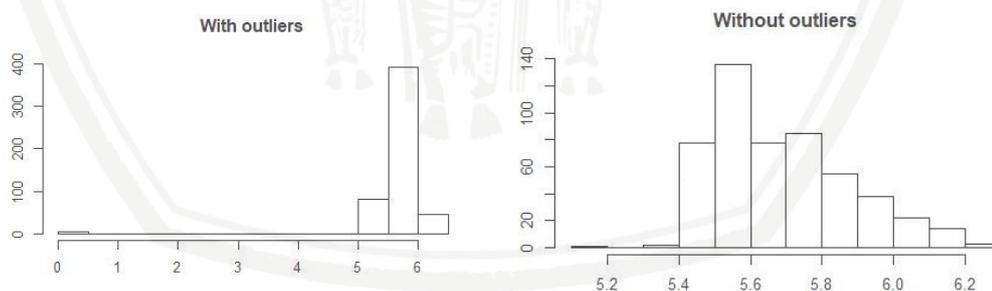


Gambar 6.4 Histogram Nilai Variabel *Monetary* Sebelum dan Sesudah Transformasi Data

Dari Gambar 6.1 hingga 6.4 dapat dilihat nilai sebelum transformasi data dan hasil setelah transformasi data. Hasil dari transformasi variabel *Length* mengubah rentang nilai menjadi 0 hingga 2,5. Hasil dari transformasi variabel *Recency* mengubah rentang nilai menjadi 0 hingga 2,5. Hasil dari transformasi variabel *Frequency* mengubah rentang nilai menjadi 0 hingga 0,8. Dan hasil transformasi variabel *Monetary* mengubah rentang nilai menjadi 5,2 hingga 6,4.

6.2.2 Data Cleaning

Hasil yang didapat dari grafik histogram adalah tidak terdapat *outliers* pada variabel *Recency*. Sedangkan pada variabel *Monetary* terdapat 1,9% atau sebesar 10 *outliers* berdasarkan Gambar 6.5. Variabel *Length* dan *Frequency* tidak dilakukan penghapusan *outliers* karena nilai yang menguntungkan jumlahnya hanya 11,8% atau sebesar 62 pelanggan dan jika dilakukan, maka nilai yang menguntungkan akan dianggap *outliers*. Gambar 6.5 merupakan perbandingan ada dan tidaknya *outliers* pada variabel *Monetary*.



```
Outliers identified: 10 from 522 observations
Proportion (%) of outliers: 1.91570881226054
Mean of the outliers: 3.80319336062655
Mean without removing outliers: 5.6481194211478
Mean if we remove outliers: 5.68415313326736
Do you want to remove outliers and to replace with NA? [yes/no]: yes
outliers successfully removed
```

Gambar 6.5 Histogram Perbandingan Ada dan Tidaknya *Outliers* pada Variabel *Monetary*

Data yang mengandung *outliers* akan diubah nilainya menjadi *NA* oleh sistem seperti dalam Tabel 6.6.

Tabel 6.6 Data Hasil Proses Penghapusan *Outliers*

| Id_pelanggan | <i>Length</i> | <i>Recency</i> | <i>Frequency</i> | <i>Monetary</i> |
|--------------|---------------|----------------|------------------|-----------------|
| 68 | 2.037462 | 2 | 0.6020600 | NA |
| 271 | 1.7242759 | 1.939519 | 0.8450980 | NA |
| 296 | 1.707570 | 1.875061 | 0.6989700 | NA |
| 337 | 0 | 1.919078 | 0.6989700 | NA |
| 357 | 1.623249 | 1.832509 | 0.9030900 | NA |
| 506 | 1.505150 | 1.968483 | 0.8450980 | NA |
| 523 | 0 | 2.394452 | 0.3010300 | NA |
| 524 | 0 | 2.369216 | 0.3010300 | NA |
| 525 | 0 | 1.643453 | 0.3010300 | NA |
| 526 | 0 | 1.447158 | 0.3010300 | NA |

Setelah mengetahui *outliers* berupa nilai *NA*, dilakukan penghapusan pada data tersebut agar sistem dapat membaca posisi data. Sisa data yang tidak terhapus atau tidak mengandung *outliers* sebesar 512 pelanggan.

6.2.3 Normalisasi Min-Max

Metode Min-Max merupakan metode normalisasi data yang mengubah rentang nilai menjadi 0 hingga 1. Selain itu, penggunaan normalisasi Min-Max dapat membuat nilai antar variabel menjadi seragam untuk dilakukan proses *clustering*. Normalisasi pada variabel *Length* dapat dilihat dalam Tabel 6.7.

Tabel 6.7 Hasil Nilai Normalisasi Data Terhadap Variabel *Length*

| Id_pelanggan | <i>Length</i> |
|--------------|---------------|
| 1 | 0.9041649 |
| 37 | 0.9307470 |
| 126 | 0 |
| 211 | 0.6819495 |

Selanjutnya, Tabel 6.8 merupakan nilai hasil dari normalisasi data terhadap variabel *Recency*.

Tabel 6.8 Hasil Nilai Normalisasi Data Terhadap Variabel *Recency*

| Id_pelanggan | <i>Recency</i> |
|--------------|----------------|
| 1 | 0.45488181 |
| 37 | 0.4171181 |
| 126 | 0.7351768 |
| 211 | 0.5831792 |

Setelah mendapatkan nilai normalisasi dari variabel *Recency*, dilakukan pengurangan 1 dengan nilai *Recency* karena berbanding terbalik dengan ketiga variabel lainnya. Hasil dari 1 dikurangi nilai *Recency* terdapat dalam Tabel 6.9.

Tabel 6.9 Hasil Pengurangan Normalisasi *Recency*

| Id_pelanggan | <i>Recency</i> (1- <i>Recency</i>) |
|--------------|-------------------------------------|
| 1 | 0.5451182 |
| 37 | 0.5828819 |
| 126 | 0.2648232 |
| 211 | 0.4168208 |

Hasil dari normalisasi pada variabel *Frequency* ditunjukkan dalam Tabel 6.10.

Tabel 6.10 Hasil Nilai Normalisasi Data Terhadap Variabel *Frequency*

| Id_pelanggan | <i>Frequency</i> |
|--------------|------------------|
| 1 | 0.4425070 |
| 37 | 1 |
| 126 | 0 |
| 211 | 0.7564708 |

Hasil normalisasi data terakhir dilakukan terhadap variabel *Monetary* yang dapat dilihat dalam Tabel 6.11.

Tabel 6.11 Hasil Nilai Normalisasi Data Terhadap Variabel *Monetary*

| Id_pelanggan | <i>Monetary</i> |
|--------------|-----------------|
| 1 | 0.9311009 |
| 37 | 1 |
| 126 | 0.5707637 |
| 211 | 0.7898489 |

Setelah mendapatkan semua data yang dinormalisasi, data tersebut digabungkan menjadi satu seperti dalam Tabel 6.12.

Tabel 6.12 Tabel Penggabungan dari Normalisasi LRFM

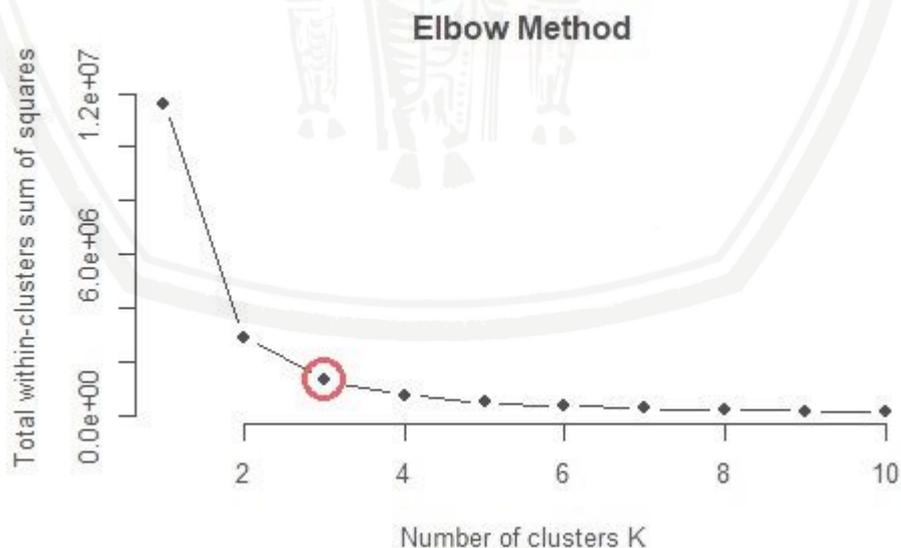
| Id_pelanggan | Length | Recency (1-Recency) | Frequency | Monetary |
|--------------|-----------|---------------------|-----------|-----------|
| 1 | 0.9041649 | 0.5451182 | 0.4425070 | 0.9311009 |
| 37 | 0.9307470 | 0.5828819 | 1 | 1 |
| 126 | 0 | 0.2648232 | 0 | 0.5707637 |
| 211 | 0.6819495 | 0.4168208 | 0.7564708 | 0.7898489 |

6.3 Clustering

Sub bab ini menjelaskan hasil dan analisis dari penentuan *cluster* dengan metode *Elbow*, dan *clustering* menggunakan *Fuzzy C-Means* pada 2 *cluster* dan 3 *cluster*.

