

**ANALISIS SEGMENTASI KONSUMEN MENGGUNAKAN METODE  
K-MEANS *CLUSTERING* BERDASARKAN MODEL RFM DAN  
PENGETAHUAN DATA LAINNYA SEBAGAI REKOMENDASI  
STRATEGI PEMASARAN  
(STUDI KASUS : YAYASAN XYZ)**

**(Skripsi)**

**Oleh**

**RIKA OKTA NABELLA**

**1815061004**



**FAKULTAS TEKNIK  
UNIVERSITAS LAMPUNG  
BANDAR LAMPUNG  
2022**

## ABSTRAK

### ANALISIS SEGMENTASI KONSUMEN MENGGUNAKAN METODE K-MEANS *CLUSTERING* BERDASARKAN MODEL RFM DAN PENGETAHUAN DATA LAINNYA SEBAGAI REKOMENDASI STRATEGI PEMASARAN (STUDI KASUS : YAYASAN XYZ)

Oleh

Rika Okta Nabella

Perusahaan yang berfokus pada penjualan sudah mulai menerapkan teknik data mining di mana tidak hanya mengembangkan suatu produk dan layanan namun juga telah mendalami kebutuhan dari konsumen yang didapatkan dari data transaksi maupun data konsumen. Berdasarkan alur pikir tersebut, peneliti akan melakukan analisis *clustering* untuk mengetahui karakteristik berdasarkan pola-pola yang terbentuk menggunakan *tools RapidMiner* dengan penerapan metode K-Means berdasarkan model RFM (*Recency, Frequency, Monetary*) dan pengetahuan data lain yang ada pada perusahaan. Penerapan visualisasi data menggunakan Google Data Studio dapat membantu dalam melakukan visualisasi dengan mudah dan cepat. Untuk metode penelitian yang digunakan dalam analisis yaitu *Cross Industry Standart Process for Data Mining* atau CRISP-DM. Metode pengumpulan data yang didapatkan berasal dari database perusahaan. Penelitian menghasilkan bahwa 3 *cluster* memiliki nilai 0.197 menggunakan uji validitas *Davies Bouldin Index*. *Cluster low priority* berjumlah 1090 konsumen, *cluster medium priority* berjumlah 73 konsumen dan *cluster high priority* berjumlah 969 konsumen. Sehingga, berdasarkan karakteristik berbeda yang dimiliki setiap *cluster* dapat menghasilkan strategi pemasaran pada setiap *cluster*. Selain itu, pengetahuan data lain yang ada pada perusahaan juga dapat membantu dalam menghasilkan strategi pemasaran dan keputusan bisnis.

Kata kunci : *Clustering, K-Means, Model RFM, CRISP-DM, Davies Bouldin Index, Google Data Studio, RapidMiner*

## **ABSTRACT**

### **CONSUMER SEGMENTATION ANALYSIS USING K-MEANS CLUSTERING METHOD BASED ON RFM MODEL AND OTHER DATA KNOWLEDGE AS MARKETING STRATEGY RECOMMENDATIONS (CASE STUDY: YAYASAN XYZ)**

**By**

**Rika Okta Nabella**

*Companies that focus on sales have started to apply data mining techniques which not only develop a product and service but also have explored the needs of consumers obtained from transaction data and consumer data. Based on that, the researcher will conduct a clustering analysis to find out the characteristics based on the patterns formed using RapidMiner tools with the application of the K-Means method based on the RFM (Recency, Frequency, Monetary) model and knowledge of other data in the company. Implementing data visualization using Google Data Studio can help visualize easily and quickly. The research method used in the analysis is Cross Industry Standard Process for Data Mining or CRISP-DM. The data collection method obtained is derived from the company's database. The research shows that 3 clusters have 0.197 using the Davies Bouldin Index validity test. There are 1090 low priority clusters, 73 medium priority clusters and 969 high priority clusters. So, based on the different characteristics that each cluster has, it can produce a marketing strategy in each cluster. In addition, knowledge of other data that exists in the company can also help in generating marketing strategies and business decisions.*

**Keywords : Clustering, K-Means, RFM Model, CRISP-DM, Davies Bouldin Index, Google Data Studio, RapidMiner**

**ANALISIS SEGMENTASI KONSUMEN MENGGUNAKAN METODE K-  
MEANS *CLUSTERING* BERDASARKAN MODEL RFM DAN  
PENGETAHUAN DATA LAINNYA SEBAGAI REKOMENDASI  
STRATEGI PEMASARAN  
(STUDI KASUS : YAYASAN XYZ)**

**Oleh**

**RIKA OKTA NABELLA**

**Skripsi**

Sebagai Salah Satu Syarat untuk Mencapai Gelar

**SARJANA TEKNIK**

**Pada**

Jurusan Teknik Elektro

Fakultas Teknik Universitas Lampung



**FAKULTAS TEKNIK  
UNIVERSITAS LAMPUNG  
BANDAR LAMPUNG  
2022**

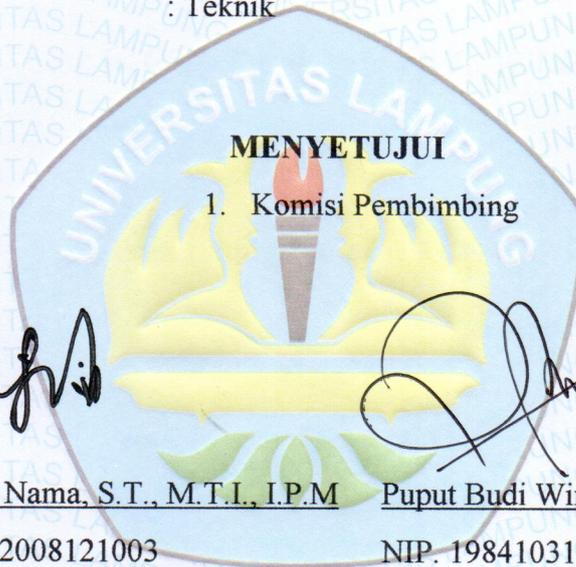
Judul Skripsi : **ANALISIS SEGMENTASI KONSUMEN  
MENGUNAKAN METODE K-MEANS  
CLUSTERING BERDASARKAN MODEL RFM  
DAN PENGETAHUAN DATA LAINNYA  
SEBAGAI REKOMENDASI STRATEGI  
PEMASARAN  
(STUDI KASUS : YAYASAN XYZ)**

Nama Mahasiswa : Rika Okta Nabella

Nomor Pokok Mahasiswa : 1815061004

Jurusan : Teknik Elektro

Fakultas : Teknik



Ir. Gigh Forda Nama, S.T., M.T.I., I.P.M

Puput Budi Wintoro S.Kom, M.T.

NIP. 198307122008121003

NIP. 198410312019031004

2. Mengetahui

Ketua Jurusan  
Teknik Elektro

Ketua Program Studi  
Teknik Informatika

Herlinawati S.T., M.T.

NIP. 197103141999032001

Mona Arif Muda, S.T., M.T.

NIP. 197111122000031002

**MENGESAHKAN**

1. Tim Penguji

Ketua

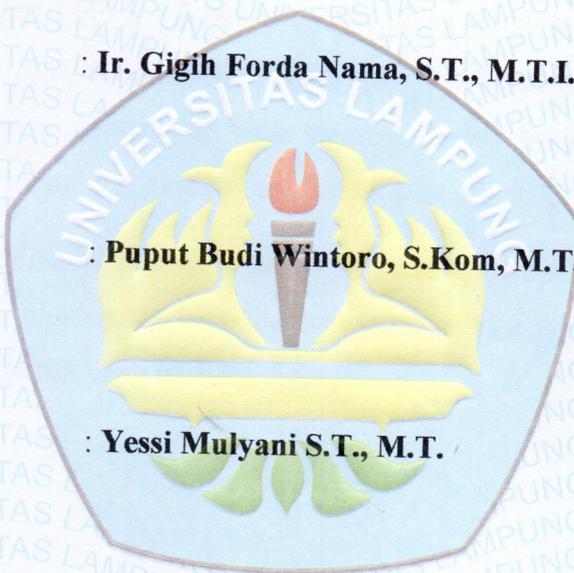
: **Ir. Gigih Forda Nama, S.T., M.T.I., I.P.M.**.....

Sekretaris

: **Puput Budi Wintoro, S.Kom, M.T.**.....

Penguji

: **Yessi Mulyani S.T., M.T.**.....



*[Handwritten signatures of the three members of the examination team]*

2. Dekan Fakultas Teknik



**Dr. Eng. Ir. Helmy Fitriawan, S.T., M.Sc.** *[Signature]*

NIP. 19750928 200112 1002

Tanggal Lulus Ujian Skripsi : 30 Mei 2022

## SURAT PERNYATAAN

Saya yang bertandatangan dibawah ini, menyatakan bahwa skripsi saya yang berjudul "ANALISIS SEGMENTASI KONSUMEN MENGGUNAKAN METODE K-MEANS CLUSTERING BERDASARKAN MODEL RFM DAN PENGETAHUAN DATA LAINNYA SEBAGAI REKOMENDASI STRATEGI PEMASARAN (STUDI KASUS : YAYASAN XYZ)" dengan ini menyatakan bahwa skripsi saya dibuat oleh saya sendiri. Semua hasil yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung.

Apabila di kemudian hari terbukti bahwa skripsi ini merupakan salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan hukum atau akademik yang berlaku.

Bandar Lampung, 5 Juni 2022

Pembuat pernyataan,



Rika Okta Nabella

NPM. 1815061004

## RIWAYAT HIDUP



Penulis bernama Rika Okta Nabella yang lahir di Bandar Lampung, 28 Oktober 1999. Penulis merupakan anak kedua dari Bapak Ruzlan dan Ibu Eriana. Penulis memiliki satu kakak perempuan bernama Ria Julianti Ariska dan satu adik perempuan bernama Keisha Rajni Ramadhini.

Penulis telah menyelesaikan pendidikannya di SD Negeri 1 Beringin Raya pada tahun 2012, SMP Negeri 14 Bandar Lampung pada tahun 2015, dan SMA Negeri 9 Bandar Lampung pada tahun 2018. Penulis terdaftar sebagai Mahasiswa Fakultas Teknik, Jurusan Teknik Elektro Program Studi Teknik Informatika Universitas Lampung pada tahun 2018 melalui jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) dan mengikuti kegiatan Kuliah Kerja Nyata (KKN) selama 40 hari di Desa Gunung Mas, Kecamatan Teluk Betung Selatan, Bandar Lampung.

Selama menempuh pendidikan di Jurusan Teknik Elektro Program Studi Teknik Informatika Universitas Lampung, penulis aktif mengikuti beberapa kegiatan, antara lain :

1. Anggota Himpunan Mahasiswa Teknik Elektro, Departemen Sosial dan Kewirausahaan, Divisi Kewirausahaan Periode 2018-2019.
2. Anggota Himpunan Mahasiswa Teknik Elektro, Departemen Sosial dan Kewirausahaan, Divisi Sosial Periode 2019-2020

3. Anggota Generasi Baru Bank Indonesia, Divisi Kerjasama dan Kemitraan Periode 2020-2021.
4. Melaksanakan Kerja Praktik di PT Telkom Indonesia yang berada di Jalan Kartini No.01, Kelapa Tiga, Tanjung Karang Pusat, Klp. Tiga, Kec. Tj Karang Pusat, Kota Bandar Lampung, Lampung pada tanggal 1 Februari 2021 hingga 12 Maret 2021.
5. Melaksanakan Kuliah Kerja Nyata (KKN) selama 40 hari di Desa Gunung Mas, Kecamatan Teluk Betung Selatan, Bandar Lampung pada bulan Agustus – September 2021.
6. Mengikuti kegiatan pelatihan *Big Data using Python* pada program Digital Talent Scholarship pada Juli – September 2021.
7. Melaksanakan Magang Bersertifikat pada program Merdeka Belajar - Kampus Merdeka (MBKM) di Yayasan Adipurna Inovasi Asia pada bulan Agustus 2021 – Januari 2022.
8. Melaksanakan kegiatan pelatihan *Path Data Analyst* pada program Generasi GIGIH 2.0 di Yayasan Anak Bangsa Bisa pada tahun 2022.

## PERSEMBAHAN

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Segala puji syukur kepada Allah SWT berkat karunia, kesehatan, rahmat serta hidayah-Nya yang telah diberikan, shalawat teriring salam kepada Nabi Muhammad SAW, suri tauladan Akhlaqul Kharimah yang kita nantikan syafa'atnya di hari akhir kelak. Dengan segala kerendahan hati, saya persembahkan skripsi ini kepada:

Ayah dan Ibu yang saya sayangi Ruzlan dan Eriana. Dan juga kakak dan adik saya Ria Julianti Ariska dan Keisha Rajni Ramadhini.

Terimakasih telah selalu memberikan doa, dukungan, dan motivasi untuk selalu melakukan yang terbaik di dalam hidup.

Almamater tercinta Universitas Lampung

## **MOTO**

“Allah tidak akan memberikan suatu cobaan di luar batas kemampuan manusia”

**-(Q.S. Al Baqarah : 286 )-**

“Barang siapa yang bersungguh-sungguh, sesungguhnya kesungguhan tersebut  
untuk kebaikan dirinya sendiri”

**-(Q.S. Al-Ankabut: 6 )-**

## SANWACANA

Dengan mengucapkan Alhamdulillahirabbil'alamin, segala puji bagi Allah SWT yang Maha Pengasih lagi Maha Penyayang, atas limpahan rahmat, taufik serta hidayah-Nya sehingga penulis dapat menyelesaikan penulisan skripsi yang berjudul: "ANALISIS SEGMENTASI KONSUMEN MENGGUNAKAN METODE K-MEANS BERDASARKAN MODEL RFM DAN PENGETAHUAN DATA LAINNYA SEBAGAI REKOMENDASI STRATEGI PEMASARAN (STUDI KASUS : YAYASAN XYZ)" Sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik pada Fakultas Teknik Universitas Lampung. Dalam penyelesaian skripsi ini, penulis mendapatkan banyak bantuan, bimbingan, saran dan dukungan dari segenap pihak, baik secara langsung maupun tidak langsung sehingga penyusunan skripsi ini berjalan dengan baik. Maka pada kesempatan ini penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Drs. Suharno, M.Sc., Ph.D., selaku Dekan Fakultas Teknik Universitas Lampung.
2. Bapak Khairudin, Ph.D., Eng.. selaku Ketua Jurusan Teknik Elektro Fakultas Teknik Universitas Lampung.
3. Bapak Mona Arif Muda, S.T., M.T., selaku Ketua Program Studi Teknik Informatika Universitas Lampung

4. Bapak Ir Gigih Forda Nama, S.T., M.T.I., I.P.M selaku Pembimbing Utama atas bimbingan, saran, motivasi dan kesabaran kepada penulis selama penyelesaian skripsi.
5. Bapak Puput Budi Wintoro, S.Kom, M.T. selaku Pembimbing Kedua atas bimbingan, saran, motivasi dan kesabaran kepada penulis selama penyelesaian skripsi.
6. Ibu Yessi Mulyani, S.T., M.T., selaku Penguji Utama dan Pembimbing Akademik atas bimbingan serta masukan dalam penyelesaian skripsi maupun bimbingan selama perkuliahan.
7. Seluruh Dosen Teknik Informatika atas bimbingan dan kesabarannya dalam mendidik penulis.
8. Mbak Rika yang telah membantu penulis dalam segala urusan administrasi.
9. Ramadhan Irawan yang selalu memberikan semangat dan selalu bersedia mendengarkan cerita sehari-hari.
10. Destiara Rizky Rahmadanti dan Almira Zavira Alami yang selalu membantu dan memberikan semangat penulis selama perkuliahan.
11. Teman-teman Teknik Informatika 2018 dan Teknik Elektro 2018 atas segala bentuk dukungan yang selalu kalian berikan selama ini.
12. Diri sendiri karena telah berusaha menyelesaikan perkuliahan dan skripsi ini tepat waktu.