6.3.1 Elbow Method

Penentuan jumlah *cluster* menggunakan metode *Elbow*. Hasil *cluster* ditunjukkan melalui grafik dalam Gambar 6.6. Pada gambar tersebut ditunjukkan penurunan drastic terakhir terletak diantara 2 dan 4, dimana titik tersebut menunjukkan angka 3 yang berarti hasil dari implementasi dari metode *elbow* menghasilkan 3 *cluster*. Untuk pengimplementasian metode *clustering*, jumlah *cluster* yang digunakan adalah 2 dan 3 *cluster* untuk membandingkan validitas dan performa *cluster* yang lebih baik.



Gambar 6.6 Hasil *Elbow Method*

Root Mean Squared Deviations (RMSD): 0.5922244
 Mean Absolute Deviation (MAD): 3.669139

Membership degrees matrix (top and bottom 5 rows):

```

  cluster 1 cluster 2
1 0.85170397 0.1482960
2 0.68586565 0.3141344
3 0.04741055 0.9525895
4 0.15579344 0.8442066
5 0.04925780 0.9507422
...
  cluster 1 cluster 2
508 0.02872648 0.9712735
509 0.55112990 0.4488701
510 0.01662394 0.9833761
511 0.16864031 0.8313597
512 0.15601375 0.8439862
  
```

Descriptive statistics for the membership degrees by clusters

| | Size | Min | Q1 | Mean | Median | Q3 | Max |
|-----------|------|-----------|-----------|-----------|-----------|-----------|-----------|
| Cluster 1 | 91 | 0.5030158 | 0.6409733 | 0.7756401 | 0.8470184 | 0.8942425 | 0.9854205 |
| Cluster 2 | 421 | 0.5185518 | 0.9108978 | 0.9297822 | 0.9515221 | 0.9777370 | 0.9995211 |

Dunn's Fuzziness Coefficients:

```

dunn_coeff normalized
0.8459982 0.6919964
  
```

within cluster sum of squares by cluster:

```

      1      2
17.69570 23.73449
(between_ss / total_ss = 53.46%)
  
```

Gambar 6.7 Hasil 2 Cluster

Tabel 6.13 Hasil 2 Cluster

| Id_pelanggan | Length | Recency (1-Recency) | Frequency | Monetary | Cluster |
|--------------|-----------|---------------------|-----------|-----------|---------|
| 1 | 0.9041649 | 0.5451182 | 0.4425070 | 0.9311009 | 1 |
| 37 | 0.9307470 | 0.5828819 | 1 | 1 | 1 |
| 126 | 0 | 0.2648232 | 0 | 0.5707637 | 2 |
| 211 | 0.6819495 | 0.4168208 | 0.7564708 | 0.7898489 | 1 |

6.3.2.2 Hasil Fuzzy C-Means 3 Segmen

Gambar 6.8 merupakan hasil dari 3 cluster dari implementasi Fuzzy C-Means. Hasil tersebut menunjukkan centroid awal dan akhir dari variabel Inorm sebagai Length, rnormbaru sebagai Recency, fnorm sebagai Frequency, dan mnorm sebagai Monetary. Selain itu, terdapat derajat keanggotaan dari setiap data pada cluster 1, 2, dan 3. Setiap data dapat dikatakan termasuk dalam cluster dengan melihat derajat keanggotaan yang mendekati nilai 1. Root Mean Squared Deviations (RMSD) memiliki nilai 0,403134 yang merupakan standar deviasi dari perbedaan antara nilai prediksi dengan nilai yang diamati. Berikutnya terdapat nilai Mean Absolute Deviations (MAD) sebesar 2,187867 yang berarti jarak rata-rata setiap titik dengan titik tengah. Hasil selanjutnya menunjukkan bahwa cluster 1 terdapat 301 anggota, cluster 2 terdapat 149 anggota, dan cluster 3 terdapat 62

Gambar 6.8 Hasil 3 Cluster

Tabel 6.14 Hasil 3 Cluster

| Id_pelanggan | Length | Recency (1-Recency) | Frequency | Monetary | Cluster |
|--------------|-----------|---------------------|-----------|-----------|---------|
| 1 | 0.9041649 | 0.5451182 | 0.4425070 | 0.9311009 | 3 |
| 37 | 0.9307470 | 0.5828819 | 1 | 1 | 3 |
| 126 | 0 | 0.2648232 | 0 | 0.5707637 | 1 |
| 211 | 0.6819495 | 0.4168208 | 0.7564708 | 0.7898489 | 3 |

6.4 Uji Validitas Cluster

Tahap ini adalah pengujian dari validitas *cluster* dengan menggunakan 3 metode pengujian, yaitu *Modified Partition Coefficient (MPC)*, *Davies Bouldin's Index*, dan package *cValid* yang didalamnya terdapat perbandingan nilai *connectivity*, *dunn index*, dan *silhouette width*.

6.4.1 Modified Partition Coefficient (MPC)

Pengujian dengan menggunakan metode *Modified Partition Coefficient (MPC)* dilakukan terhadap 2 dan 3 *cluster*. Script 6.1 adalah *script* pengujian MPC pada 2 *cluster*.

```
> hasil2 <- ppclust2(r2cluster, "fclust")
[1] 0.691996359151531
```

Script 6.1 Validitas MPC 2 Cluster dan Hasilnya

Hasil yang diperoleh pada 2 *cluster* dengan menggunakan MPC menghasilkan nilai sebesar 0,691996359151531. Script 6.2 adalah *script* pengujian MPC pada 3 *cluster*.

```
> hasil3 <- ppclust2(r3cluster, "fclust")
[1] 0.559808864881957
```

Script 6.2 Validitas MPC 3 Cluster dan Hasilnya

Hasil yang diperoleh pada 3 *cluster* dengan menggunakan MPC menghasilkan nilai sebesar 0,559808864881957. Hasil terbaik diperoleh *cluster* dua karena memiliki nilai yang lebih mendekati angka 1 yang berarti *cluster* tersebut memiliki kecenderungan nilai yang tidak monoton pada setiap data.

6.4.2 Davies Bouldin's Index

Pengujian kedua adalah dengan *DB Index* atau *Davies Bouldin's Index* dengan hasil seperti dalam Tabel 6.15.

Tabel 6.15 Nilai Hasil dari *DB Index*

| Total Cluster | <i>Davies Bouldin's Index</i> |
|---------------|-------------------------------|
| 2 | 0.86 |
| 3 | 1.04 |
| 4 | 0.9 |
| 5 | 0.91 |
| 6 | 0.92 |
| 7 | 1.08 |
| 8 | 1.088 |
| 9 | 1.07 |
| 10 | 1.05 |

Hasil dari tabel diatas menunjukkan bahwa *cluster* 2 adalah *cluster* terbaik dengan nilai 0.86 menurut *Davies Bouldin's Index* karena memiliki nilai terendah yang berarti *cluster* tersebut memiliki jarak data satu sama lain yang paling sesuai.

6.4.3 CValid

Pengujian terakhir menggunakan *packages* *cValid* yang terdapat dalam tools R yang berisi perbandingan nilai dari *Connectivity*, *Dunn Index*, dan *Silhouette Width*. Tabel 6.16 menunjukkan nilai dari hasil implementasi *packages* *CValid* pada 2 hingga 10 *cluster*.

Tabel 6.16 Nilai Hasil dari *Connectivity*, *Dunn Index*, dan *Silhouette Width*

| Total Cluster | <i>Connectivity</i> | <i>Dunn Index</i> | <i>Silhouette</i> |
|---------------|---------------------|-------------------|-------------------|
| 2 | 0 | 0.4 | 0.61 |
| 3 | 8.38 | 0.12 | 0.56 |
| 4 | 10.71 | 0.16 | 0.549 |
| 5 | 29.96 | 0.04 | 0.542 |
| 6 | 45.89 | 0.014 | 0.472 |
| 7 | 48.03 | 0.016 | 0.46 |
| 8 | 54.21 | 0.016 | 0.471 |
| 9 | 68.7 | 0.027 | 0.373 |
| 10 | 73.81 | 0.028 | 0.371 |

Connectivity menghasilkan nilai dengan rentang 0 hingga tidak terbatas, *Dunn Index* menghasilkan nilai dengan rentang 0 hingga 1, dan *Silhouette Width*

menghasilkan nilai dengan rentang -1 hingga 1. Hasil dari tabel diatas menunjukkan *cluster 2* adalah *cluster* terbaik karena sesuai dengan kriteria pengujian *Connectivity*, *Dunn Index*, dan *Silhouette Width*, dimana nilai *Connectivity* semakin mendekati 0, maka data berada pada *cluster* yang tepat sedangkan semakin jauh dari 0, maka data berada *cluster* yang kurang tepat. Selain itu, nilai *Dunn Index* dilihat dari hasil yang mendekati nilai 1 karena semakin tinggi maka rasio jarak akan semakin kecil antara data pada *cluster* yang sama dengan *cluster* yang berbeda. Dan hasil *Silhouette Width* dilihat dari hasil yang mendekati 1 dengan alasan semakin tinggi tingkat kepercayaan *cluster* yang terkelompok dan tidak terkelompok.

6.5 Uji Verifikasi

Tahap ini terbagi menjadi dua, yaitu uji verifikasi pada 2 dan 3 *cluster* untuk setiap data terhadap nilai *Euclidean Distance* dan nilai derajat keanggotaan.

6.5.1.1 Verifikasi 2 Cluster

Pengujian pertama menggunakan metode perhitungan *Euclidean Distance* mencari *cluster* yang tepat pada data pelanggan berdasarkan titik data dengan titik pusat *cluster*. Hasil uji dapat dilihat dalam Tabel 6.17.

Tabel 6.17 Verifikasi *Euclidean Distance* 2 Cluster

| Id_pelanggan | Fuzzy C-Means | Jarak data ke <i>centroid</i> | | Nilai Minimal | Hasil Validasi |
|--------------|---------------|-------------------------------|----------|---------------|----------------|
| | | 1 | 2 | | |
| 1 | 1 | 0.229169 | 1.315027 | 0.229169 | 1 |
| 37 | 1 | 0.471145 | 2.188278 | 0.471145 | 1 |
| 126 | 2 | 0.640156 | 0.037135 | 0.037135 | 2 |
| 211 | 1 | 0.115408 | 1.197627 | 0.115408 | 1 |

Hasil pada id_pelanggan 1 merupakan bagian dari *cluster* pertama berdasarkan hasil implementasi *Fuzzy C-Means*. Kemudian verifikasi pelanggan tersebut dilihat dari jarak terdekat dengan *centroid*. Perbandingan id_pelanggan 1 pada *cluster* pertama adalah 0,229169 dan pada *cluster* kedua adalah 1,315027, sehingga pelanggan tersebut termasuk ke dalam *cluster* pertama karena jarak ke *centroid* lebih dekat. Pengecekan selanjutnya dilakukan pada hasil verifikasi dengan hasil *Fuzzy C-Means* menggunakan *Script 6.1*.

```
> identical(hasil2cluster$cluster2, hasil2cluster$hasilcek
[1] TRUE
```

Script 6.3 Verifikasi 2 Cluster Hasil Pengecekan dengan *Fuzzy C-Means*

Hasil implementasi *function* *identical* adalah TRUE yang berarti bahwa hasil verifikasi dengan melihat perhitungan *Euclidean Distance* dengan hasil



implementasi *Fuzzy C-Means* bernilai sama dan dapat diartikan penempatan *cluster* telah sesuai.

6.5.1.2 Verifikasi 3 Cluster

Pengujian pertama menggunakan metode perhitungan *Euclidean Distance* mencari *cluster* yang tepat pada data pelanggan berdasarkan titik data dengan titik pusat *cluster*. Hasil uji dapat dilihat dalam Tabel 6.18.

Tabel 6.18 Verifikasi *Euclidean Distance* 3 Cluster

| Id_pelanggan | Fuzzy C-Means | Jarak data ke <i>centroid</i> | | | Nilai Minimal | Hasil Validasi |
|--------------|---------------|-------------------------------|----------|----------|---------------|----------------|
| | | 1 | 2 | 3 | | |
| 1 | 1 | 1.366808 | 1.24758 | 0.133233 | 0.133233 | 3 |
| 37 | 1 | 2.24913 | 2.076658 | 0.310683 | 0.310683 | 3 |
| 126 | 2 | 0.022947 | 0.160713 | 0.84939 | 0.022947 | 1 |
| 211 | 1 | 1.219531 | 1.194137 | 0.0489 | 0.0489 | 3 |

Hasil pada id_pelanggan 1 merupakan bagian dari *cluster* pertama berdasarkan hasil implementasi *Fuzzy C-Means*. Kemudian verifikasi pelanggan tersebut dilihat dari jarak terdekat dengan *centroid*. Perbandingan id_pelanggan 1 pada *cluster* pertama adalah 1,366808, *cluster* kedua adalah 1,24758, dan *cluster* ketiga adalah 0,133233, sehingga pelanggan tersebut termasuk ke dalam *cluster* ketiga karena jarak ke *centroid* lebih dekat. Pengecekan selanjutnya dilakukan pada hasil verifikasi dengan hasil *Fuzzy C-Means* menggunakan *Script* 6.1.

```
> identical(hasil3cluster$cluster3, hasil3cluster$hasilcek)
[1]TRUE
```

Script 6.4 Verifikasi 3 Cluster Hasil Pengecekan dengan *Fuzzy C-Means*

Hasil implementasi *function* *identical* adalah TRUE yang berarti bahwa hasil verifikasi dengan melihat perhitungan *Euclidean Distance* dengan hasil implementasi *Fuzzy C-Means* bernilai sama dan dapat diartikan penempatan *cluster* telah sesuai.

6.6 Analisis Hasil *Clustering*

Sub bab ini berisi analisis dari 2 *cluster* dan 3 *cluster* berdasarkan nilai LRFM dan segmen yang dihasilkan berupa perbandingan antara klasifikasi segmen dengan rincian nilai LRFM. Analisis hasil *clustering* menghasilkan peringkat setiap *cluster*. Peringkat tertinggi pada setiap *cluster* memiliki nilai *Length*, *Frequency*, dan *Monetary* tertinggi dan *Recency* terendah. Analisa hasil 2 *cluster* menunjukkan bahwa *cluster* 1 adalah kelompok pelanggan yang berpotensi sedangkan analisa hasil 3 *cluster* menunjukkan bahwa *cluster* 2 adalah kelompok pelanggan yang



berpotensi. Penjelasan lengkap dari analisis hasil *clustering* dapat dilihat dalam subbab 6.6.1 dan 6.6.2.

6.6.1 Analisis 2 Cluster

Data yang digunakan pada analisis *cluster* adalah data denormalisasi ke bentuk awal. Data yang dicari adalah nilai rentang dan rata-rata dari setiap variabel LRFM dan dikelompokkan berdasarkan *cluster*.

6.6.1.1 Analisis Cluster 1

Analisis lanjutan dari *cluster* pertama ditunjukkan dalam Tabel 6.19.

Tabel 6.19 Analisis Cluster 1

| Cluster | | 1 |
|----------------|----------------------------|-----------------------|
| Jumlah Anggota | | 91 |
| Karakteristik | Rentang <i>Length</i> | 0-256 hari |
| | Rentang <i>Recency</i> | 19-250 hari |
| | Rentang <i>Frequency</i> | 2-4 kali |
| | Rentang <i>Monetary</i> | 538000-1683000 rupiah |
| | Rata-rata <i>Length</i> | 38 hari |
| | Rata-rata <i>Recency</i> | 94 hari |
| | Rata-rata <i>Frequency</i> | 2 kali |
| | Rata-rata <i>Monetary</i> | 997385 rupiah |
| Peringkat | | 1 |

Dari tabel tersebut dapat dilihat bahwa *cluster* 1 merupakan *cluster* dengan peringkat pertama pada penelitian 2 segmen. *Cluster* 1 memiliki anggota atau pelanggan lebih sedikit dibandingkan dengan *cluster* 2, yaitu 91 pelanggan. Anggota *cluster* 1 memiliki rata-rata 38 hari pada variabel *Length* yang menunjukkan bahwa rentang pembelian pertama dan terakhir memiliki jarak yang tinggi karena dalam periode Juli 2017 hingga Maret 2018 anggota sudah bertransaksi lebih dari satu kali. Selanjutnya, rata-rata *Recency* pada *cluster* 1 adalah 94 hari yang menunjukkan waktu yang untuk pembelian terakhir dari pelanggan, yaitu 3 bulan sebelum penelitian dilakukan. Hasil lain menunjukkan bahwa rata-rata *Frequency* adalah 2 kali. Hal tersebut menunjukkan bahwa setiap pelanggan dalam *cluster* 1 melakukan transaksi minimal 2 kali dalam periode Juli 2017 hingga Maret 2018. Hasil terakhir adalah rata-rata *Monetary*, yaitu 997385 rupiah yang menunjukkan bahwa nominal yang dikeluarkan setiap pelanggan terbilang tinggi untuk mendapatkan produk dari Ride Inc.