Semoga Allah SWT membalas segala kebaikan hati dan jasa yang telah kalian berikan kepada saya. Saya menyadari meskipun skripsi ini sudah disusun dengan sebaik mungkin, skripsi ini masih terdapat kekurangan dan masih jauh dari kesempurnaan, namun saya sangat berharap skripsi ini akan membawa manfaat bagi siapapun yang membacanya dan bagi penulis dalam mengembangkan dan mengamalkan ilmu pengetahuan.

Bandar Lampung 5 Juni 2022

Penulis

A handwritten signature in black ink, appearing to read 'Rika Okta Nabella', with a stylized flourish at the end.

Rika Okta Nabella

NPM. 1815061004

## DAFTAR ISI

Halaman

|  |            |
|--|------------|
| <b>DAFTAR ISI</b> .....                      | <b>i</b>   |
| <b>DAFTAR GAMBAR</b> .....                   | <b>iv</b>  |
| <b>DAFTAR TABEL</b> .....                    | <b>vi</b>  |
| <b>DAFTAR SYNTAX</b> .....                   | <b>vii</b> |
| <b>I. PENDAHULUAN</b> .....                  | <b>1</b>   |
| 1.1 Latar Belakang .....                     | 1          |
| 1.2 Rumusan Masalah .....                    | 3          |
| 1.3 Tujuan Penelitian.....                   | 4          |
| 1.4 Manfaat Penelitian.....                  | 4          |
| 1.5 Batasan Masalah.....                     | 4          |
| 1.6 Hipotesis Penelitian .....               | 5          |
| 1.7 Sistematika Penulisan.....               | 5          |
| <b>II. TINJAUAN PUSTAKA</b> .....            | <b>7</b>   |
| 2.1 Profil Perusahaan.....                   | 7          |
| 2.1.1 Visi dan Misi Perusahaan.....          | 8          |
| 2.1.2 Nilai-nilai Perusahaan .....           | 8          |
| 2.1.3 Struktur Organisasi Perusahaan .....   | 9          |
| 2.2 Segmentasi Konsumen .....                | 10         |
| 2.3 Strategi Pemasaran .....                 | 11         |
| 2.4 <i>Data Mining</i> .....                 | 12         |
| 2.4.1 Pengertian <i>Data Mining</i> .....    | 12         |
| 2.4.2 Pengelompokan <i>Data Mining</i> ..... | 13         |
| 2.4.3 Tahapan <i>Data Mining</i> .....       | 14         |
| 2.5 <i>Clustering</i> .....                  | 15         |

|             |  |           |
|-------------|--|-----------|
| 2.6         | <i>K-Means</i> .....                                 | 16        |
| 2.6.1       | Keunggulan dan Kelemahan Metode <i>K-Means</i> ..... | 19        |
| 2.7         | Model RFM .....                                      | 20        |
| 2.8         | <i>Detect Outlier (Distance)</i> .....               | 21        |
| 2.9         | <i>Min-Max Normalization</i> .....                   | 22        |
| 2.10        | Penentuan Jumlah <i>Cluster</i> .....                | 23        |
| 2.11        | RapidMiner.....                                      | 25        |
| 2.12        | JupyterLab .....                                     | 26        |
| 2.13        | Google Data Studio .....                             | 27        |
| 2.14        | <i>Google Sheet</i> .....                            | 28        |
| 2.15        | CRISP-DM .....                                       | 29        |
| 2.16        | Kategori Usia.....                                   | 32        |
| 2.17        | Penelitian Terkait .....                             | 33        |
| <b>III.</b> | <b>METODE PENELITIAN.....</b>                        | <b>39</b> |
| 3.1         | Waktu dan Tempat .....                               | 39        |
| 3.2         | Alat dan Bahan .....                                 | 40        |
| 3.2.1       | Alat.....  | 40        |
| 3.2.2       | Bahan.....   | 41        |
| 3.3         | Tahap Penelitian .....                               | 42        |
| 3.3.1       | Studi Literatur .....                                | 42        |
| 3.3.2       | Penggunaan Metode CRISP-DM .....                     | 43        |
| <b>IV.</b>  | <b>HASIL DAN PEMBAHASAN.....</b>                     | <b>54</b> |
| 4.1         | Pemahaman Bisnis .....                               | 54        |
| 4.1.1       | Menentukan Tujuan Bisnis .....                       | 54        |
| 4.1.2       | Menilai Situasi .....                                | 55        |
| 4.1.3       | Menentukan Tujuan Data Mining .....                  | 56        |
| 4.1.4       | Menghasilkan Rencana Project .....                   | 56        |
| 4.2         | Pemahaman Data .....                                 | 57        |
| 4.2.1       | Mengumpulkan Data.....                               | 57        |
| 4.2.2       | Mendeskripsikan Data.....                            | 57        |
| 4.2.3       | Mengeksplor Data .....                               | 60        |
| 4.2.4       | Memverifikasi Kualitas Data .....                    | 60        |
| 4.3         | Persiapan Data .....                                 | 62        |

|           |  |            |
|-----------|--|------------|
| 4.3.1     | Persiapan Data Segmentasi Konsumen.....                        | 62         |
| 4.3.2     | Persiapan Data Potensial Konsumen.....                         | 72         |
| 4.3.3     | Persiapan Data Kategori, Kursus, dan Tipe Pembayaran .....     | 78         |
| 4.3.4     | Persiapan Data Distribusi Usia dan <i>Gender</i> .....         | 83         |
| 4.3.5     | Persiapan Data Top Rating Kursus .....                         | 87         |
| 4.3.6     | Persiapan Data Top Pendapatan <i>Affiliate Marketing</i> ..... | 91         |
| 4.3.7     | Persiapan Data Trend Akses Website .....                       | 94         |
| 4.4       | Pemodelan .....  | 96         |
| 4.4.1     | Pemodelan K-Means Menggunakan Perhitungan Excel.....           | 96         |
| 4.4.2     | Pemodelan K-Means Menggunakan RapidMiner .....                 | 100        |
| 4.5       | Evaluasi .....   | 103        |
| 4.6       | Penyebaran ( <i>Deployment</i> ) .....                         | 107        |
| 4.6.1     | Analisis Data Segmentasi Konsumen .....                        | 110        |
| 4.6.2     | Analisis Data Potensial Konsumen .....                         | 120        |
| 4.6.3     | Analisis Data Kategori, Kursus dan Tipe Pembayaran.....        | 122        |
| 4.6.4     | Analisis Data Distribusi Usia dan <i>Gender</i> .....          | 124        |
| 4.6.5     | Analisis Data Top Rating Kursus.....                           | 126        |
| 4.6.6     | Analisis Data Top Pendapatan <i>Affiliate Marketing</i> .....  | 128        |
| 4.6.7     | Analisis Data Trend Akses Website.....                         | 130        |
| <b>5.</b> | <b>KESIMPULAN DAN SARAN.....</b>                               | <b>132</b> |
| 5.1       | Kesimpulan.....  | 132        |
| 5.2       | Saran.....   | 133        |
|           | <b>DAFTAR PUSTAKA .....</b>                                    | <b>134</b> |

## DAFTAR GAMBAR

|   | Halaman |
|---|---------|
| Gambar 2.1 Struktur Organisasi Perusahaan .....                   | 9       |
| Gambar 2.2 Tahapan Data Mining.....                               | 14      |
| Gambar 2.3 Contoh Cluster K-Means.....                            | 19      |
| Gambar 2.4 Tampilan RapidMiner .....                              | 26      |
| Gambar 2.5 Tampilan JupyterLab .....                              | 27      |
| Gambar 2.6 Tampilan Google Data Studio.....                       | 28      |
| Gambar 2.7 Tampilan Google Sheet.....                             | 29      |
| Gambar 2.8 Alur Model CRISP-DM [35] .....                         | 30      |
| Gambar 2.9 Dashboard PT Arthamas Citra Mandiri .....              | 34      |
| Gambar 2.10 Dashboard Medical Center ITS.....                     | 36      |
| Gambar 3.1 Hirarchical dan Non Hirarchical Clustering.....        | 16      |
| Gambar 3.2 Flowchart Algoritma K-Means .....                      | 18      |
| Gambar 3.3 Tahap Penelitian Metode CRISP-DM.....                  | 43      |
| Gambar 3.4 Record Tabel Users .....                               | 47      |
| Gambar 3.5 Record Tabel Users_detail .....                        | 47      |
| Gambar 3.6 Record Tabel Course.....                               | 48      |
| Gambar 3.7 Record Tabel Category .....                            | 48      |
| Gambar 3.8 Record Tabel Payment .....                             | 49      |
| Gambar 3.9 Record Tabel Payment_balance .....                     | 49      |
| Gambar 3. 10 Record Tabel Rating .....                            | 49      |
| Gambar 3. 11 Record Tabel Ci_sessions .....                       | 50      |
| Gambar 4.1 Class Diagram Database Perusahaan .....                | 59      |
| Gambar 4.2 Hasil Integrasi Data Segmentasi Konsumen .....         | 65      |
| Gambar 4.3 Data Segmentasi Konsumen .....                         | 66      |
| Gambar 4.4 Konfigurasi RapidMiner pada Tahap Detect Outlier ..... | 67      |
| Gambar 4.5 Hasil Konfigurasi Detect Outlier .....                 | 68      |

|  |     |
|--|-----|
| Gambar 4.6 Konfigurasi RapidMiner pada Tahap Normalisasi.....              | 69  |
| Gambar 4.7 Hasil Konfigurasi Normalisasi.....                              | 69  |
| Gambar 4. 8 Hasil Integrasi Data Potensial Konsumen.....                   | 75  |
| Gambar 4.9 Data Potensial Konsumen.....                                    | 77  |
| Gambar 4.10 Hasil Integrasi Data Kategori, Kursus dan Tipe Pembayaran..... | 80  |
| Gambar 4.11 Data Kategori, Kursus dan Tipe Pembayaran.....                 | 82  |
| Gambar 4. 12 Hasil Integrasi Data Distribusi Usia dan Gender.....          | 85  |
| Gambar 4.13 Data Distribusi Usia dan Gender.....                           | 86  |
| Gambar 4.14 Hasil Integrasi Data Top Rating Kursus.....                    | 89  |
| Gambar 4.15 Data Top Rating Kursus.....                                    | 90  |
| Gambar 4.16 Hasil Integrasi Data Top Pendapatan Affiliate Marketing.....   | 92  |
| Gambar 4.17 Data Top Pendapatan Affiliate Marketing.....                   | 93  |
| Gambar 4.18 Data Trend Akses Website.....                                  | 95  |
| Gambar 4.19 Hasil Perhitungan K-Means Menggunakan Microsoft Excel.....     | 98  |
| Gambar 4. 20 Konfigurasi RapidMiner.....                                   | 100 |
| Gambar 4. 21 Konfigurasi Loop dengan $K = 2 - 10$ .....                    | 103 |
| Gambar 4. 22 Menambahkan Sumber Data pada Google Data Studio.....          | 108 |
| Gambar 4.23 Memilih Sheet.....   | 108 |
| Gambar 4.24 Menyajikan Grafik pada Google Data Studio.....                 | 109 |
| Gambar 4.25 Plot Cluster Berdasarkan Nilai Centroid.....                   | 110 |
| Gambar 4.26 Visualisasi Segmentasi Konsumen.....                           | 116 |
| Gambar 4.27 Control List Cluster.....                                      | 117 |
| Gambar 4.28 Visualisasi Potensial Users.....                               | 120 |
| Gambar 4.29 Visualisasi Kategori, Kursus dan Tipe Pembayaran.....          | 122 |
| Gambar 4.30 Visualisasi Distribusi Usia dan Gender.....                    | 124 |
| Gambar 4.31 Visualisasi Top Rating Kursus.....                             | 126 |
| Gambar 4.32 Visualisasi Top Pendapatan Affiliate marketing.....            | 128 |
| Gambar 4.33 Visualisasi Data Trend Akses Website.....                      | 130 |

## DAFTAR TABEL

|  | Halaman |
|--|---------|
| Tabel 3.1 Jadwal Penelitian .....  | 39      |
| Tabel 3.2 Analisis Data dan Tujuan.....  | 44      |
| Tabel 3.3 Tabel-tabel Database.....  | 46      |
| Tabel 3.4 Tabel dan Field yang digunakan .....   | 51      |
| Tabel 4.1 Tabel dan Field pada Data Segmentasi Konsumen.....   | 62      |
| Tabel 4.2 Data Segmentasi Konsumen .....   | 70      |
| Tabel 4.3 Perbandingan Normalisasi Perhitungan Excel dan RapidMiner .....                                | 71      |
| Tabel 4.4 Tabel dan Field pada Data Potensial Konsumen .....   | 72      |
| Tabel 4.5 Tabel dan Field pada Data Kategori, Kursus dan Tipe Pembayaran .....                           | 78      |
| Tabel 4.6 Tabel dan Field pada Data Distribusi Usia dan Gender.....                                      | 83      |
| Tabel 4.7 Tabel dan Field Data Top Rating Kursus.....  | 87      |
| Tabel 4.8 Tabel dan Field pada Data Top Pendapatan Affiliate Marketing .....                             | 91      |
| Tabel 4.9 Tabel dan Field pada Data Trend Akses Website .....  | 94      |
| Tabel 4.10 Inisialisasi Awal Cluster .....   | 97      |
| Tabel 4.11 Contoh Data Perhitungan.....  | 97      |
| Tabel 4.12 Hasil Centroid Iterasi ke-1 .....   | 98      |
| Tabel 4.13 Hasil Centroid Iterasi ke-2.....  | 99      |
| Tabel 4.14 Hasil Centroid Iterasi ke-3 .....   | 99      |
| Tabel 4.15 Hasil Centroid Iterasi ke -4.....   | 99      |
| Tabel 4.16 Hasil Cluster pada Tahapan Membangun Model.....   | 101     |
| Tabel 4.17 Perbandingan Centroid Cluster Menggunakan Perhitungan Microsoft Excel<br>dan RapidMiner ..... | 101     |
| Tabel 4.18 Hasil Cluster pada Tahapan Evaluasi .....   | 104     |
| Tabel 4.19 Nilai Centroid RFM Cluster.....   | 111     |
| Tabel 4.20 Karakteristik Konsumen pada Cluster Low priority .....  | 112     |
| Tabel 4.21 Karakteristik Konsumen pada Cluster Medium Priority.....                                      | 113     |
| Tabel 4.22 Karakteristik Konsumen pada Cluster High Priority .....                                       | 114     |
| Tabel 4.23 Perbandingan Karakteristik 3 Cluster .....  | 115     |