6.6.1.2 Analisis Cluster 2

Analisis lanjutan dari *cluster* kedua ditunjukkan dalam Tabel 6.20.

Tabel 6.20 Analisis *Cluster 2*

| <i>Cluster</i> | | 2 |
|----------------|----------------------------|----------------------|
| Jumlah Anggota | | 421 |
| Karakteristik | Rentang <i>Length</i> | 0-1 hari |
| | Rentang <i>Recency</i> | 18-447 hari |
| | Rentang <i>Frequency</i> | 1-2 kali |
| | Rentang <i>Monetary</i> | 140000-855000 rupiah |
| | Rata-rata <i>Length</i> | 0.0023 hari |
| | Rata-rata <i>Recency</i> | 138 hari |
| | Rata-rata <i>Frequency</i> | 1 kali |
| | Rata-rata <i>Monetary</i> | 444783 rupiah |
| Peringkat | | 2 |

Dari tabel tersebut dapat dilihat bahwa cluster 2 merupakan cluster dengan peringkat pertama pada penelitian 2 segmen. Cluster 2 memiliki anggota atau pelanggan lebih banyak dibandingkan dengan cluster 2, yaitu 421 pelanggan. Anggota cluster 2 memiliki rata-rata 0.0023 hari pada variabel length yang menunjukkan bahwa rentang pembelian pertama dan terakhir memiliki jarak yang sangat rendah karena dalam periode Juli 2017 hingga Maret 2018 anggota hanya melakukan transaksi satu kali dan sangat sedikit pelanggan yang memiliki rentang 1 hari. Selanjutnya, rata-rata recency pada cluster 2 adalah 138 hari yang menunjukkan waktu yang lama untuk pembelian terakhir dari pelanggan, yaitu 4 bulan sebelum penelitian dilakukan. Hasil lain menunjukkan bahwa rata-rata frequency adalah 1 kali. Hal tersebut menunjukkan bahwa setiap pelanggan dalam cluster 2 melakukan transaksi hanya 1 kali dalam periode Juli 2017 hingga Maret 2018. Hasil terakhir adalah rata-rata monetary, yaitu 444783 rupiah yang menunjukkan bahwa nominal yang dikeluarkan pelanggan terbilang rendah untuk mendapatkan produk dari Ride Inc.

6.6.2 Analisis 3 *Cluster*

Data yang digunakan pada analisis 3 *cluster* adalah data denormalisasi ke bentuk awal. Data yang dicari adalah nilai rentang dan rata-rata dari setiap variabel LRFM dan dikelompokkan berdasarkan *cluster*.

6.6.2.1 Analisis *Cluster 1*

Analisis *cluster* pertama pada penelitian 3 segmen ditunjukkan dalam Tabel 6.21.

Tabel 6.21 Analisis Cluster 1

| | | |
|----------------|----------------------------|-----------------------|
| <i>Cluster</i> | | 1 |
| Jumlah Anggota | | 301 |
| Karakteristik | Rentang <i>Length</i> | 0-1 hari |
| | Rentang <i>Recency</i> | 82-447 hari |
| | Rentang <i>Frequency</i> | 1-2 kali |
| | Rentang <i>Monetary</i> | 140000-1030000 rupiah |
| | Rata-rata <i>Length</i> | 0.009 hari |
| | Rata-rata <i>Recency</i> | 175 hari |
| | Rata-rata <i>Frequency</i> | 1 kali |
| Peringkat | | 3 |

Dari tabel tersebut dapat dilihat bahwa cluster 1 merupakan cluster dengan anggota atau pelanggan terbanyak, yaitu 301 pelanggan. Anggota cluster 1 memiliki rata-rata 0.009 hari pada variabel length yang menunjukkan bahwa rentang pembelian pertama dan terakhir memiliki jarak yang sangat rendah karena dalam periode Juli 2017 hingga Maret 2018 anggota hanya melakukan transaksi satu kali dan sangat sedikit pelanggan yang memiliki rentang 1 hari atau bisa disimpulkan hanya sedikit pelanggan yang melakukan transaksi lebih dari satu kali. Selanjutnya, rata-rata recency pada cluster 1 adalah 175 hari yang menunjukkan waktu yang lama untuk pembelian terakhir dari pelanggan, yaitu hampir 6 bulan sebelum penelitian dilakukan. Hasil lain menunjukkan bahwa rata-rata frequency adalah 1 kali pada periode tersebut. Hasil terakhir adalah rata-rata monetary, yaitu 448286 rupiah yang menunjukkan bahwa nominal yang dikeluarkan setiap pelanggan terbilang rendah untuk mendapatkan produk dari Ride Inc. Dari nilai LRFM tersebut, *cluster 1* menduduki peringkat ketiga pada penelitian 3 segmen karena memiliki nilai *Recency* yang lebih tinggi dan *Monetary* lebih rendah dari *cluster 3*.

6.6.2.2 Analisis Cluster 2

Analisis *cluster* kedua pada penelitian 3 segmen ditunjukkan dalam Tabel 6.22.

Tabel 6.22 Analisis Cluster 2

| | | |
|----------------|------------------------|-------------|
| <i>Cluster</i> | | 2 |
| Jumlah Anggota | | 149 |
| Karakteristik | Rentang <i>Length</i> | 0-256 hari |
| | Rentang <i>Recency</i> | 19-234 hari |

| | | |
|-----------|----------------------------|-----------------------|
| | Rentang <i>Frequency</i> | 2-4 kali |
| | Rentang <i>Monetary</i> | 538000-1683000 rupiah |
| | Rata-rata <i>Length</i> | 57 hari |
| | Rata-rata <i>Recency</i> | 98 hari |
| | Rata-rata <i>Frequency</i> | 2 kali |
| | Rata-rata <i>Monetary</i> | 996118 rupiah |
| Peringkat | | 1 |

Dari tabel tersebut dapat dilihat bahwa cluster 2 memiliki anggota atau pelanggan sebesar 149 pelanggan. Anggota cluster 1 memiliki rata-rata 57 hari pada variabel *length* yang menunjukkan bahwa rentang pembelian pertama dan terakhir memiliki jarak yang sangat tinggi karena dalam periode Juli 2017 hingga Maret 2018 anggota dapat melakukan transaksi minimal dua kali, yaitu pertama dan terakhir. Selanjutnya, rata-rata *recency* pada cluster 2 adalah 98 hari yang menunjukkan waktu yang tidak terlalu lama untuk pembelian terakhir dari pelanggan, yaitu 3 bulan sebelum penelitian dilakukan. Hasil lain menunjukkan bahwa rata-rata *frequency* adalah 2 kali transaksi setiap pelanggan pada periode tersebut. Hasil terakhir adalah rata-rata *monetary*, yaitu 996118 rupiah yang menunjukkan bahwa nominal yang dikeluarkan setiap pelanggan terbilang tinggi untuk mendapatkan produk dari Ride Inc. Dari nilai LRFM tersebut, *cluster 2* menduduki peringkat pertama pada penelitian 3 segmen karena memiliki nilai *Length*, *Frequency*, dan *Monetary* tertinggi yang berarti *cluster* kedua merupakan kelompok yang menguntungkan.

6.6.2.3 Analisis Cluster 3

Analisis *cluster* ketiga pada penelitian 3 segmen ditunjukkan dalam Tabel 6.23.