## DAFTAR SYNTAX

|  | Halaman |
|--|---------|
| Syntax 4.1 Koneksi ke Local Database.....  | 63      |
| Syntax 4. 2 Memilih Local Database .....   | 63      |
| Syntax 4.3 Mengintegrasikan Data Segmentasi Konsumen .....                                       | 64      |
| Syntax 4.4 Pembersihan Data Segmentasi Konsumen.....   | 66      |
| Syntax 4.5 Mengintegrasikan Data Potensial Konsumen .....  | 74      |
| Syntax 4.6 Membersihkan Data Potensial Konsumen Hanya Login dan Hanya Checkout<br>.....          | 75      |
| Syntax 4.7 Membersihkan Data Potensial Konsumen Pembayaran Expired dan<br>Pembayaran Batal ..... | 76      |
| Syntax 4.8 Mengintegrasikan Data Kategori, Kursus dan Tipe Pembayaran .....                      | 80      |
| Syntax 4.9 Membersihkan Data Kategori, Kursus dan Tipe Pembayaran .....                          | 81      |
| Syntax 4.10 Mengintegrasikan Data Distribusi Usia dan Gender.....                                | 84      |
| Syntax 4.11 Membersihkan Data Distibusi Usia dan Gender.....                                     | 85      |
| Syntax 4.12 Mengintegrasikan Data Top Rating Kursus.....   | 88      |
| Syntax 4.13 Membersihkan Data Top Rating Kursus.....   | 89      |
| Syntax 4.14 Mengintegrasikan Data Top Pendapatan Affiliate Marketing .....                       | 92      |
| Syntax 4.15 Membersihkan Data Top Pendapatan Affiliate Marketing .....                           | 93      |
| Syntax 4.16 Mengintegrasikan Data Trend Akses Website.....                                       | 94      |

## I. PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan teknologi informasi saat ini membawa pengaruh terhadap meningkatnya perkembangan di sektor bisnis. Teknologi informasi memiliki peran penting dalam membantu kegiatan bisnis sehingga lebih efektif dan efisien, diantaranya memberikan kemudahan dalam meningkatkan inovasi, melakukan komunikasi, proses penyebaran informasi, dan tentunya menghemat biaya.

Selain itu, data transaksi konsumen menjadi hal yang penting yang dapat dimanfaatkan dalam menemukan pola-pola konsumen. Dengan begitu, sifat dan perilaku konsumennya dapat diketahui oleh perusahaan sehingga perusahaan dapat melakukan bentuk strategi pemasaran berdasarkan pola konsumen tersebut. Teknologi yang digunakan yaitu teknologi *Data Mining*. *Data Mining* akan memungkinkan pengguna menemukan pengetahuan berdasarkan hasil ekstraksi dari database yang mungkin sebelumnya belum diketahui.

Di Indonesia, perusahaan-perusahaan yang berfokus pada penjualan sudah mulai menerapkan data mining dimana tidak hanya mengembangkan suatu produk dan layanan namun juga telah mendalami kebutuhan dari konsumen yang didapatkan dari data konsumen maupun data transaksi yang sudah dilakukan sebelumnya.

Yayasan XYZ bergerak dalam bidang pendidikan dengan kegiatan penjualan jasa kursus *online* dengan target pasarnya pelajar, mahasiswa dan karyawan yang ingin meningkatkan keterampilan dan keahlian yang terintegrasi dengan dunia kerja dan bisnis. Selain itu, menyediakan platform *marketplace* dan *affiliate marketing* di bidang edukasi.

Yayasan XYZ memiliki beberapa kompetitor sehingga sebagai langkah dalam mempertahankan konsumen agar tetap memiliki hubungan yang baik dengan perusahaan, seperti users yang belum pernah melakukan transaksi berbayar berubah menjadi pernah melakukan transaksi berbayar selain itu konsumen yang pernah melakukan transaksi berbayar kembali melakukan transaksi berbayar diperlukan pengelolaan data transaksi konsumen dengan teknologi *data mining* yang tepat agar dapat unggul dibandingkan kompetitor lain karena selama ini ribuan data *history* transaksi belum dimanfaatkan secara maksimal oleh perusahaan untuk diketahui pola-pola yang dimiliki dan kegiatan marketing perusahaan masih menerapkan perlakuan yang sama kepada seluruh konsumennya.

Pemanfaatan teknologi *data mining* dalam melakukan segmentasi konsumen berdasarkan model *Recency, Frequency, Monetary* (RFM) yaitu variabel terakhir kali melakukan transaksi dalam hari, jumlah transaksi, dan total nilai pembelian dalam kurun waktu tertentu adalah teknik *data mining clustering* salah satunya dengan metode K-Means untuk mengetahui pola konsumen di setiap *cluster* sehingga kegiatan marketing yang dilakukan perusahaan menjadi terarah.

Selain data transaksi, data-data lain yang mendukung penelitian seperti data *gender* dan usia, data rating kursus, data pendapatan *affiliate marketing*, data trend akses website dan data potensial konsumen belum dimanfaatkan oleh perusahaan

untuk dijadikan pengetahuan dalam membantu membuat strategi pemasaran. Sehingga, data-data lain tersebut perlu dilakukan analisis untuk diketahui pengetahuan datanya agar dapat digunakan sebagai bentuk upaya dalam meningkatkan *revenue* perusahaan dengan cara memberikan rekomendasi strategi pemasaran

Dari uraian di atas, peneliti tertarik dalam melakukan penelitian yang akan digunakan sebagai tugas akhir dengan judul **“ANALISIS SEGMENTASI KONSUMEN MENGGUNAKAN METODE K-MEANS *CLUSTERING* BERDASARKAN MODEL RFM DAN PENGETAHUAN DATA LAINNYA SEBAGAI REKOMENDASI STRATEGI PEMASARAN”**.

## **1.2 Rumusan Masalah**

Berdasarkan penjelasan dasar masalah tersebut, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana mendapatkan pengetahuan pola-pola yang terbentuk berdasarkan data transaksi konsumen menggunakan *tools* dengan penerapan metode K-Means berdasarkan model RFM dan pengetahuan data lainnya ?
2. Bagaimana penerapan google data studio dapat mempermudah visualisasi data ?
3. Bagaimana bentuk rekomendasi strategi pemasaran yang dapat dilakukan berdasarkan analisis yang dilakukan ?

### **1.3 Tujuan Penelitian**

Tujuan dari penelitian ini adalah:

1. Dapat mengetahui karakteristik konsumen berdasarkan pola-pola yang terbentuk menggunakan *tools* dengan penerapan metode K-Means berdasarkan model RFM dan pola-pola pengetahuan data lainnya.
2. Dapat membuat visualisasi dengan penerapan google data studio.
3. Dapat merekomendasikan bentuk strategi pemasaran yang akan dilakukan.

### **1.4 Manfaat Penelitian**

Penelitian ini dapat memberikan manfaat sebagai berikut:

1. Dapat dipahami proses *clustering* menggunakan metode K-Means berdasarkan model RFM dan proses mencari pola pengetahuan data lainnya.
2. Dapat dipahami proses visualisasi dengan penerapan google data studio.
3. Dapat diberikan rekomendasi strategi pemasaran berdasarkan analisis data yang telah dibuat.

### **1.5 Batasan Masalah**

Penelitian ini memiliki batasan masalah sebagai berikut:

1. Tidak menjelaskan cara kerja perangkat lunak RapidMiner dan cara pengoperasian.

2. Metode *data mining* yang digunakan hanya satu yaitu *clustering* menggunakan metode K-Means berdasarkan model RFM.
3. Hanya membahas data transaksi penjualan produk dan jasa perusahaan untuk periode 1 Januari 2021 sampai dengan 25 November 2021.
4. Visualisasi menggunakan penerapan google data studio.

### **1.6 Hipotesis Penelitian**

Berdasarkan rumusan masalah maupun literatur, peneliti mempunyai hipotesis bahwa teknik *data mining* menggunakan metode K-Means berdasarkan model RFM (*Recency*, *Frequency* dan *Monetary*) dapat digunakan untuk melakukan *cluster* dalam analisis tingkat prioritas konsumen di Yayasan XYZ guna di dapatkannya data valid terkait *cluster high priority*, *medium priority*, dan *low priority*. Penerapan google data studio dapat membantu dalam melakukan visualisasi. Selain itu, dengan data konsumen lainnya yaitu data potensial konsumen, data *gender* dan usia, data rating kursus, data *affiliate marketing* dan data trend akses website kedepannya membantu perusahaan menentukan strategi pemasaran yang tepat dalam upaya meningkatkan pendapatan perusahaan.

### **1.7 Sistematika Penulisan**

Adapun sistematika penulisan laporan penelitian ini adalah sebagai berikut

1. PENDAHULUAN

Pada bagian ini membahas terkait latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, hipotesis dan sistematika penulisan.

## 2. TINJAUAN PUSTAKA

Pada bagian ini membahas teori-teori penunjang penelitian seperti Profil Perusahaan, Strategi Pemasaran, Segmentasi Konsumen, Data Mining, *Clustering*, K-Means, Model RFM, *Detect Outlier* (Distance), *Min-Max Normalization*, Penentuan Jumlah *Cluster*, RapidMiner, JupyterLab, Google Data Studio, Google Sheet, CRISP-DM, Kategori Usia dan penelitian terkait.

## 3. METODOLOGI PENELITIAN

Pada bagian ini membahas waktu dan tempat penelitian, alat dan bahan, metode yang digunakan dan diagram alir metode yang dihasilkan.

## 4. PEMBAHASAN

Pada bagian ini membahas tahap pemahaman data, tahap persiapan data, tahap pemodelan, tahap evaluasi, dan tahap penyebaran data dalam memberikan analisis terhadap data serta rekomendasi strategi yang dapat dilakukan.

## 5. KESIMPULAN DAN SARAN

Pada bagian ini berisi kesimpulan berdasarkan temuan penelitian dan saran untuk pengembangan lebih lanjut.

## II. TINJAUAN PUSTAKA

### 2.1 Profil Perusahaan

Yayasan XYZ merupakan *online educational platform* untuk berbagi keahlian dan pengetahuan yang dapat digunakan secara luas baik untuk komunitas maupun individu. Platform ini dirancang sebagai pilihan yang tepat untuk membuat pelatihan bagi dan oleh siapapun. Pada platform ini dapat mencari dan mendapatkan kursus dengan cara yang jauh simple dan efisien. Aktivitas pembelajaran akan memberikan pengalaman yang luar biasa dengan dukungan konten pembelajaran yang terstruktur, mudah digunakan, dan menarik.

Yayasan XYZ terdapat tiga kegiatan yaitu kursus daring untuk umum, pelatihan untuk korporasi dan perguruan tinggi. Pada kegiatan kursus daring untuk umum menyediakan wadah bagi siapapun untuk bisa mencari, meningkatkan keahlian, belajar hal yang baru sesuai dengan minat dan keahlian yang diinginkan, menyediakan wadah bagi siapapun khususnya bagi pengajar yang memiliki keahlian atau pengetahuan dengan membuat kursusnya sendiri dan mendapatkan jangkauan yang lebih luas dan juga memberikan kesempatan bagi siapapun untuk bisa mendapatkan penghasilan tambahan dari hasil penjualan kursusnya atau bekerjasama dalam penjualan kursusnya pengajar lainnya di *platform*.

Pada kegiatan pelatihan untuk korporasi dimana Yayasan XYZ membantu perusahaan untuk meningkatkan kemampuan maupun keahlian dari SDM nya. Sedangkan pada kegiatan perguruan tinggi dimana bekerjasama dengan beberapa perguruan tinggi serta para praktisi & *expert* di berbagai bidang dan jurusan di Indonesia. Program ini membuka peluang bagi para mahasiswa untuk bisa meraih pengalaman dan keahlian yang sesuai dengan kebutuhan dari dunia kerja dan mencetak lulusan yang siap kerja.

### **2.1.1 Visi dan Misi Perusahaan**

Visi dari Yayasan XYZ yaitu menjadi *platform* komunitas edukasi terdepan berbasis keterampilan dan keahlian. Sedangkan misi dari Yayasan XYZ yaitu menciptakan efisiensi pembelajaran yang bermutu tinggi, menjadi *platform* terintegrasi dengan dunia kerja dan dunia bisnis, terus menerus menciptakan individu dengan kemampuan yang berkualitas, meningkatkan ekosistem pembelajaran yang terdepan dengan dukungan teknologi, dan menciptakan keselarasan kemampuan dan keahlian tenaga kerja dengan kepentingan perusahaan dan pelanggan.

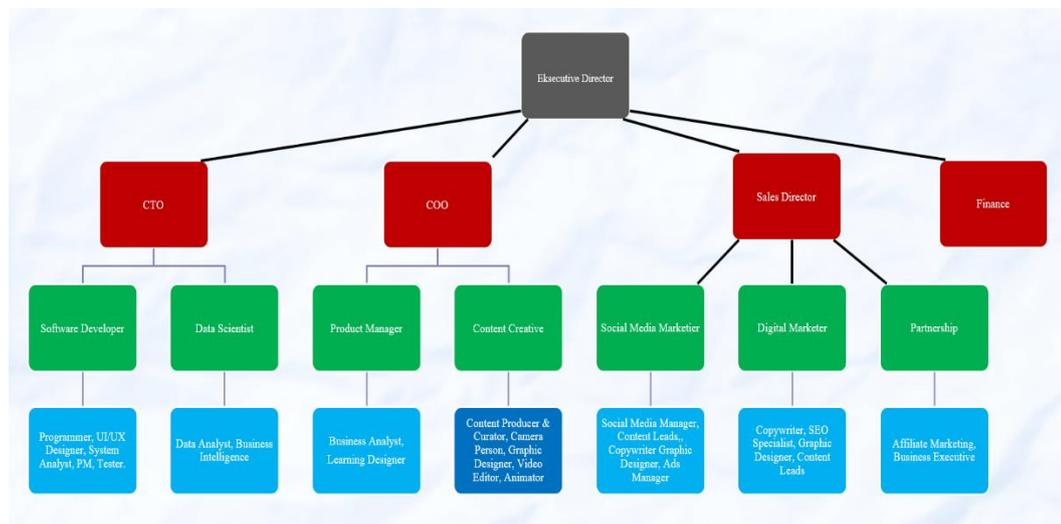
### **2.1.2 Nilai-nilai Perusahaan**

Nilai perusahaan Yayasan XYZ terdiri dari 5 nilai yaitu *leadership*, *equality*, *dependable*, *sustainability*, dan *teamwork*. Berikut penjelasan dari 5 nilai yang dijalankan :

1. *Leadership*, terdepan dalam kreativitas dan inovasi

2. *Equality*, mendukung prinsip kesetaraan
3. *Dependable*, memberikan pelayanan terbaik yang handal
4. *Sustainability*, mendukung prinsip pembangunan berkelanjutan
5. *Teamwork*, menjunjung tinggi kerjasama tim untuk mencapai tujuan bersama

### 2.1.3 Struktur Organisasi Perusahaan



Gambar 2.1 Struktur Organisasi Perusahaan

Berdasarkan gambar 2.1, dapat dilihat bahwa perusahaan dikepalai oleh seorang *Executive Director*. Kemudian terdapat 4 wakil direktur yaitu *Chief Technology Office (CTO)*, *Chief Operating Officer (COO)*, *Sales Director* dan *Finance*. *Chief Technology Office (CTO)* bertanggung jawab dalam kegiatan teknologi informasi dan menangani departemen yang bergerak dibidang IT perusahaan, seperti *Software Developer* dan *Data Scientist*. *Chief Operating Officer (COO)* bertanggung jawab dalam kebijakan operasional perusahaan dan menangani departemen *Product*

*Manager dan Content Creative. Sales Director* menangani departemen *Social Media Marketer, Digital Marketer dan Partnership*. Sedangkan *Finance* bertanggung jawab dalam bagian keuangan perusahaan.

## **2.2 Segmentasi Konsumen**

Segmentasi adalah proses membagi pasar yang heterogen menjadi kelompok-kelompok yang homogen. Pengelompokan tersebut dilakukan berdasarkan sifat kelompok demografi, geografi, psikografi dan perilaku [1].

Untuk menjangkau konsumen agar lebih efektif, segmentasi konsumen dapat menghasilkan kelompok-kelompok yang memiliki beberapa kriteria. Hal tersebut dilakukan dengan tujuan strategi yang diterapkan pada setiap segmen konsumennya menjadi lebih terstruktur dan spesifik. Karakteristik yang dapat menjadi acuan antara lain [2]:

1. Segmentasi geografis akan akan membagi konsumen ke dalam kelompok pasar yang berbeda sesuai dengan ukuran wilayah geografis mereka, seperti wilayah, iklim, kota atau desa.
2. Segmentasi demografis akan mengelompokkan konsumen ke dalam beberapa segmen pasar berdasarkan variabel demografis seperti usia, ukuran keluarga, kehidupan keluarga, jenis kelamin, pekerjaan, pendapatan, agama, kelas sosial, dan ras.
3. Segmentasi psikografis akan mengelompokkan konsumen pada kelompok pasar dengan perubahan gaya hidup dan kepribadian.

4. Segmentasi perilaku atau *behavior* akan mengelompokkan konsumen menjadi kelompok-kelompok pasar berdasarkan pengetahuan konsumen, sikap, dan respon yang diberikan konsumen terhadap suatu produk.

Dengan menerapkan segmentasi konsumen, kegiatan pemasaran yang dilakukan lebih tepat sasaran dan sumber daya perusahaan yang ada dapat digunakan secara efisien dan efektif guna menjamin kepuasan pelanggan [3]. Perusahaan-perusahaan di Indonesia juga sudah banyak menerapkan segmentasi terhadap konsumennya.

Keuntungan lain yang diberikan dengan adanya segmentasi konsumen yaitu menjadikan perusahaan selangkah lebih maju dibandingkan kompetitor karena telah melakukan kegiatan yang lebih spesifik dalam mengidentifikasi konsumen yang *high priority* dan *low priority* maupun konsumen yang berpotensi dan tidak berpotensi sehingga dapat menentukan strategi pemasaran berdasarkan karakteristik yang dimiliki pada setiap segmen.

### **2.3 Strategi Pemasaran**

Sebuah perusahaan atau suatu organisasi harus mengembangkan strategi agar dapat bertahan dan berkembang di masa depan. Hal tersebut diperlukan untuk berpikir bagaimana nanti di masa depan, bukan berpikir tentang bagaimana dan apa yang harus dilakukan perusahaan dalam jangka panjang. Rencana strategi yang disusun oleh perusahaan dengan memaksimalkan sumber daya tersebut digunakan untuk mencapai tujuan akhir perusahaan yaitu meningkatkan pendapatan perusahaan [4].

Pemasaran adalah proses dan desain untuk menciptakan, mengomunikasikan, menyampaikan, dan mempertukarkan penawaran nilai tambah kepada pelanggan dan masyarakat umum. Selain itu, pemasaran juga mencakup kegiatan untuk memenuhi keinginan dan kebutuhan dalam upaya mempengaruhi konsumen agar mau membeli barang dan jasa perusahaan [2].

Oleh karena itu, strategi pemasaran adalah rencana yang disusun secara sistematis dan dapat sepenuhnya digunakan sebagai panduan dalam melakukan kegiatan pemasaran untuk mencapai tujuan bisnis akhir yaitu mendapatkan keuntungan atau menghasilkan pendapatan. Langkah-langkah dalam menerapkan strategi pemasaran antara lain melakukan segmentasi pasar, menentukan target pasar dan *market positioning* [5]. Dengan hal tersebut, artinya terdapat hubungan bahwa kegiatan segmentasi merupakan langkah dalam implementasi suatu strategi pemasaran.

## **2.4 Data Mining**

### **2.4.1 Pengertian Data Mining**

*Data mining* adalah proses mengekstraksi/menggali/mencari informasi yang tersembunyi dari dalam database yang besar serta bermanfaat bagi pihak yang berkepentingan. *Data mining* dapat membantu pemegang data untuk menganalisis dan menemukan hubungan yang tidak terduga antar data, pada akhirnya akan membantu pengambilan keputusan [6].

### 2.4.2 Pengelompokan *Data Mining*

*Data mining* dapat dikelompokkan menjadi 2 kategori utama, yaitu [7]:

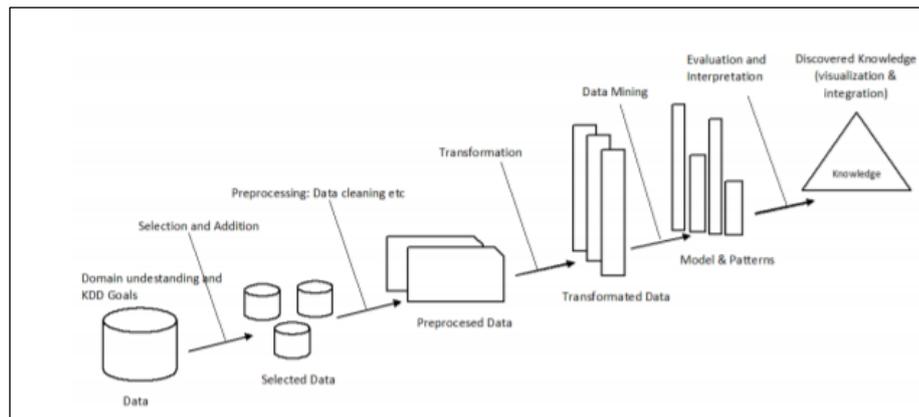
1. *Descriptive mining*, merupakan suatu proses dalam menemukan suatu karakteristik dari data yang masih dalam satu basis data. Contoh teknik data mining yang termasuk kelompok *descriptive mining* yaitu *clustering*, *association*, dan *sequential mining*.
2. *Predictive mining*, merupakan suatu proses dalam menemukan pola dari data dengan menggunakan beberapa variabel lain. Contoh teknik data mining yang termasuk kelompok *predictive mining* yaitu klasifikasi.

Sedangkan pengelompokan berdasarkan tugas, data mining dapat dikelompokkan sebagai berikut [8] :

1. *Classification* adalah teknik untuk melihat perilaku dan sifat atribut dari kelompok tertentu yang telah didefinisikan sebelumnya. Dengan menggunakan teknik ini, dimungkinkan untuk mengklasifikasikan data baru dengan memanipulasi data yang sebelumnya telah diklasifikasi dan hasilnya akan digunakan untuk memberikan sejumlah aturan.
2. *Association* adalah teknik mengenali sifat peristiwa khusus yang terjadi pada waktu tertentu. Sehingga dapat menganalisa kemungkinan yang akan terjadi setelah mengenali kelakuan tersebut.
3. *Clustering* adalah teknik menganalisis kelompok yang berbeda dengan data yang memiliki kesamaan atau kemiripan suatu objek. *Clustering* akan membagi objek ke dalam beberapa kelompok berdasarkan apa yang ditemukan dari data yang berjumlah besar tersebut.

### 2.4.3 Tahapan *Data Mining*

Tahapan data mining yang digunakan yaitu model proses *Knowledge Discovery Databases* atau KDD yang memiliki 9 langkah yang dapat dilihat pada Gambar 2.1 [9].



Gambar 2.2 Tahapan Data Mining

Adapun setiap tahapan dalam metode ini adalah sebagai berikut :

1. *Domain Understanding and KDD Goals* adalah proses dalam mengembangkan dan pemahaman tentang domain aplikasi dan pengetahuan sebelumnya.
2. *Selection and Additions* adalah proses penentuan data target dan subnet dari data sampel atau variabel.
3. *Data cleaning* adalah proses untuk menghilangkan noise sehingga menghasilkan sekumpulan data yang konsisten tanpa adanya noise dan kesalahan cetak (*tipografi*).
4. *Data transformation* adalah proses mengubah data dari satu bentuk ke bentuk lainnya sehingga data tersebut siap digunakan dalam proses data mining.

5. *Data mining* adalah proses memilih metode, algoritma yang tepat dalam pencarian pola-pola data, dan algoritma yang telah dipilih yaitu algoritma K-Means kemudian akan diimplementasikan sesuai dengan tujuan yang telah didefinisikan untuk mengekstrak pola data.
6. *Evaluation and Interpretation* adalah proses penyajian informasi/pola /pengetahuan dari hasil pengolahan dalam bentuk yang lebih dapat dipahami oleh pihak tertentu yang membutuhkan.
7. *Discovered Knowledge* adalah proses memutuskan hal apa yang akan dilakukan terhadap pengetahuan yang telah dihasilkan.

## 2.5 Clustering

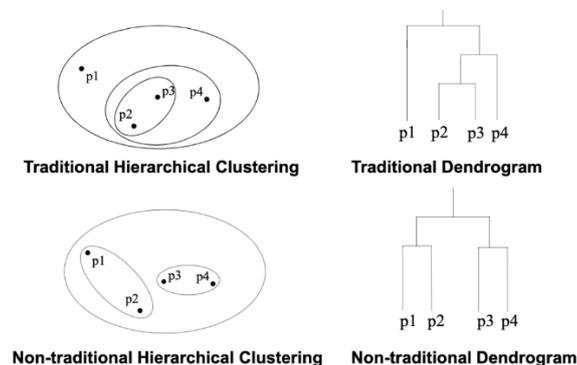
*Clustering* merupakan studi formal dalam melakukan partisi atau mengelompokkan data yang tidak memiliki label kategori. *Clustering* bersifat *unsupervised learning* artinya *clustering* tidak memiliki tahap dalam melakukan pelatihan data, hal tersebut berbeda dengan klasifikasi yang bersifat *supervised learning*. *Clustering* biasanya digunakan dalam mengelompokkan data yang memiliki kemiripan pada objek data dan sebaliknya meminimalkan kemiripan terhadap *cluster* lainnya [10].

Terdapat dua jenis metode *clustering* dalam melakukan pengelompokkan data, yaitu *hierarchical clustering* dan *non-hierarchical clustering* [11]. Pada *hierarchical clustering*, diasumsikan bahwa semua data merupakan *cluster*. Selanjutnya, pada pengulangan berikutnya, *cluster* akan berkurang secara iteratif dan *cluster* yang mempunyai jarak terdekat akan digabungkan, hingga iterasi

tersebut berhenti ketika jumlah *cluster* sudah sesuai dengan jumlah yang diinginkan. Contoh dari *hierarchical clustering* yaitu *Centroid Linkage Clustering* [12]. Sedangkan *non-hierarchical clustering*, pengelompokan dimana objek akan dikelompokkan ke dalam beberapa jumlah *cluster* yang telah ditentukan sebelumnya, sehingga data dengan karakteristik yang sama dapat dikelompokkan ke dalam kelompok yang sama dan data dengan karakteristik yang berbeda dapat dikelompokkan ke dalam cluster yang lain [11].

## 2.6 K-Means

Algoritma K-Means adalah salah satu metode *non-hierarchical clustering* dan dapat digunakan pada data yang tidak memiliki label. K-Means berusaha mempartisi atau membagi data menjadi satu atau lebih *cluster* sehingga terbentuk satu atau lebih *cluster* dengan masing-masing mempunyai karakteristik data yang sama sedangkan antar *cluster* memiliki karakteristik data yang berbeda. Tujuan utama dilakukan *clustering* yaitu untuk meminimalisasikan perbedaan karakteristik di setiap *cluster* dan memaksimalkan perbedaan karakteristik antar *cluster* [13].