Tabel 6.23 Analisis Cluster 3

| | | |
|----------------|----------------------------|-----------------------|
| <i>Cluster</i> | | 3 |
| Jumlah Anggota | | 62 |
| Karakteristik | Rentang <i>Length</i> | 0-1 hari |
| | Rentang <i>Recency</i> | 18-108 hari |
| | Rentang <i>Frequency</i> | 1-2 kali |
| | Rentang <i>Monetary</i> | 240000-1560000 rupiah |
| | Rata-rata <i>Length</i> | 0.006 hari |
| | Rata-rata <i>Recency</i> | 57 hari |
| | Rata-rata <i>Frequency</i> | 1 kali |

| | | |
|-----------|---------------------------|---------------|
| | Rata-rata <i>Monetary</i> | 538878 rupiah |
| Peringkat | | 2 |

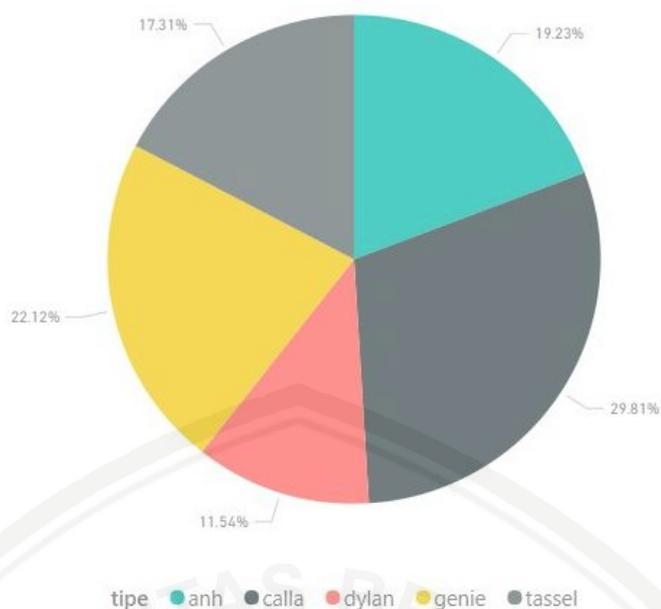
Dari tabel tersebut dapat dilihat bahwa cluster 3 memiliki anggota atau pelanggan terendah, yaitu 62 pelanggan. Anggota cluster 3 memiliki rata-rata 0.006 hari pada variabel length yang menunjukkan bahwa rentang pembelian pertama dan terakhir memiliki jarak yang sangat rendah karena dalam periode Juli 2017 hingga Maret 2018 anggota hanya melakukan transaksi satu kali dan sangat sedikit pelanggan yang memiliki rentang 1 hari atau bisa disimpulkan hanya sedikit pelanggan yang melakukan transaksi lebih dari satu kali. Selanjutnya, rata-rata recency pada cluster 1 adalah 57 hari yang menunjukkan waktu yang terbilang cepat untuk pembelian terakhir dari pelanggan, yaitu hampir 2 bulan sebelum penelitian dilakukan. Hasil lain menunjukkan bahwa rata-rata frequency adalah 1 kali transaksi setiap pelanggan pada periode tersebut. Hasil terakhir adalah rata-rata monetary, yaitu 538878 rupiah yang menunjukkan bahwa nominal yang dikeluarkan setiap pelanggan terbilang rendah untuk mendapatkan produk dari Ride Inc, tetapi lebih tinggi dibandingkan dengan cluster 1. Dari nilai LRFM tersebut, cluster 3 menduduki peringkat kedua pada penelitian 3 segmen karena memiliki nilai *Recency* terendah dan *Monetary* yang lebih tinggi dari cluster 1, tetapi memiliki nilai *Length*, *Frequency*, dan *Monetary* yang lebih rendah dari cluster 2.

6.7 Visualisasi Hasil *Clustering*

Visualisasi hasil *clustering* ditampilkan berupa grafik dan diagram untuk mempermudah Ride Inc. dalam pembacaannya. Hasil *clustering* terbaik sesuai uji validitas dan verifikasi *cluster* yang ditampilkan, yaitu 2 *cluster*. Terdapat tiga halaman visualisasi, yaitu halaman utama, halaman 2 *cluster*, dan halaman data 2 *cluster*.

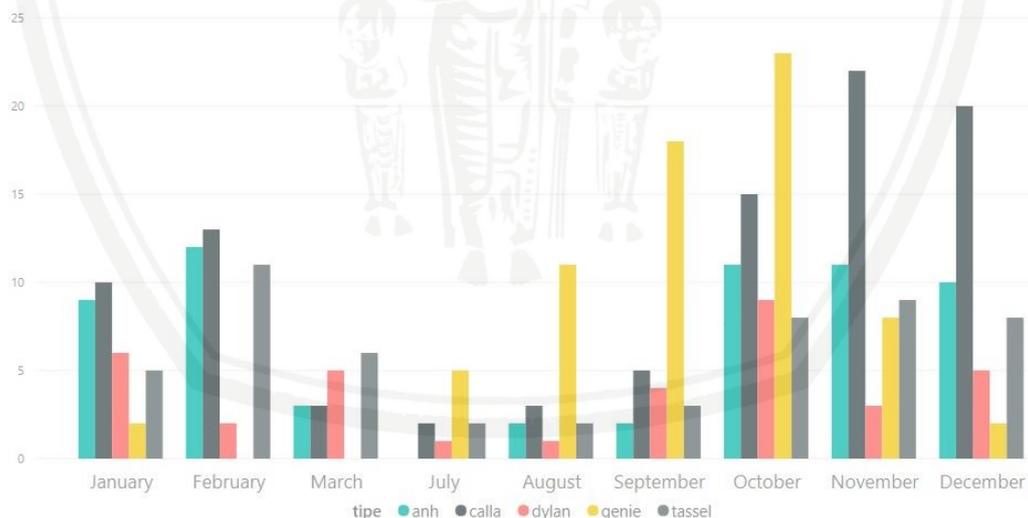
6.7.1 Halaman Utama

Gambar 6.9 menunjukkan diagram donat penjualan seluruh tipe sepatu dalam periode Juli 2017 hingga Maret 2018. Berdasarkan diagram tersebut, persentase tipe sepatu yang paling banyak dipesan adalah calla sebesar 14,59% atau sebanyak 96 transaksi. Transaksi terbanyak selanjutnya terdapat dalam tipe sepatu genie sebanyak 69 kali transaksi dan anh sebanyak 61 kali transaksi. Transaksi paling rendah terdapat dalam tipe avena, barre, moccasine, oxford dan jeslyn sebesar 0.15% atau sebanyak 1 transaksi.



Gambar 6.9 Persentase Total Setiap Tipe Sepatu

Selanjutnya, Gambar 6.10 menunjukkan diagram batang penjualan tipe sepatu setiap bulannya. Dari gambar dibawah, dapat diketahui bahwa tipe sepatu Genie merupakan penjualan tertinggi pada bulan Juli 2017 hingga Oktober 2017 dengan total 5, 11, 18, dan 23. Namun, tipe sepatu Calla yang baru tersedia dibulan Agustus 2017 menduduki penjualan terbanyak pada bulan November 2017 hingga Februari 2018 dengan total 22, 20, 10, dan 13. Penjualan terbanyak pada bulan Maret 2018 diperoleh tipe tassel dengan total 7 pasang sepatu.



Gambar 6.10 Diagram Batang Penjualan Tipe Sepatu Setiap Bulan

Terakhir, Gambar 6.11 menunjukkan grafik penjualan setiap bulan dalam periode Juli 2017 hingga Maret 2018. Dimulai bulan Juli, Ride Inc menghasilkan Rp6.544.000 dari total penjualan. Peningkatan tertinggi terjadi pada bulan

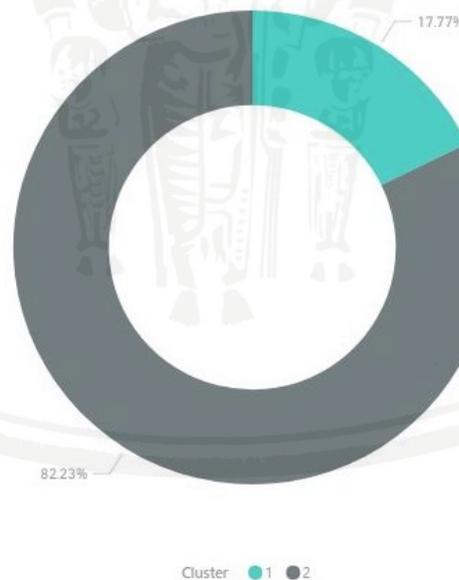
Oktober 2017 sebesar Rp28.885.975 dengan total Rp49.860.975. Dan penurunan terendah terjadi pada bulan November 2017 dengan selisih Rp11.800.975.