Gambar 3.1 Hirarchical dan Non Hirarchical Clustering

Pada tahap awal algoritma K-Means, pusat cluster dipilih secara acak dari kumpulan data populasi. Metode K-Means kemudian menguji setiap komponen dari populasi data dan menetapkan komponen ke salah satu pusat *cluster* yang ditentukan berdasarkan jarak minimum antara komponen dan setiap *cluster*. Terakhir, pusat *cluster* akan dihitung ulang hingga semua komponen data digolongkan ke dalam tiap-tiap pusat *cluster* dan terbentuk pusat cluster baru. Berikut langkah-langkah dalam pengelompokan data menggunakan algoritma K-Means [14] :

1. Tentukan jumlah *cluster*
2. Inisialisasi awal dan pusat *cluster* dilakukan secara random
3. Menghitung jarak dari setiap objek dengan pusat *cluster* (*centroid*) berdasarkan kemiripan atau ketidakmiripan data menggunakan metode jarak *Euclidean Distance*. *Euclidean Distance* ini berfungsi untuk mendapatkan nilai pusat *cluster* (*centroid*) yang baru hingga nilai pusat *cluster* (*centroid*) tetap dengan rumus berikut :

$$d_{(x,y)} = \sqrt{\sum_{i=1}^n (x_i - y_j)^2}$$

Keterangan :

$d_{(x,y)}$  = ukuran ketidakmiripan

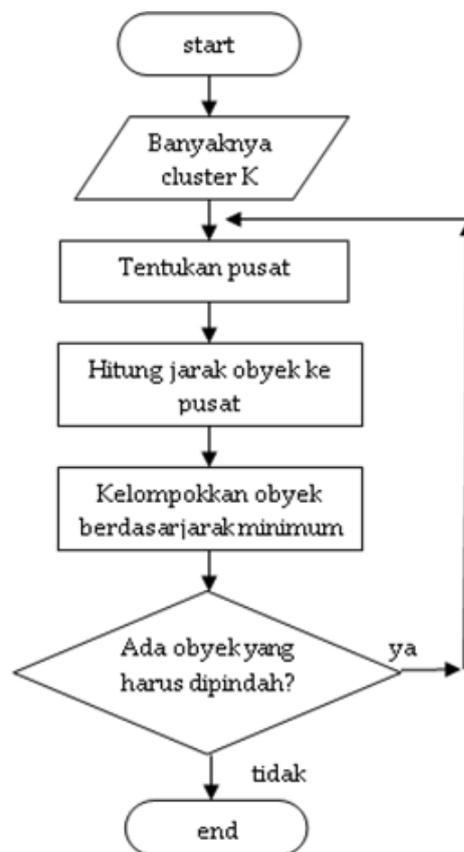
$x_i$  =  $(x_1, x_2, x_3, \dots, x_i)$  yaitu variabel data

$y_j$  =  $(y_1, y_2, y_3, \dots, y_j)$  yaitu variabel pada titik pusat

Jarak terdekat dari satu data ke satu *cluster* akan menentukan data tersebut masuk ke dalam *cluster* yang mana [15].

4. Hitung pusat *cluster* yang baru dengan keanggotaan yang baru di setiap *cluster* dengan cara menghitung rata-rata objek pada *cluster*.
5. Hitung kembali jarak tiap objek dengan pusat *cluster* yang baru, hingga *cluster* tidak berubah, maka proses pengklasteran selesai.

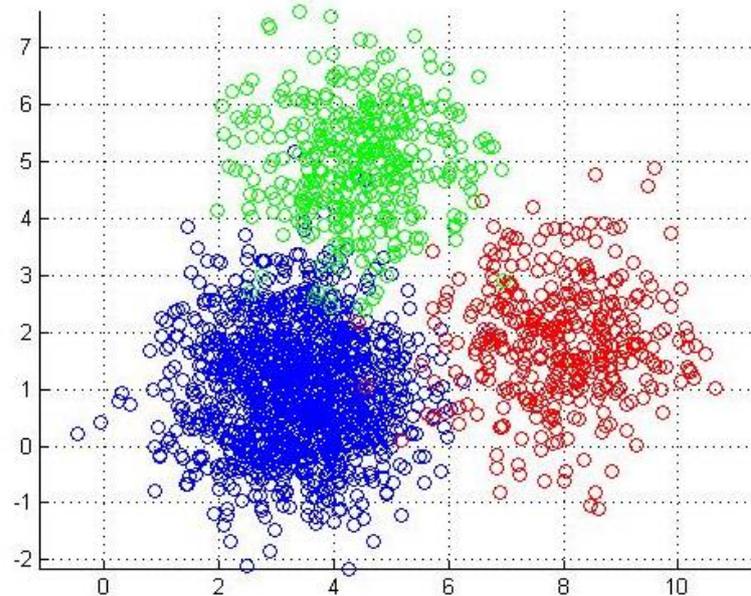
Untuk lebih rinci, diagram alir algoritma K-Means yaitu:



Gambar 3.2 Flowchart Algoritma K-Means

Pada Gambar 3.2 Flowchart Algoritma K-Means, algoritma K-Means pada awalnya menentukan banyaknya cluster, kemudian mengambil sebagian objek dari populasi untuk dijadikan titik pusat yang dipilih secara random sebagai pusat cluster. Setelah itu, algoritma K-Means akan menghitung jarak setiap objek ke titik pusat yang telah dipilih secara random. Nilai jarak objek ke suatu cluster yang

paling minimum akan dikelompokkan ke dalam cluster tersebut. Kemudian dihitung kembali jarak setiap objek ke pusat cluster yang baru, jika terdapat perubahan cluster maka kembali menghitung jarak setiap objek ke *centroid* yang baru namun jika tidak terdapat perubahan maka iterasi selesai.



Gambar 2.3 Contoh *Cluster K-Means*

### 2.6.1 Keunggulan dan Kelemahan Metode *K-Means*

Metode K-Means memiliki kelebihan dan kekurangan, dan kelebihan dari metode K-Means adalah dapat mengelompokkan data yang besar dengan cepat. Namun kelemahan dari metode k-means adalah jumlah cluster harus diketahui terlebih dahulu [16].

Selain itu, kelebihan yang dimiliki k-means yaitu pengimplementasian mudah untuk dilakukan dan dijalankan, prinsip yang digunakan sederhana sehingga dapat dijelaskan dalam non-statistik, dan k-means sudah sangat umum digunakan.

Sedangkan kelemahan dari metode k-means adalah penentuan nilai k yang dilakukan secara acak, pengelompokan data yang diperoleh akan berbeda dan penentuan nilai k tidak memiliki jaminan merupakan jumlah k yang optimal [17].

## 2.7 Model RFM

Model RFM merupakan singkatan dari *Recency*, *Frequency* dan *Monetary*, model ini sering digunakan perusahaan untuk melakukan segmentasi pelanggan sebagai salah satu upaya pemasaran yang dapat dilakukan. Model RFM akan menggunakan data informasi konsumen di masa lalu untuk mengevaluasi dan memprediksi perilaku dan nilai konsumen [18].

Selain itu, model RFM ini dapat digunakan untuk mengidentifikasi konsumen yang lebih mungkin cepat merespons terhadap penawaran baru [19].

Model RFM melibatkan 3 variabel yang terdiri atas *Recency*, *Frequency* dan *Monetary* yang digunakan untuk mengukur riwayat transaksi konsumen [20].

Penjelasan dari ketiga variabel RFM yaitu :

a. *Recency*

*Recency* merupakan jarak dari waktu konsumen terakhir kali melakukan transaksi dengan waktu saat ini atau seberapa baru konsumen melakukan transaksi

b. *Frequency*

*Frequency* merupakan jumlah transaksi yang dilakukan konsumen dalam periode tertentu atau seberapa sering konsumen melakukan transaksi

c. *Monetary*

*Monetary* merupakan total nilai transaksi yang dilakukan konsumen

Model RFM memiliki beberapa alasan mengapa model tersebut sering digunakan dan populer, antara lain hemat biaya dalam memperoleh *behavior* atau perilaku konsumen untuk dianalisis dan diukur, digunakan untuk memprediksi respon sehingga dapat meningkatkan keuntungan perusahaan, efektif dalam memodelkan variabel menjadi sejumlah variabel yang lebih kecil, dan terakhir model RFM ini sudah lama dikenal dalam mengukur dan mengidentifikasi konsumen yang berharga.

Namun model RFM juga memiliki kelemahan yaitu hanya berfokus dengan konsumen yang telah melakukan transaksi dan tidak bisa diimplementasikan oleh pelanggan baru [21].

## **2.8 Detect Outlier (Distance)**

*Detect outlier* atau *anomaly detection* merupakan proses dalam menemukan pola dari kumpulan data tertentu yang menyimpang dari sifat mayoritas di data tersebut. *Detect outlier* ini akan menemukan *noise* agar memastikan model yang diusulkan tersebut paling cocok dengan data. Selain itu, *detect outlier* merupakan salah satu upaya dalam menghilangkan anomali sebagai pembersihan data [22].

*Anomaly detection* atau *detect outlier* sebenarnya merupakan salah satu pendekatan dalam melakukan klusterisasi. Metode *machine learning* yang dapat menggunakan *anomaly detection* yaitu K-Nearest Neighbor, K-Means, DBSCAN dan lain-lain [23]. Pada *tools* RapidMiner, selain *Detect Outlier (Distances)* juga

terdiri atas *Detect Outlier (Class Outlier Factors)*, *Detect Outlier (Densities)*, dan *Detect Outlier (Local Outlier Factors)*.

## 2.9 *Min-Max Normalization*

Normalisasi setiap atribut pada data diperlakukan secara mandiri dan strategi normalisasi diterapkan untuk menghilangkan bias tertentu yang ada pada data. Normalisasi data penting dalam transformasi data, dikarenakan data dalam dunia nyata biasanya direkam di skala yang berbeda, sehingga dengan melalui normalisasi, data tersebut dikonversi menjadi bentuk universal untuk perbandingan [24].

*Min-Max Normalization* merupakan pendekatan normalisasi data yang melakukan transformasi linear terhadap data. Normalisasi *Min-Max* dapat dihitung dengan perhitungan seperti berikut :

$$x'_i = \frac{(x_i - x_{min})}{(x_{max} - x_{min})} ((new_{max} - new_{min}) + new_{min})$$

Keterangan :

$x'_i$  = Nilai yang telah dinormalisasikan

$x_i$  = Nilai asli

$x_{min}$  = Nilai terkecil atribut

$x_{max}$  = Nilai terbesar atribut

$new_{max}$  = Rentang maksimal dengan nilai 1

$new_{min}$  = Rentang minimal dengan nilai 0

Beberapa metode data normalization lainnya yaitu, *Z-Score*, *Decimal Scaling*, *Sigmoidal* dan *Softmax*. Dalam kasus tertentu, metode *Min-Max* dan *Z-Score* dapat meningkatkan kinerja model [23].

## 2.10 Penentuan Jumlah Cluster

Penentuan jumlah *cluster* menggunakan *Davies Bouldin Index*, nilai tersebut akan menunjukkan jumlah *cluster* yang optimal. *Davies Bouldin Index* (DBI) diperkenalkan oleh David L. Davies dan Donald W. Boildin pada tahun 1979. DBI adalah suatu ukuran yang digunakan dalam menentukan jumlah *cluster* terbaik setelah proses clusterisasi selesai [25]. Jika nilai DBI yang diperoleh semakin kecil mendekati 0 dan tidak negatif ( $\geq 0$ ) maka semakin baik *cluster* yang diperoleh dari *Cluster K-Means* yang digunakan [26].

Langkah-langkah untuk menghitung *Davies Bouldin Index* adalah sebagai berikut [27] :

### 1. *Sum of Square Within-Cluster* (SSW)

SSW digunakan dalam mengetahui kohesi dari sebuah *cluster* ke-i salah satunya adalah dengan menghitung nilai dari *Sum of Square Within-Cluster* (SSW) dengan rumus sebagai berikut :

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(X_j, C_i)$$

Keterangan :

$m_i$  = jumlah data dalam *cluster* ke-i

$c_i$  = *centroid cluster* ke-i

$d(X_j, C_i)$  = jarak setiap data ke *centroid* i yang dihitung menggunakan jarak *euclidean*

## 2. *Sum of Square Between-Cluster (SSB)*

SSB digunakan untuk mengetahui jarak antar *cluster* dengan rumus sebagai berikut :

$$SSB_{i,j} = d(C_i, C_j)$$

Keterangan :

$d(C_i, C_j)$  = jarak *centroid* ke-i dengan *centroid* ke-j di *cluster* lain

## 3. *Ratio (Rasio)*

Rasio digunakan untuk mengetahui nilai perbandingan antara *cluster* ke-i dan *cluster* ke-j untuk menghitung nilai rasio yang dimiliki oleh masing-masing *cluster*. Indeks i dan j akan merepresentasikan jumlah *cluster*, dalam menentukan nilai rasio menggunakan rumus sebagai berikut :

$$R_{i,j,n} = \frac{SSW_i + SSW_j + \dots + SSW_n}{SSB_{i,j} + \dots + SSB_{n,i,j}}$$

Keterangan :

$SSW_i$  = *Sum of Square Within-Cluster* pada *centroid* i

$SSB_{i,j}$  = *Sum of Square Between Cluster centroid* ke i dengan j pada *cluster* yang berbeda

## 4. *Davies Bouldin Index (DBI)*

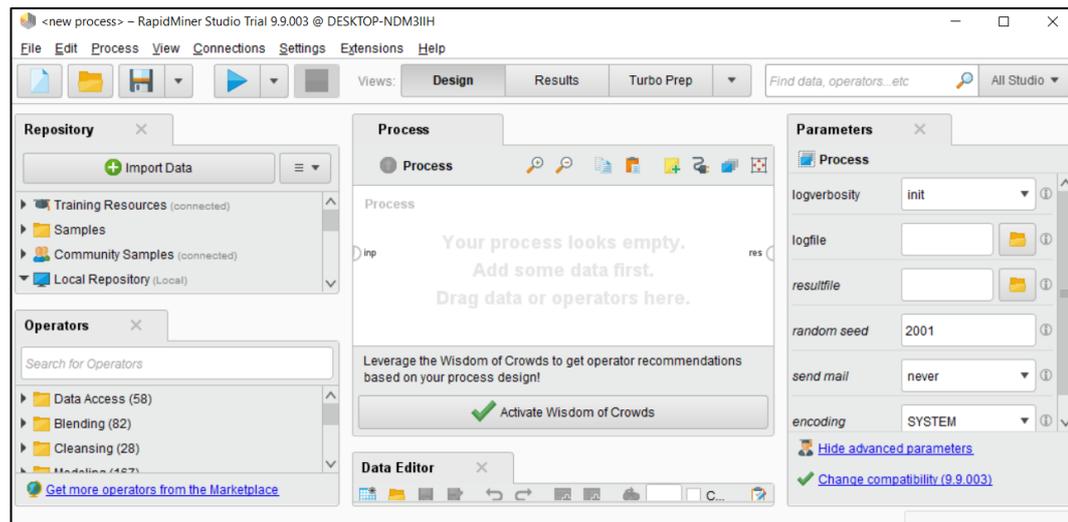
Nilai rasio yang telah didapatkan maka akan digunakan untuk mengetahui nilai DBI dengan rumus sebagai berikut :

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j,\dots,k})$$

## 2.11 RapidMiner

RapidMiner merupakan *tools* yang dibuat oleh Dr. Markus Hofmann dari *Institute of Technology Blanchardstown* dan Raif Klinkenberg dari rapid-i.com. RapidMiner ini dirancang dengan tampilan GUI (*Graphical User Interface*) yang akan mempermudah pengguna dalam mengaplikasikan *tools* RapidMiner. Selain itu, sifatnya yang *open source*, bahasa java dibawah lisensi GNU *Public Licence* dan dapat dijalankan dalam sistem operasi jenis apapun merupakan suatu kelebihan dari *tools* RapidMiner. Selain itu, dalam menggunakan RapidMiner tidak membutuhkan keahlian coding yang khusus dikarenakan seluruh fasilitas pada *tools* RapidMiner telah disediakan. Namun *tools* RapidMiner ini dikhususkan dalam penggunaan jumlah data yang besar [28].

RapidMiner menyediakan lingkungan terpadu dalam sebagai pembelajaran mesin (*machine learning*), pembelajaran mendalam (*deep learning*), penambangan teks (*teks mining*), dan analisis prediktif. *Tools* RapidMiner cocok digunakan untuk aplikasi bisnis maupun komersial, penelitian, pendidikan, pelatihan, pembuatan *prototype* dengan cepat, dan pengembangann aplikasi serta sangat mendukung dalam semua langkah proses pembelajaran mesin termasuk dalam mempersiapkan data, visualisasi hasil, validasi dan pengoptimalan [29].

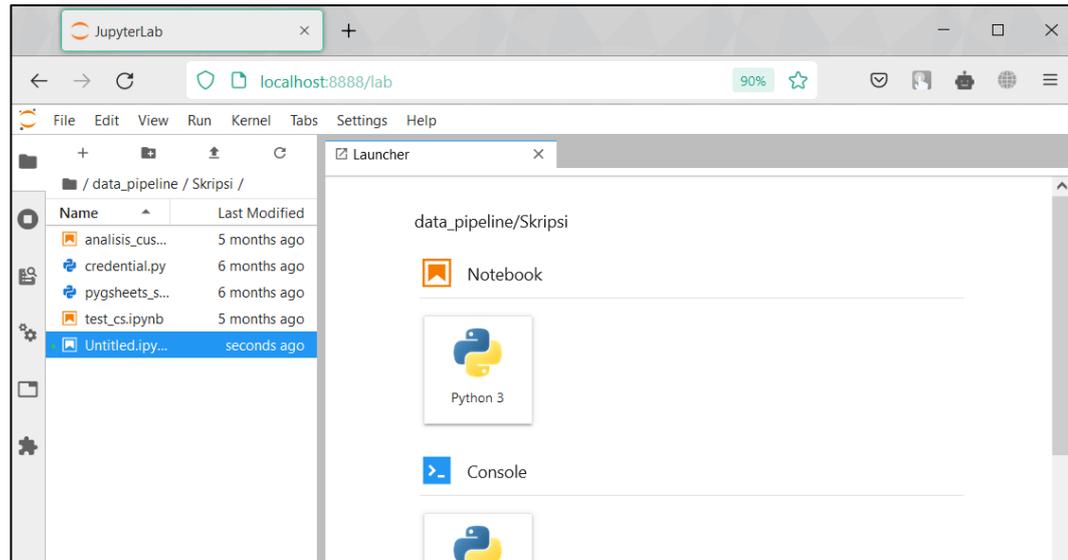


Gambar 2.4 Tampilan RapidMiner

## 2.12 JupyterLab

Jupyterlab memungkinkan untuk mendokumentasikan dan reproduibilitas alur kerja dalam data science dimana data tersebut besar dan kompleks. Sehingga dengan adanya jupyterlab, maka peneliti dapat mendokumentasikan dalam melakukan pencatatan langkah-langkah dalam proses penelitian termasuk sumber data, transformasi dan hingga akhir menjawab pertanyaan dari manajemen atau klien. Selain itu, reproduibilitas dimana memiliki kemampuan untuk menjalankan kembali analisis sehingga data tersebut dapat diperbarui tanpa harus mengingat apa yang telah dilakukan. Terdapat dua cell dalam jupyter, yaitu [30]:

1. *Code cell* merupakan tempat kode dimasukkan
2. *Markdown cell* merupakan tempat teks dimasukkan



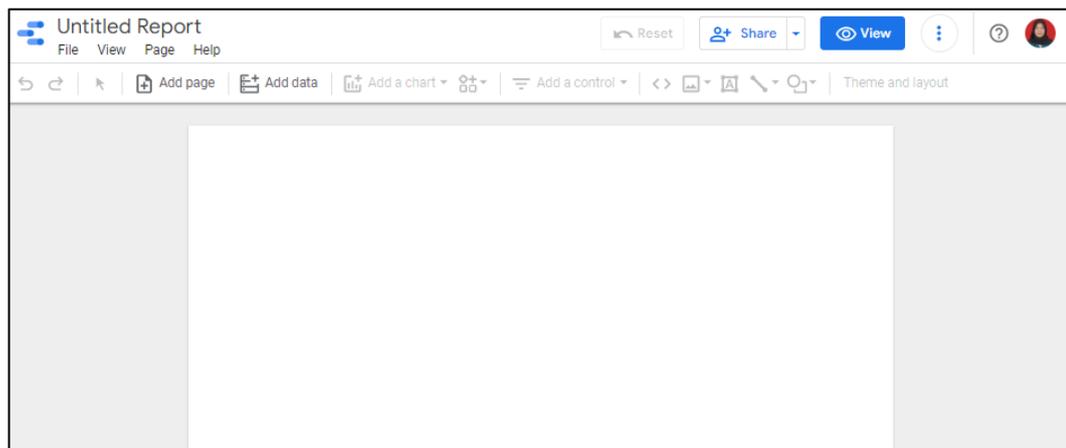
Gambar 2.5 Tampilan JupyterLab

### 2.13 Google Data Studio

Google Data Studio merupakan program yang digunakan dalam memvisualisasikan data agar mudah digunakan dan dapat mempresentasikan data yang kompleks menjadi menarik dan jelas. Google Data Studio diluncurkan secara gratis pada tahun 2016. Tersedia secara gratis dan untuk siapa saja yang memiliki akun Google, Google Data Studio ini juga merupakan solusi potensial dalam visualisasi data dan berfitur lengkap [31].

Keunggulan google data studio yaitu fleksibilitas, kemudahan menggunakan dan aksesibilitas. Kombinasi fitur yang baik membuat google data studio fleksibel dalam menangani tugas yang besar. Kemudahan dalam menggunakannya membuat pengguna merasa cepat memahaminya. Serta aksesibilitas, dimana google data studio dapat digunakan kepada siapa saja yang memiliki koneksi internet dan tidak perlu melakukan instalasi atau konfigurasi perangkat lunak [32].

Selain itu, google data studio ditawarkan secara gratis. Berbeda dengan kompetitor lainnya yaitu Tableau dan Power BI dalam membuat visualisasi data user perlu mengeluarkan biaya dalam menggunakan fitur-fitur yang ada didalamnya.

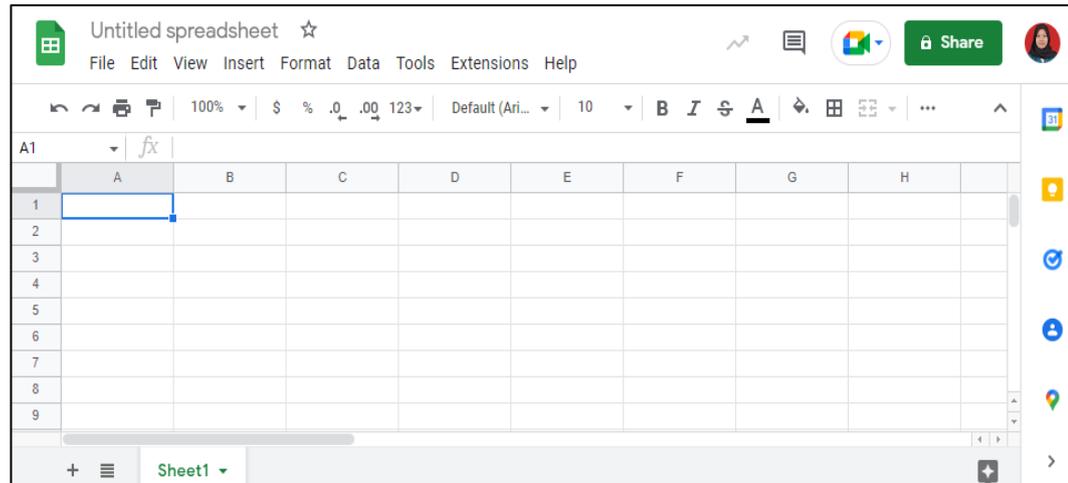


Gambar 2.6 Tampilan Google Data Studio

## 2.14 Google Sheet

*Google sheet* merupakan *software* yang setara dengan Microsoft Excel yang mulai dikembangkan di tahun 2005 oleh perusahaan 2Web Technologies [33]. Google sheet memiliki limit maksimal 256 kolom dan 20.000 baris per sheet. Sedangkan, file yang diunggah dikonversi ke format google sheet tidak boleh lebih besar dari 20 MB [34].

*Google sheet* mampu membuat spreadsheet yang rumit secara online, berbasis web sehingga menggunakan web browser, dan memiliki fungsi yang canggih. Selain itu, google sheet memungkinkan pengguna melakukan kolaborasi dan mengimpor data dari database secara *real time*.

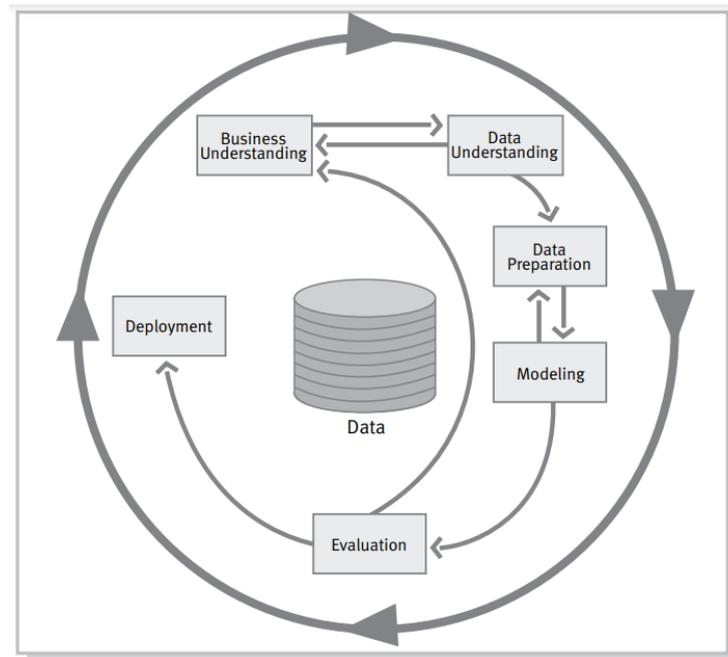


Gambar 2.7 Tampilan Google Sheet

## 2.15 CRISP-DM

*Cross Industry Standard Process for Data-Mining* atau CRISP-DM digunakan dalam dunia industri untuk menyelesaikan berbagai permasalahan bisnis dan memiliki keterkaitan dengan data mining.

Model CRISP-DM menampilkan siklus proyek dari data mining, dimana model CRISP-DM dibedakan menjadi 6 tahapan yaitu *Business Understanding* (Pemahaman Bisnis), *Data Understanding* (Pemahaman Data), *Data Preparation* (Persiapan Data), *Modeling* (Pemodelan), *Evaluation* (Evaluasi), dan terakhir *Deployment*



Gambar 2.8 Alur Model CRISP-DM [35]

Tahapan awal pada alur Model CRISP-DM yaitu tahapan *business understanding*, berdasarkan data yang dimiliki akan dilihat perspektif dari segi bisnis untuk mencapai tujuan bisnis. Selanjutnya, tahapan *data understanding*, pada tahapan ini data-data dikumpulkan dan dipahami, pada tahapan ini juga dapat kembali ke tahapan sebelumnya yaitu *business understanding* untuk memastikan bahwa data-data yang dikumpulkan dapat digunakan untuk mencapai tujuan. Selanjutnya *data preparation*, pada tahapan ini data-data yang telah dikumpulkan dilakukan persiapan data seperti menyeleksi, mengintegrasikan, membersihkan dan memformat data. Selanjutnya, tahapan *modeling*, pada tahapan ini melakukan pemilihan teknik model data mining yang ingin diterapkan, pada tahapan ini juga dapat kembali ke tahapan sebelumnya yaitu tahapan *data preparation*, hal tersebut dikarenakan jika dalam tahapan *modeling* ditemukan data-data yang tidak diinginkan dan mengganggu hasil pemodelan maka dapat kembali ke tahapan *data*

*preparation* untuk memperbaiki permasalahan data. Setelah tahapan *modeling* selanjutnya tahapan *evaluation*, pada tahapan ini akan melakukan evaluasi model yang telah dibangun untuk memastikan bahwa tujuan bisnis telah terjawab di tahapan *business understanding* dan telah memiliki nilai performa model yang baik. Pada tahapan *evaluation* terdapat siklus balik kembali ke *business understanding* dikarenakan tidak semua kasus harus dilakukan deployment, terdapat kasus jika sudah mendapatkan informasi pada tahapan *evaluation* akan melakukan iterasi kembali di tahapan *business understanding*.

Pada alur model CRISP-DM ini tidak statis akan ada iterasi-iterasi selanjutnya, dimana jika tujuan bisnis berkembang pada tahapan *business understanding* dikarenakan perubahan waktu maka dapat menyebabkan perubahan karakteristik data, data yang digunakan juga akan berkembang sehingga tahapan data preparation akan berubah bahkan model data mining yang digunakan juga dapat berubah dikarenakan perkembangan. Hal tersebut membuat alur Metode CRISP-DM terus melakukan iterasi.

Berikut adalah penjelasan dari setiap tahapan pada model CRISP-DM :

1. *Business Understanding* (Pemahaman Bisnis)

Pada tahapan ini bertujuan untuk memahami perspektif dari segi bisnis, cara mendapatkan data yang dibutuhkan dan mencocokkannya untuk memperoleh model pengetahuan terbaik yang dapat dihasilkan untuk mencapai tujuan.

2. *Data Understanding* (Pemahaman Data)

Pada tahapan ini mulai melakukan pengumpulan data awal, mengidentifikasi kualitas data yang ada, mendeteksi pengetahuan dari data untuk membentuk hipotesis dari informasi yang tersembunyi.

3. *Data Preparation* (Persiapan Data)

Pada tahapan ini melakukan persiapan data mencakup seluruh kegiatan dalam menyusun dataset akhir yang akan dijadikan pemodelan berasal dari data mentah. Tahapan ini juga cenderung dilakukan dalam beberapa kali dan contoh kegiatan yang dilakukan termasuk pemilihan atribut, pembersihan data, pembuatan atribut baru dan transformasi data.

4. *Modeling* (Pemodelan)

Pada tahapan ini melakukan pemilihan teknik pemodelan yang akan diterapkan.

5. *Evaluation* (Evaluasi)

Pada tahapan ini melakukan evaluasi model secara menyeluruh dan meninjau langkah-langkah yang dijalankan untuk membangun model serta memastikannya mencapai tujuan bisnis yang benar.

6. *Deployment* (Penyebaran)

Pada tahapan ini menyajikan data sehingga mudah dipahami dan dapat digunakan oleh pengguna [35].