Gambar 6.11 Grafik Penjualan Setiap Bulan

6.7.2 Halaman 2 Cluster

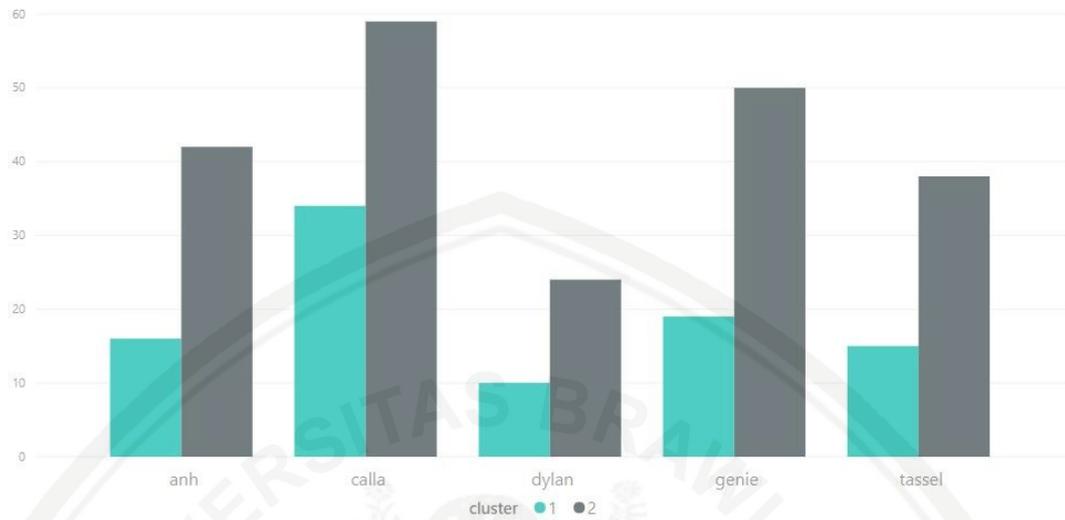
Gambar 6.12 menunjukkan diagram donat dari persentase pelanggan setiap cluster. Cluster terbagi menjadi 2, cluster pertama terdapat 17.77% atau sebesar 91 pelanggan. Dan cluster kedua terdapat 82.23% atau sebesar 421 pelanggan.



Gambar 6.12 Presentase 2 Cluster

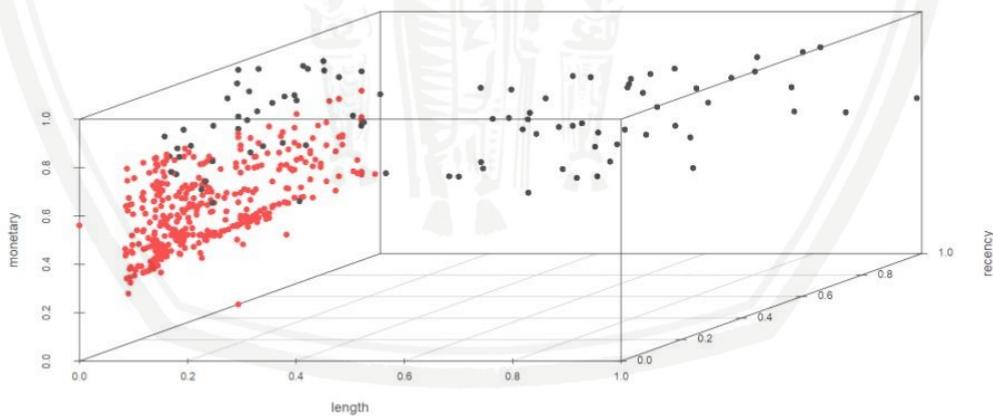
Selanjutnya terdapat diagram batang dari perbandingan jumlah transaksi setiap tipe sepatu berdasarkan cluster seperti pada Gambar 6.13. Dari gambar dibawah dapat dilihat perbedaan tertinggi terdapat tipe sepatu genie, yaitu 50 transaksi pada cluster 2 dan 19 pada cluster 1. Pembelian terbanyak pada tipe

sepatu dari *cluster 1* dan *cluster 2* adalah tipe *calla*, *cluster 1* sebanyak 34 dan *cluster 2* sebanyak 59. Selain itu, pembelian terendah pada *cluster 1* adalah 0 transaksi pada tipe sepatu *avena*, *barre*, *classic*, *jeslyn*, *jessie*, *lona*, *moccasin*, dan *oxford*. Dan pembelian terendah pada *cluster 2* adalah 0 transaksi pada tipe sepatu *alberta*, *hampton*, dan *model*.



Gambar 6.13 Diagram Batang Jumlah Transaksi Setiap Tipe Sepatu Berdasarkan Cluster

Terakhir, terdapat tampilan *scatterplot 3D* yang menunjukkan lokasi data berdasarkan *cluster* yang dengan perbedaan warna, yaitu merah dan hitam. Gambar 6.14 menampilkan *scatterplot 3D* dari variabel *Length*, *Recency*, dan *Monetary* (LRM) dan selain itu terdapat RFM dan LRF.



Gambar 6.14 Scatterplot 3D Variabel Length, Recency, dan Monetary

6.7.3 Halaman Data 2 Cluster

Gambar 6.15 menunjukkan diagram batang dari perbandingan 2 *cluster* terhadap variabel *Frequency* transaksi. Posisi *x-axis* dengan nilai 1,2,3, dan 4 menunjukkan total penjualan dan *y-axis* dengan nilai rentang 0 hingga 450 menunjukkan total pelanggan. *Cluster 1* tidak terdapat pelanggan yang melakukan



transaksi hanya 1 kali. Perbandingan pada *Frequency* yang kedua terdapat 69 pelanggan pada *cluster 1* dan 2 pelanggan pada *cluster 2*. Dan hasil *Frequency* ketiga dan keempat tidak terdapat pelanggan dari *cluster 2*.



Gambar 6.15 Diagram Batang Perbandingan 2 Cluster Berdasarkan *Frequency*

Selanjutnya terdapat tabel yang berisi urutan pelanggan terbaik berdasarkan variabel *Frequency*. Tabel tersebut menampilkan 20 pelanggan terbaik dan didalamnya tidak terdapat satupun pelanggan dari *cluster 2*. Pelanggan terbaik memiliki *Frequency* 3 dan 4 kali transaksi. Tabel pelanggan dapat dilihat dalam Gambar 6.16.

| id_pelanggan | frequency | monetary | cluster2 |
|--------------|-----------|----------|----------|
| 19 | 4 | 1600000 | 1 |
| 37 | 4 | 1683000 | 1 |
| 43 | 3 | 1335000 | 1 |
| 45 | 3 | 1163000 | 1 |
| 99 | 3 | 1070000 | 1 |
| 124 | 3 | 1254125 | 1 |
| 138 | 3 | 1028000 | 1 |
| 168 | 4 | 1484000 | 1 |
| 211 | 3 | 998000 | 1 |
| 212 | 3 | 1382000 | 1 |
| 227 | 3 | 790000 | 1 |
| 242 | 3 | 1100000 | 1 |

Gambar 6.16 Tabel Pelanggan Terbaik Berdasarkan *Frequency*

Selain tabel diatas, terdapat tabel yang berisi 20 pelanggan terbaik berdasarkan variabel *Monetary*. Semua pelanggan yang ditampilkan berasal dari *cluster 1* dengan *Monetary* tertinggi sebesar Rp1.683.000. Tabel pelanggan dapat dilihat dalam Gambar 6.17.

| id_pelanggan | frequency | monetary | cluster2 |
|--------------|-----------|----------|----------|
| 1 | 2 | 1418000 | 1 |
| 19 | 4 | 1600000 | 1 |
| 37 | 4 | 1683000 | 1 |
| 43 | 3 | 1335000 | 1 |
| 124 | 3 | 1254125 | 1 |
| 158 | 2 | 1318800 | 1 |
| 168 | 4 | 1484000 | 1 |
| 201 | 2 | 1291000 | 1 |
| 212 | 3 | 1382000 | 1 |
| 285 | 3 | 1450500 | 1 |
| 329 | 2 | 1296000 | 1 |
| 336 | 4 | 1276000 | 1 |

Gambar 6.17 Tabel Pelanggan Terbaik Berdasarkan *Monetary*

Terakhir, terdapat kesimpulan dari halaman data 2 *cluster* yang akan diisi manual berdasarkan hasil diagram dan tabel yang ditampilkan pada halaman tersebut. Kesimpulan dapat dilihat dalam Gambar 6.18.

Kesimpulan

Cluster 1 merupakan cluster terbaik karena memiliki rata-rata nilai length yang tinggi, recency terendah, frequency tertinggi dan monetary tertinggi dibandingkan cluster 2.