## **2.16 Kategori Usia**

Menurut Kementerian Kesehatan Republik Indonesia tahun 2009 kategori usia yaitu [36]:

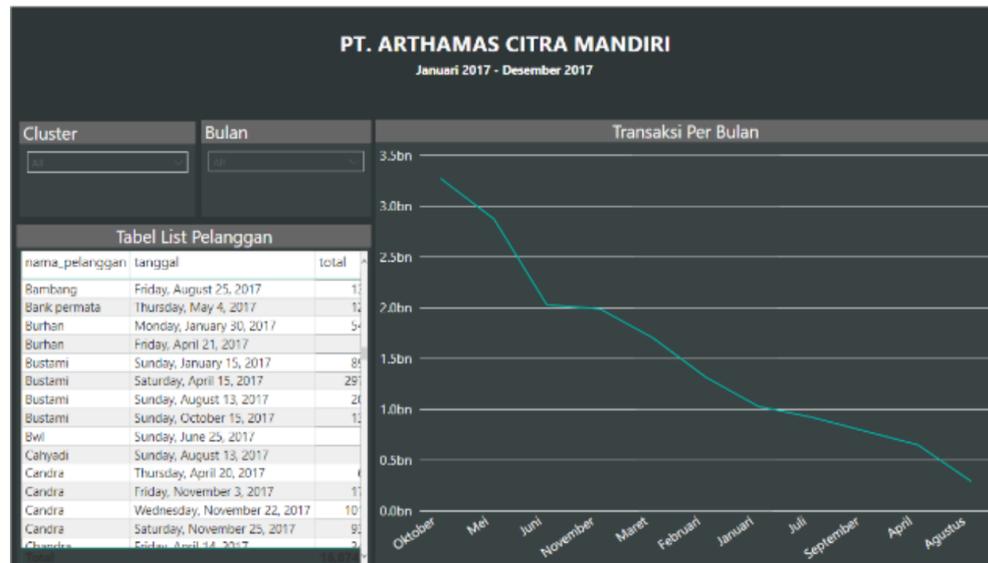
1. Masa balita usia 0 – 5 tahun
2. Masa kanak-kanak usia 5 – 11 tahun
3. Masa remaja awal usia 12 – 16 tahun

4. Masa remaja akhir usia 17 – 25 tahun
5. Masa dewasa awal usia 26 – 35 tahun
6. Masa dewasa akhir usia 36 – 45 tahun
7. Masa lansia awal usia 46 – 55 tahun
8. Masa lansia akhir usia 56 – 65 tahun
9. Masa manula usia 65 – ke atas

### 2.17 Penelitian Terkait

Dalam penyusunan skripsi ini sedikit banyak terinspirasi dari penelitian-penelitian terdahulu. Adapun penelitian yang berkaitan mengenai teknik data mining, model RFM, Algoritma K-Means dan metode CRISP-DM antara lain :

1. Penelitian yang dilakukan oleh Wiratama Ahsani Taqwim, Nanang Yudi Setiawan, dan Fitra A. Bachtiar yang berjudul “Analisis Segmentasi Pelanggan dengan RFM Model Pada PT. Arthamas Citra Mandiri Menggunakan Metode Fuzzy C-Means *Clustering*” pada tahun 2019. Metode yang digunakan yaitu Fuzzy C-Means *Clustering* berdasarkan model RFM (*Recency*, *Frequency*, dan *Monetary*). Data yang digunakan di dapat dari riwayat transaksi pelanggan periode Januari 2017 hingga Desember 2017 dengan jumlah 981 transaksi. Pada penelitian ini menghasilkan *cluster* 3 merupakan hasil terbaik. Kemudian hasil *cluster* divisualisasikan dengan dashboard dan mendapatkan hasil usability testing sebesar 77.5 artinya dashboard tersebut masuk kedalam kategori *acceptable* atau bisa diterima oleh perusahaan [37].



Gambar 2.9 Dashboard PT Arthamas Citra Mandiri

2. Penelitian selanjutnya yang dilakukan oleh Yulison Herry Chrisnanto dan Ade Kanianingsih yang berjudul “Pengelompokan Ekuitas Pelanggan Berbasis *Recency Frequency Monetary* (RFM) Menggunakan *K-Means Clustering*” pada tahun 2019. Menggunakan metode *K-Means Clustering* berdasarkan analisis RFM. Data yang digunakan merupakan transaksi yang berjumlah 20.814 selama tahun 2018, dengan jumlah pelanggan 894. Pada penelitian ini dihasilkan *cluster-1* sebanyak 7 pelanggan, *cluster-2* sebanyak 368 pelanggan dan *cluster-3* sebanyak 390 pelanggan [38].
3. Penelitian selanjutnya yang dilakukan oleh Christina Deni Rumiarti dan Indra Budi yang berjudul “Segmentasi Pelanggan pada Customer Relationship Management di Perusahaan Ritel : Studi Kasus PT Gramedia Asri Media” pada tahun 2017. Tujuan penelitian ini yaitu membuat segmentasi pelanggan di PT Gramedia Asri Media menggunakan algoritma *K-Means* berdasarkan RFM sedangkan dalam evaluasi menggunakan *elbow method*, *silhouette*

*method*, dan *Calinski-Harabasz index*. Data yang digunakan didapat dari data transaksi buku member Kompas Gramedia *Value Card* (KGVC). Pada penelitian ini berdasarkan RFM menghasilkan 2 *cluster* sedangkan berdasarkan banyaknya jenis buku yang dibeli menghasilkan 3 *cluster* [39].

4. Penelitian selanjutnya yang dilakukan oleh Setiawan, Husni Amani, dan Wawan Tripiawan yang berjudul “Perancangan Segmentasi Pelanggan dengan Metode *Clustering* K-Means dan Model RFM pada Klinik Kecantikan Seoul Secret” pada tahun 2021. Tujuan penelitian ini yaitu melakukan segmentasi pelanggan pengguna program paket treatment dan memberikan rekomendasi strategi pemasaran menggunakan metode *Clustering* K-Means dan penerapan model RFM. Data yang digunakan didapat dari data transaksi Klinik Kecantikan Seoul Secret. Pada penelitian ini menghasilkan rekomendasi strategi pemasaran yang akan diberikan oleh pihak Klinik Kecantikan Seoul Secret [40].
5. Penelitian selanjutnya yang dilakukan oleh Firman Maulana Aditya yang berjudul “Perancangan Dashboard sebagai Pendukung Keputusan Berdasarkan Data Antrean pada Aplikasi Antribos (Studi Kasus : Medical Center ITS)” pada tahun 2018. Tujuan penelitian ini merancang dashboard yang membantu pemilik layanan dalam mengambil keputusan dengan metode *User Centered Design* agar dashboard yang dibuat mudah dipahami oleh pengguna. Data yang digunakan didapat dari data antrean pada aplikasi Antribos. Pada penelitian ini menghasilkan dashboard Medical Center ITS dengan menggunakan *User Centered Design* dan metode pengujian *System*

*Usability Scale* menunjukkan pemahaman pengguna menggunakan dashboard dengan nilai 90% [41].



Gambar 2.10 Dashboard Medical Center ITS

6. Penelitian selanjutnya yang dilakukan oleh Donny Fernando yang berjudul “Visualisasi Data Menggunakan Google Data Studio” pada tahun 2018. Tujuan penelitian ini membuat visualisasi data penjualan untuk meringkas kinerja penjualan dengan lebih cepat dan mudah dipahami. Pada penelitian ini menghasilkan solusi alternatif dalam melakukan visualisasi data menggunakan Google Data Studio dengan memberikan kemudahan untuk mengintegrasikan dan membagikan laporan tanpa mengabaikan keamanan laporan tersebut [42].
7. Penelitian selanjutnya yang dilakukan oleh P. Anitha dan Malini M.Patil yang berjudul “RFM Model for Customer Purchase Behavior using K-means

*Algorithm*” pada tahun 2019. Tujuan penelitian ini menerapkan business intelligence dalam mengidentifikasi calon pelanggan dengan menyediakan data yang relevan. Data akan dilakukan analisis untuk meningkatkan penjualan, keuntungan bisnis, menambah wawasan dalam memprediksi perilaku dan pola konsumen dalam melakukan pembelian dengan menerapkan algoritma K-Means berdasarkan RFM (*Recency, Frequency, dan Monetary*) serta berdasarkan perhitungan *Silhouette Coefficient* untuk menentukan jumlah *cluster* yang divalidasi. Hasil yang diperoleh pada penelitian ini berkaitan dengan transaksi penjualan yang dianalisis dengan berbagai parameter seperti kekinian transaksi, jumlah transaksi dan total nilai transaksi [43].

8. Penelitian selanjutnya yang dilakukan oleh Genifer Snipes yang berjudul “*Product Review Google Data Studio*” pada tahun 2018. Tujuan penelitian ini melakukan *product review* dari Google Data Studio. Pada penelitian ini menghasilkan bahwa Google Data Studio merupakan solusi potensial untuk melakukan visualisasi data yang berfitur lengkap dan tersedia secara gratis dengan tujuan untuk membantu pengguna dalam membuat laporan visual yang dinamis dan menarik serta mengintegrasikan sumber data eksternal dapat dengan mudah [31].
9. Penelitian selanjutnya yang dilakukan oleh K.Archana dan K.G.Saranya yang berjudul “*Mall Customer Segmentation Using Clustering Algorithm*” pada tahun 2019. Tujuan penelitian ini yaitu melakukan analisis keranjang pasar untuk memprediksi target pelanggan sehingga memungkinkan tim pemasaran dalam merencanakan strategi untuk memasarkan produk baru dengan sasaran

kepada pelanggan yang memiliki minat serupa. Penelitian ini menghasilkan 5 *cluster* dimana setiap *cluster* memiliki karakteristik yang berbeda dan model pengelompokan yang digunakan mampu secara efektif memproses informasi yang besar [44].

10. Penelitian selanjutnya dilakukan oleh Onur Dogan, Ejder Aycin, Zeki Atıl Bulut yang berjudul “*Customer Segmentation by Using RFM Model and Clustering Methods : A Case Study in Retail Industry*” pada tahun 2018. Tujuan penelitian ini yaitu memberikan pemahaman pelanggan yang lebih baik, strategi yang dirancang dengan baik dan keputusan yang lebih efisien. Penelitian segmentasi pelanggan ini menggunakan *clustering* berdasarkan model RFM. Penelitian ini menghasilkan 4 *cluster* yaitu *cluster* Reguler berjumlah 644.081 pelanggan, *cluster High priority* berjumlah 514 pelanggan, *cluster Star* berjumlah 97 pelanggan dan *cluster Advanced* berjumlah 55.340 pelanggan [45].

### III. METODE PENELITIAN

#### 3.1 Waktu dan Tempat

Adapun waktu dan tempat penelitian dan pembuatan tugas akhir ini dilakukan pada :

Waktu : November 2021 – April 2022

Tempat : Yayasan XYZ, Jl. Lembah Nyiur Raya, Pondok Kelapa, Durensawit, East Jakarta City, Jakarta 13450.

Tabel 3.1 Jadwal Penelitian

| No | Aktifitas                | Waktu |   |   |   |     |   |   |   |     |   |   |   |     |   |   |   |     |   |
|----|--------------------------|-------|---|---|---|-----|---|---|---|-----|---|---|---|-----|---|---|---|-----|---|
|    |                          | Nov   |   |   |   | Des |   |   |   | Jan |   |   |   | ... |   |   |   | Apr |   |
|    |                          | 1     | 2 | 3 | 4 | 1   | 2 | 3 | 4 | 1   | 2 | 3 | 4 | 1   | 2 | 3 | 4 | 1   | 2 |
| 1. | Studi Literatur          |       |   |   |   |     |   |   |   |     |   |   |   |     |   |   |   |     |   |
| 2. | Fase Pemahaman<br>Bisnis |       |   |   |   |     |   |   |   |     |   |   |   |     |   |   |   |     |   |
| 3. | Fase Pemahaman<br>Data   |       |   |   |   |     |   |   |   |     |   |   |   |     |   |   |   |     |   |



pre-processing data seperti menghilangkan data yang tidak digunakan, menghitung nilai RFM, memeriksa kesalahan pada data yang lainnya.

## 2. RapidMiner

RapidMiner digunakan sebagai aplikasi dalam melakukan analisis data mining menggunakan metode dan algoritma yang diinginkan.

## 3. Google Sheet

Google Sheet digunakan untuk menyimpan data yang sudah dilakukan proses *pre-processing* pada aplikasi JupyterLab untuk dilakukan visualisasi pada Google Data Studio dan melakukan pengunduhan data untuk dianalisis pada aplikasi RapidMiner menggunakan algoritma yang diinginkan.

## 4. Google Data Studio

Google Data Studio digunakan untuk memvisualisasikan data berupa dashboard yang informatif agar lebih mudah dipahami dan dibagikan kepada *stakeholder*.

### **3.2.2 Bahan**

Dalam penelitian ini bahan penelitian didapat dari perusahaan yang terdiri atas data transaksi penjualan konsumen perusahaan dari tanggal 1 Januari 2021 hingga 25 November 2021 dan data lainnya akan dilakukan analisis sebagai bentuk rekomendasi strategi pemasaran.