Gambar 6.18 Kesimpulan

6.8 Analisis Hasil *System Usability Scale (SUS)*

Pengujian *System Usability Scale (SUS)* terhadap visualisasi *dashboard* yang dilakukan pada pihak Ride Inc. menghasilkan 2 responden. Setelah mendapatkan skor akhir, selanjutnya melakukan perhitungan rata-rata menggunakan rumus dari Soleh (2005) yang dapat dilihat pada persamaan 6.1.

$$rata - rata = \sum_{i=1}^n \frac{x_i}{N} \dots (6.1)$$

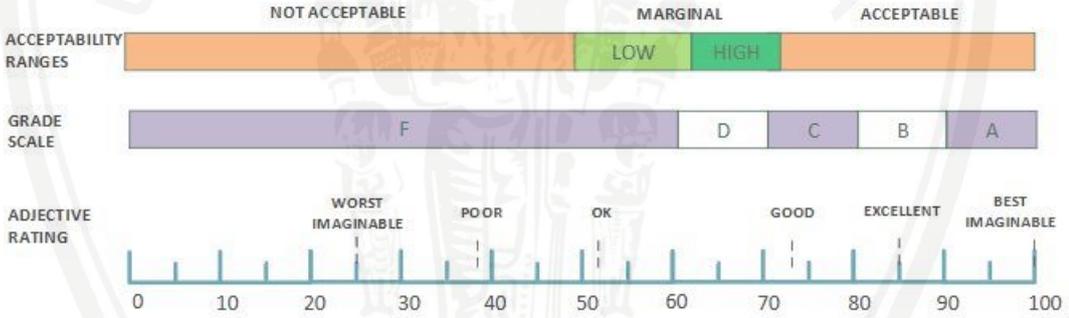


Nilai x_i merupakan nilai penjumlahan dari responden dan N merupakan jumlah responden. Hasil perhitungan dan rata-rata dari 2 responden dapat dilihat dalam Tabel 6.24.

Tabel 6.24 Perhitungan Hasil System Usability Scale (SUS)

| Responden | Pernyataan | | | | | | | | | | Skor SUS (Total * 2.5) |
|-----------|------------|---|---|---|---|---|---|---|---|----|------------------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| 1 | 3 | 1 | 4 | 0 | 4 | 3 | 3 | 3 | 3 | 2 | 65 |
| 2 | 3 | 2 | 2 | 2 | 3 | 4 | 2 | 3 | 3 | 2 | 65 |
| Rata-rata | | | | | | | | | | | 65 |

Menurut Brooke (2013) terdapat dua cara untuk menentukan tingkatan dari hasil, pertama dari tingkatan sisi pengguna dan tingkatan sisi skala. Tingkatan sisi pengguna memiliki tiga kategori yaitu *not acceptable*, *marginal*, dan *acceptable*. Sedangkan tingkatan sisi skala memiliki lima skala yaitu A atau memiliki skor lebih dari sama dengan 80.3, B atau memiliki skor lebih dari sama dengan 74 dan lebih kecil dari 80.3, C atau memiliki skor lebih besar sama dengan 68 dan lebih kecil dari 74, D atau memiliki skor sama dengan 51 dan lebih kecil dari 68, dan F lebih kecil dari 51. Gambaran tersebut dapat dilihat dalam Gambar 6.19.



Gambar 6.19 Penentuan Hasil

Sumber: Bangor, 2009

Hasil rata-rata yang diperoleh dari dua responden adalah 65. Menurut teori dari Brooke (2013) dan Gambar 6.19 skor 65 termasuk pada *acceptability range* kategori *marginal* tipe *high* yang berarti dapat diterima oleh pihak Ride Inc serta termasuk *grade scale* tipe D dan *adjective rating* tipe *good*. Namun, hasil tersebut tidak seperti saat melakukan presentasi visualisasi *dashboard* kepada pihak Ride Inc. dimana kedua responden sangat tertarik untuk segera bisa menggunakan sistem tersebut untuk membantu mereka dalam membuat strategi bisnis yang tepat untuk setiap produk yang akan dikeluarkan.



BAB 7 PENUTUP

7.1 Kesimpulan

Setelah mendapatkan hasil dan melakukan analisis terhadap hasil, maka kesimpulan yang didapat adalah:

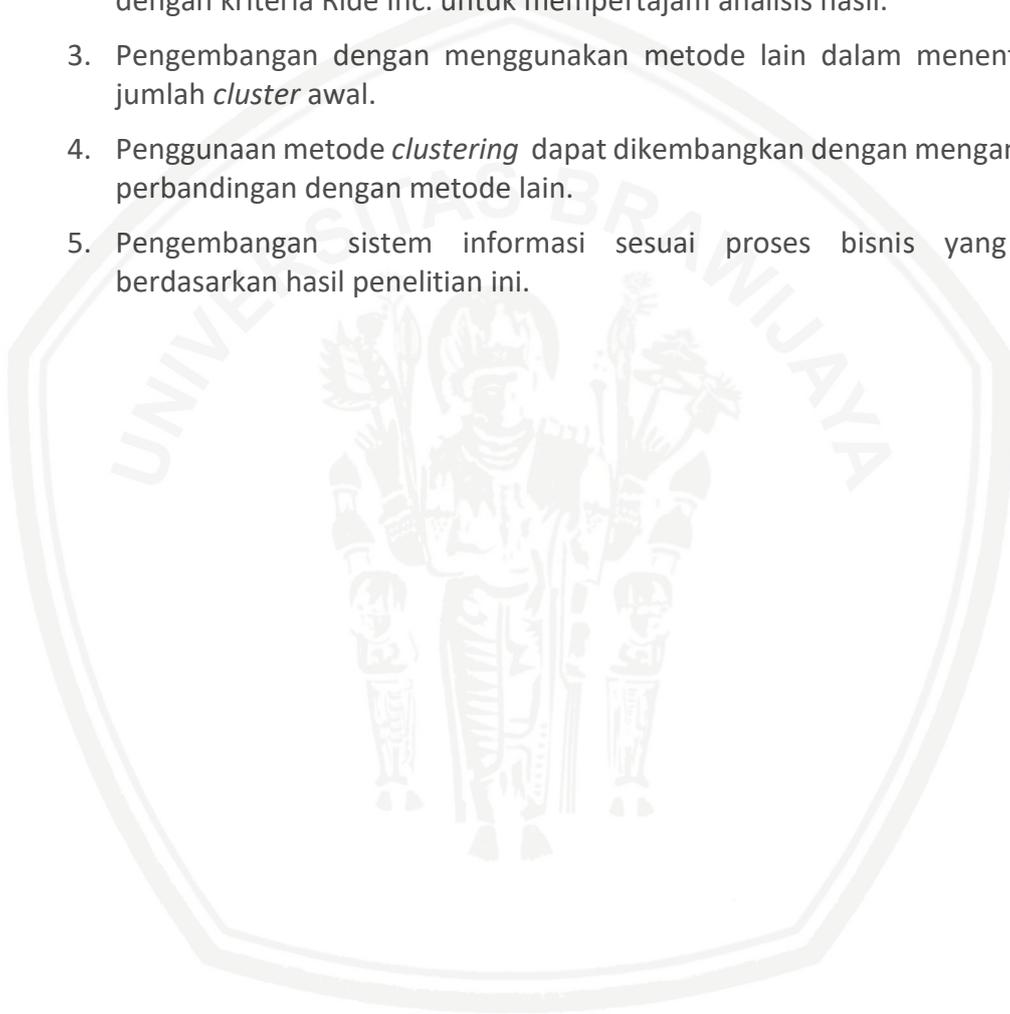
1. Hasil dari pemodelan LRFM dapat mendeskripsikan data pelanggan berdasarkan setiap segmen yang dapat dilihat pada perilaku setiap pelanggan pada setiap variabel *Length*, *Recency*, *Frequency*, dan *Monetary* dalam transaksi yang dilakukan di Ride Inc.
2. Metode *clustering* digunakan dengan melewati beberapa proses, yaitu data *preprocessing*, penentuan cluster dengan metode *elbow* dan implementasi *Fuzzy C-Means*. Data *preprocessing* meliputi data selection dari data pelanggan, data transformasi yang mengubah nilai menjadi stabil dan data normalisasi mengubah data transformasi menjadi normal dengan range 0 hingga 1. Setelah itu, penerapan metode *elbow* menunjukkan penurunan signifikan terakhir pada jumlah 3 cluster. Dan implementasi *Fuzzy C-Means* menggunakan jumlah 2 dan 3 cluster untuk mendapatkan perbandingan jumlah cluster terbaik.
3. Pengujian cluster terdiri dari uji validitas cluster dan uji verifikasi cluster. Tahap uji validitas cluster menggunakan metode *Modified Partition Coefficient*, *Davies Bouldin's Index*, dan packages *CVvalid* yang berisi *Connectivity*, *Dunn Index*, dan *Silhouette Width*. Hasil dari implementasi ketiga metode pengujian menunjukkan bahwa 2 cluster adalah hasil yang optimal. Namun, hasil dari metode *elbow* menunjukkan nilai 3 untuk jumlah cluster terhadap penurunan terakhir yang signifikan, sehingga proses *clustering* dilakukan dua kali dengan jumlah 2 cluster. Selanjutnya, penerapan uji verifikasi cluster dengan *Euclidean Distance* menghasilkan cluster yang sesuai dengan implementasi *Fuzzy C-Means*.
4. Visualisasi *dashboard* ditampilkan dalam bentuk interaktif dengan menggabungkan beberapa diagram dan grafik. Diagram yang ditampilkan berupa diagram donat dan diagram batang. Diagram donat menampilkan persentase dari total penjualan setiap tipe sepatu dan persentase dari cluster yang dihasilkan. Diagram batang menampilkan penjualan seluruh tipe sepatu setiap bulannya, jumlah transaksi setiap tipe sepatu berdasarkan cluster, dan total pelanggan berdasarkan variabel *frequency* pada setiap cluster. Dan grafik yang ditampilkan berupa total penjualan dari Ride Inc setiap bulannya pada periode Juli 2017 hingga Maret 2018 yang mengalami kenaikan dan penurunan. Selain itu, hal yang ditampilkan adalah scatterplot 3D dari LRM, RFM, serta LRF untuk melihat persebaran data dan tabel yang menunjukkan pelanggan terbaik berdasarkan *frequency* dan *monetary*. Dan dilakukan pengujian *usability* dengan *System Usability Scale (SUS)* terhadap visualisasi *dashboard*. Terdapat dua responden yaitu *founder* dan *co-founder* Ride Inc dan menghasilkan nilai

rata-rata 65 yang berarti visualisasi *dashboard* dapat diterima dengan kategori *marginal* tipe *high*.

7.2 Saran

Saran diambil dari hasil analisis yang telah dilakukan untuk rekomendasi terhadap penelitian selanjutnya.

1. Penelitian selanjutnya dapat dikembangkan dengan menambahkan data pelanggan Ride Inc. dan rentang waktu menjadi 1 tahun atau lebih.
2. Pengembangan penelitian dapat menambahkan variabel yang sesuai dengan kriteria Ride Inc. untuk mempertajam analisis hasil.
3. Pengembangan dengan menggunakan metode lain dalam menentukan jumlah *cluster* awal.
4. Penggunaan metode *clustering* dapat dikembangkan dengan menganalisis perbandingan dengan metode lain.
5. Pengembangan sistem informasi sesuai proses bisnis yang ada berdasarkan hasil penelitian ini.



DAFTAR PUSTAKA

- Afrin, Fahmida dkk. 2015. *Comparative Performance of Using PCA with K-Means and Fuzzy C Means Clustering for Customer Segmentation*. International Journal of Scientific and Technology Research, (Online), Volume 4, Issue 10.
- Angelie, Annisa Veronica. 2017. *Segmentasi Pelanggan Menggunakan Clustering K-Means dan Model RFM (Studi Kasus: PT.Bina Adidaya Surabaya)*. Skripsi. Sepuluh November Institute of Technology.
- Ansari, Azarnoush dan Riasi, Arash. 2016. *Customer Clustering Using a Combination of Fuzzy C-Means and Genetic Algorithms*. International Journal of Business and Management, (Online), Volume 11, Nomor 7, (www.ccsenet.org/ijbm).
- Astria, Dewi dan Suprayogi. 2017. *Penerapan Algoritma Fuzzy C-Means untuk Clustering Pelanggan pada CV Mataram Jaya Bawen*. Eksplora Informatika, (Online), Jurusan Teknik Informatika, Universitas Dian Nuswantoro Semarang.
- Bezdek, James. 1981. *Pattern Recognition with Fuzzy Objective Function*. New York: Plenum Press.
- Brock, Guy dkk. 2008. *clValid: An R Package for Cluster Validation*. Journal of Statistical Software, (Online), Volume 25, Issue 4, (<http://jstatsoft.org/>).
- Brooke, John. 1986. *SUS – A Quick and Dirty Usability Scale*. Redhatch Consulting, (Online), United Kingdom.
- Buttle, Francis. 2009. *Customer Relationship Management: Concepts and Technologies. Second Edition. (Ebook)*. Oxford: Elsevier.
- Dave, Rajesh N. 1996. *Validating Fuzzy Partitions Obtained Through C-Shells Clustering*. New Jersey Institute of Technology, Volume 17, Issue 6.
- Dunn, J. C. 1974. *Well-Separated Clusters and Optimal Fuzzy Partitions*. Journal of Cybernetics. 4(1). 58-73.
- Elliott, J. M. 1977. *Some Methods for The Statistical Analysis of Samples of Benthic Invertebrates. Second Edition*. Ambleside: Freshwater Biological Association.
- Few, Stephen. 2004. *Dashboard Confusion*. Perceptual Edge.
- Few, Stephen. 2006. *Information Dashboard Design. First Edition. (Ebook)*. Italy: O'Reilly Media, Inc.
- Han, Jiawei dan Kamber, Micheline. 2006. *Data Mining: Concepts and Technologies. Second Edition. (Ebook)*. Oxford: Elsevier.
- Han, Jiawei dkk. 2012. *Data Mining: Concepts and Technologies. Third Edition. (Ebook)*. Oxford: Elsevier.
- Hand, David dkk. 2001. *Principles of Data Mining*. Cambridge: The MIT Press.

- Handl, Julia dkk. 2005. *Computational Cluster Validation in Post-Genomic Data Analysis*. Oxford University Press. 21(05). 3201-3212.
- Hsin, Huang Wu dkk. 2014. *Analyzing Patients' Values by Applying Cluster Analysis and LRFM Model in a Padiatric Dental Clinic in Taiwan*. Hindawi Publishing Corporation, The Scientific World Journal, (Online), Volume 2014, (<http://dx.doi.org/10.1155/2014/685495>).
- James, C. Bezdek. 1974. *Cluster Validity with Fuzzy Sets*. Journal of Cybernetics, Volume 3, Issue 3.
- Karomi, M. Adib Al dan Darmawan, Wachid. 2016. *Meningkatkan Hasil Segmentasi Pelanggan Menggunakan Algoritma Fuzzy C-Means dan Outlier Removal Clustering*. IC-Tech, (Online), Volume XI, Nomor 2, (<http://jurnal.stmik-wp.ac.id>).
- Kodinariya, Trupti M. dan Makwana, Prashant R. 2013. *Review on Determining Number of Cluster in K-Means Clustering*. International Journal Advance Research in Computer Science and Management Studies, (Online), Volume 1, Issue 6, (www.ijarcsms.com).
- Loochach, Richa dan Garg, Kanwal. *Effect of Distance Functions on K-Means Clustering Algorithm*. International Journal of Computer Applications. 49(6).
- Malik, Shadan. 2005. *Enterprise Dashboards: Design and Best Practices for IT*. (Ebook). Canada: John Wiley & Sons, Inc.
- Rousseeuw, Peter J. *Silhouettes: A Graphical Aid To The Interpretation and Validation of Cluster Analysis*. Journal of Computational and Applied Mathematics. 53-65.
- Saitta, Sandro dkk. Tanpa Tahun. *A Bounded Index for Cluster Validity*. (Online). Switzerland: Ecole Polytechnique Federale de Lausanne (EPFL).
- Santosa, B. 2007. *Data Mining Teknik Pengumpulan Data Untuk Keperluan Bisnis*. Yogyakarta: Grahallmu.
- Sauro, Jeff. 2011. *Measuring Usability with The System Usability Scale (SUS)*, <https://measuringu.com/sus/>. Diakses pada 29 Oktober 2018.
- Selviana, Nur Indah dan Mustakim. 2016. *Analisis Perbandingan K-Means dan Fuzzy C-Means untuk Pemetaan Motivasi Belajar Mahasiswa*. SNTIKI, (Online), Volume 8.
- Soleh, Achmad Zanbar. 2005. *Ilmu Statistik. Cetakan Pertama*. Bandung: Rekayasa Sains.
- Suganya, R. dan Shanthi, R. 2012. *Fuzzy C-Means Algorithm- A review*. International Journal of Scientific and Research Publications, Volume 2, Issue 11.

- Suwitho, Maldi. 2014. *Pengaruh Customer Relationship Management Terhadap Loyalitas Pelanggan PT Moga Djaja Di Surabaya*. Jurnal Ilmu dan Riset Manajemen, Volume 3, Nomor 8.
- Wilson, D. Randall dan Martinez, R. Tony. 1997. *Improved Heterogeneous Distance Functions*. Journal of Artificial Intelligence Research, (Online), Volume 6.
- Wu, Hsin-Hung dkk. 2014. *Analyzing Patients' Values by Applying Cluster Analysis and LRFM Model in a Pediatric Dental Clinic in Taiwan*. The Scientific World Journal, Volume 2014.
- Yuliar, Ni Putu Putri dkk. 2015. *Customer Segmentation Through Fuzzy C-Means and Fuzzy RFM Method*. Journal of Theoretical and Applied Information Technology, (Online), Volume 78, Nomor 3, (www.iatit.org).