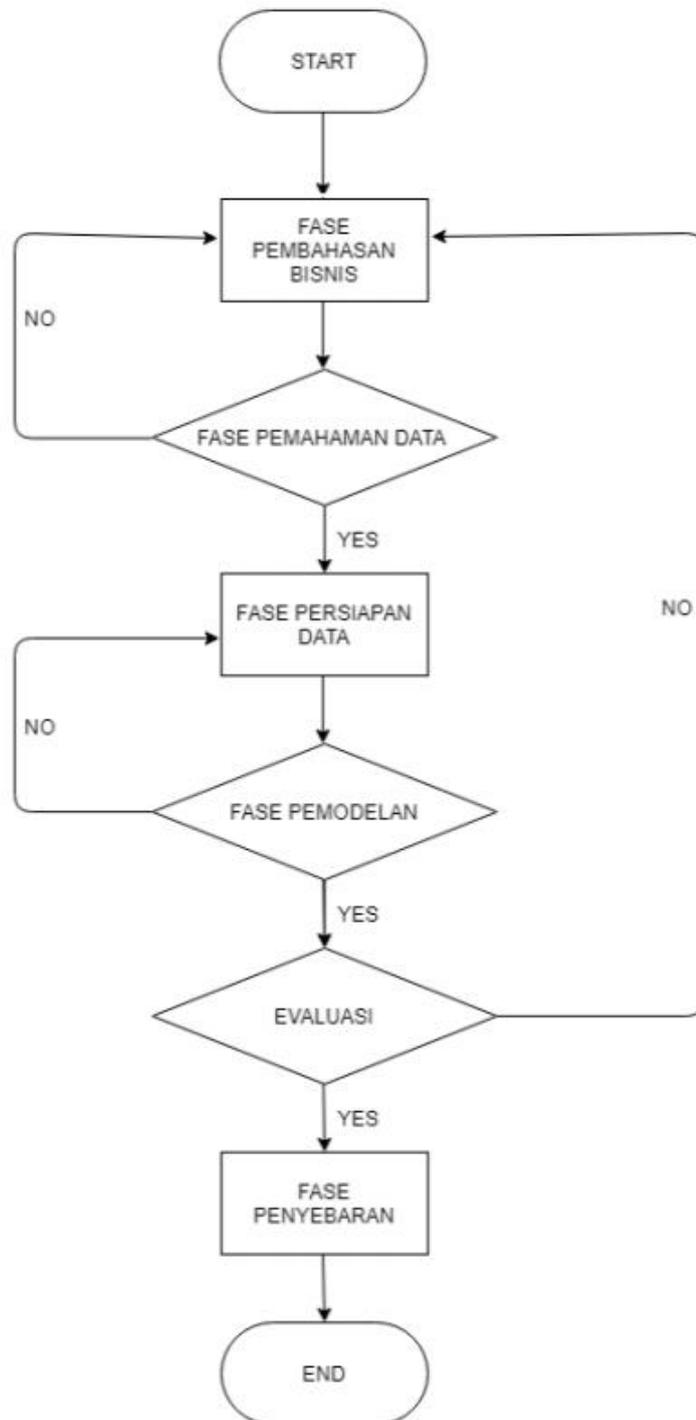
### **3.3 Tahap Penelitian**

Adapun tahap penelitian ini menggunakan metode *Cross Industry Standard Process for Data Mining* (CRISP-DM) dengan pengolahan data menggunakan Metode K-Means berdasarkan Model RFM . Berikut adalah tahap penelitian yang dilakukan yaitu :

#### **3.3.1 Studi Literatur**

Pada tahap studi literatur dilakukan untuk mempelajari ilmu dan penelitian yang terkait sebelumnya yang bersumber dari buku, jurnal, artikel dan prosiding. Ilmu yang dipelajari pada tahap studi literatur digunakan untuk mendukung penelitian ini, terkait data mining, klasifikasi, algoritma *K-Nearest Neighbor* dan metode *Cross Industry Standard Process for Data Mining* (CRISP-DM).

### 3.3.2 Penggunaan Metode CRISP-DM



Gambar 3.3 Tahap Penelitian Metode CRISP-DM

### 3.3.2.1 Fase Pemahaman Bisnis (*Business Understanding Phase*)

Tahap pemahaman bisnis ini melakukan beberapa hal seperti menentukan tujuan bisnis, menilai situasi, menentukan tujuan data mining, dan menghasilkan rencana project. Pada tahapan ini diketahui bahwa tujuan bisnis adalah memberikan rekomendasi strategi pemasaran berdasarkan analisis data transaksi perusahaan dan data lainnya yang dapat meningkatkan pendapatan perusahaan. Situasi yang berlangsung dalam perusahaan bahwa masih belum berfokus pada konsumen yang memberikan pendapatan lebih kepada perusahaan dikarenakan masih belum diketahui konsumen mana yang memberikan pendapatan tinggi kepada perusahaan sehingga kegiatan marketing tidak terarah dimana tidak memaksimalkan *behavior* yang dimiliki konsumen. Tujuan data mining dalam tahapan ini adalah melakukan cluster dalam mengetahui karakteristik *behavior* konsumen di setiap cluster, selain itu menemukan pola-pola yang terbentuk berdasarkan data perusahaan. Sehingga dihasilkan rencana project untuk melakukan analisis segmentasi konsumen menggunakan metode *k-means clustering* berdasarkan model RFM dan pengetahuan data lainnya sebagai rekomendasi strategi pemasaran untuk perusahaan. Sehingga perusahaan dapat dengan tepat melakukan pemasaran secara lebih efektif dan efisien. Berikut ketujuh analisis pengetahuan data yang akan dilakukan berdasarkan database perusahaan :

Tabel 3.2 Analisis Data dan Tujuan

| No | Analisis                 | Tujuan   |
|----|--------------------------|--|
| 1. | Data Segmentasi Konsumen | Melakukan analisis data transaksi menggunakan teknik data mining clustering dengan metode K-Means berdasarkan model RFM sehingga |

|    |  |   |
|----|--|---|
|    |  | dihasilkan cluster-cluster berdasarkan segmentasi perilaku transaksi yang telah dilakukan.  |
| 2  | Data Potensial Konsumen                        | Melakukan analisis terhadap pengguna yang belum pernah melakukan transaksi sukses, namun kedepannya berpotensi untuk membeli produk perusahaan atau melakukan transaksi sukses. |
| 3. | Data Kategori, Kursus dan Tipe Pembayaran      | Melakukan analisis kategori, kursus yang paling banyak terjual dan tipe pembayaran yang paling banyak digunakan.  |
| 4. | Data Distribusi Usia dan Gender                | Melakukan analisis usia dan gender dari pengguna untuk mengetahui <i>range</i> usia dan gender terbanyak pengguna website.  |
| 5. | Data Top Rating Kursus                         | Melakukan analisis kursus untuk mengetahui kursus apa saja yang memiliki rating teratas.  |
| 6. | Data Top Pendapatan <i>Affiliate Marketing</i> | Melakukan analisis <i>affiliate marketing</i> yang mendapatkan total komisi terbesar berdasarkan penyebaran kode referral kursus.   |
| 7. | Data Trend Akses Website                       | Melakukan analisis penggunaan website untuk mengetahui trend akses website terbanyak di setiap jamnya.  |

### 3.3.2.2 Fase Pemahaman Data (*Data Understanding Phase*)

Tahap pemahaman data ini dimulai dengan melakukan pengumpulan data yang dibutuhkan, mengidentifikasi kualitas data yang digunakan, mencari

informasi tersembunyi yang dapat membentuk hipotesis baru. Pemahaman data dilakukan pada database perusahaan, dimana pada database tersebut terdapat 36 tabel, namun hanya 8 tabel yang akan digunakan yaitu tabel users, tabel users\_detail, tabel course, tabel category, tabel payment, tabel payment\_balance, tabel rating dan tabel ci\_session. Berikut tabel dalam database yang akan digunakan dalam penelitian :

Tabel 3.3 Tabel-tabel Database

| No | Nama Tabel      | Record             | Keterangan  |
|----|-----------------|--------------------|---|
| 1. | Users           | 4149 <i>record</i> | Tabel yang menyimpan informasi seluruh pengguna website yang telah melakukan registrasi |
| 2. | Users_detail    | 641 <i>record</i>  | Tabel yang menyimpan informasi pengguna website lebih detail                            |
| 3. | Course          | 124 <i>record</i>  | Tabel yang menyimpan informasi kursus yang ada pada website                             |
| 4. | Category        | 154 <i>record</i>  | Tabel yang menyimpan informasi kategori kursus yang ada pada website                    |
| 5. | Payment         | 7574 <i>record</i> | Tabel yang menyimpan informasi detail terkait transaksi yang dilakukan oleh user        |
| 6. | Payment_balance | 2395 <i>record</i> | Tabel yang menyimpan informasi detail terkait komisi penjualan dan penarikan saldo      |

|    |             |            |  |
|----|-------------|------------|--|
| 7. | Rating      | 205 record | Tabel ini menyimpan informasi detail terkait rating kursus       |
| 8. | Ci_sessions | 574 record | Tabel ini menyimpan informasi detail terkait waktu akses website |

Berikut *record* dari setiap tabel yang digunakan :

| id  | first_name | last_name  | email        | no_hp      | password | date_added | role_id | last_modified | an_bank | verified | status |
|-----|------------|------------|--------------|------------|----------|------------|---------|---------------|---------|----------|--------|
| 161 | steven     | wijaya     | stevenw...   | 0..e. {... | NUL      | [.. 2 1..  | NUL     | [.. 0         | fc...   | 0        | 1.     |
| 162 | Andy       | Senjaya    | andysen...   | a. {...    | NUL      | [.. 2 1..  | NUL     | [.. 0         | e...    | 1        | 1.     |
| 163 | Hendra...  |            | hendras...   | nu {...    | NUL      | [.. 2 1..  | NUL     | [.. 0         | NUL     | 0        | 1.     |
| 164 | Amel       | Lia        | sakuran...   | 0..9. {... | NUL      | [.. 2 1..  | NUL     | [.. 0         | 3...    | 0        | 1.     |
| 166 | Zul        | Pikri      | fikri7310... | 0..f. {... | NUL      | [.. 2 1..  | NUL     | [.. 0         | 5...    | 0        | 1.     |
| 167 | Mahdi      | Farhan     | mahdifar...  | 0..b. {... | NUL      | [.. 2 1..  | NUL     | [.. 0         | f3...   | 0        | 1.     |
| 168 | melisa     | novitasari | novitasa...  | 0..6. {... | NUL      | [.. 2 1..  | NUL     | [.. 0         | e...    | 0        | 1.     |
| 170 | Wahyu      | Prasetyo   | sellenak...  | b. {... <. | NUL      | [.. 2 1..  | 1...    | [.. 0         | a...    | 0        | 1.     |
| 171 | Ryan       | Tryana     | ryantrya...  | e. {...    | NUL      | [.. 2 1..  | NUL     | [.. 0         | dc...   | 0        | 1.     |
| 172 | Milda P... |            | mildapur...  | nu {...    | NUL      | [.. 2 1..  | NUL     | [.. 0         | NUL     | 0        | 1.     |
| 173 | Dian       | Putra      | dianputr...  | 5. {...    | NUL      | [.. 2 1..  | NUL     | [.. 0         | 9...    | 0        | 1.     |

Gambar 3.4 *Record* Tabel Users

Berdasarkan Gambar 3.4 *Record* Tabel Users menampilkan sebagian *record* pada tabel users, tabel users terdiri atas 26 atribut diantaranya adalah id, first\_name, last\_name, email, no\_hp, password, date\_added, role\_id, last\_modified, an\_bank.

| id | user_id | address       | usia | birthdat | jr | ref    | status |
|----|---------|---------------|------|----------|----|--------|--------|
| 1  | 349     | Jlan Yab...   | 27   | 0000...  | 1  | T... 3 |        |
| 2  | 350     | Stasiun ...   | 46   | 0000...  | 1  | T... 3 |        |
| 3  | 351     | Komp. M...    | 25   | 0000...  | 1  | T... 3 |        |
| 4  | 352     | Jl adisuci... | 26   | 0000...  | 1  | T... 3 |        |
| 5  | 353     | Kota Sor...   | 28   | 0000...  | 0  | T... 3 |        |
| 6  | 355     | Kantor B...   | 26   | 0000...  | 1  | R... 3 |        |
| 7  | 356     | Jl. Hara...   | 23   | 0000...  | 0  | R... 3 |        |
| 8  | 357     | Jalan TA...   | 26   | 0000...  | 1  | T... 3 |        |

Gambar 3.5 *Record* Tabel Users\_detail

Berdasarkan Gambar 3.5 *Record* Tabel Users\_detail menampilkan sebagian *record* pada tabel users\_detail, tabel users\_detail terdiri atas 8 atribut diantaranya adalah id, user\_id, address, usia, dan jns\_kel.

| id | title         | short_description          | description | b | ou        | lan | cat | sut         | se   | requi | pri | dis  | dis  | lev | use |
|----|---------------|----------------------------|-------------|---|-----------|-----|-----|-------------|------|-------|-----|------|------|-----|-----|
| 65 | Training i... | Membuat sebuah Webs...     | <p>Bah...   | 0 | [... b... | 15  | 33  | [... ["...  | 4... | 1     | 2.. | b... | 2... |     |     |
| 70 | Android ...   | Membuat Aplikasi Andr...   | <p><sp...   | 0 | [... b... | 15  | 35  | [... ["...  | 7... | 1     | 3.. | b... | 2... |     |     |
| 71 | Web Des...    | Web Design - Membuat...    | <p>Belaj... | 0 | [... b... | 15  | 33  | [... ["...  | 3... | 1     | 1.. | b... | 2... |     |     |
| 72 | Intensiv...   | Membuat Website Siste...   | <p xss=r... | 0 | [... b... | 15  | 33  | [... [L...  | 6... | 1     | 3.. | i... | 2... |     |     |
| 77 | Training i... | Membuat aplikasi andro...  | <p xss=r... | 0 | [... b... | 15  | 35  | [... ["...  | 4... | NULL  | 0   | b... | 2... |     |     |
| 83 | Kursus O...   | Di sini kamu tidak hany... | <p><sp...   | 0 | [... b... | 47  | 85  | [... ["]... | 3... | 1     | 1.. | b... | 2... |     |     |

Gambar 3.6 *Record* Tabel Course

Berdasarkan Gambar 3.6 *Record* Tabel Course menampilkan sebagian *record* pada tabel course, tabel course terdiri atas 33 atribut diantaranya adalah id, title, description, language, level, date\_added, thumbnail, section, user\_id, dan status.

| id | cod  | name | parent | slug | seo_t    | date_add | last_modi | font_a                 | thumbnail |
|----|------|------|--------|------|----------|----------|-----------|------------------------|-----------|
| 2  | NULL | NULL | NULL   | n-a  | 15767... | NULL     | fas f...  | category-thumbnail.png |           |
| 4  | NULL | NULL | NULL   | n-a  | 15767... | NULL     | fas f...  | category-thumbnail.png |           |
| 6  | NULL | NULL | NULL   | n-a  | 15767... | NULL     | fas f...  | category-thumbnail.png |           |
| 8  | NULL | NULL | NULL   | n-a  | 15767... | NULL     | fas f...  | category-thumbnail.png |           |
| 10 | NULL | NULL | NULL   | n-a  | 15767... | NULL     | fas f...  | category-thumbnail.png |           |
| 11 | 4... | CMS  | 1      | cms  | 15767... | NULL     | NULL      | NULL                   |           |
| 12 | NULL | NULL | NULL   | n-a  | 15767... | NULL     | fas f...  | category-thumbnail.png |           |

Gambar 3.7 *Record* Tabel Category

Berdasarkan Gambar 3.7 *Record* Tabel Category menampilkan sebagian *record* pada tabel category, tabel category terdiri atas 10 atribut diantaranya adalah id, name, date\_added, last\_modified dan thumbnail.

| id  | id_user | payment_type  | pay  | pay  | coupi | cour | amc  | dat  | last_ | admi | ins  | insti | stat | status |
|-----|---------|---------------|------|------|-------|------|------|------|-------|------|------|-------|------|--------|
| 45  | 5... 5  | Free          |      |      |       | 1    | 0    | 1... | NULL  | 0    | 0    | 1     | 0    | 1      |
| 54  | 7... 7  | Free          |      |      |       | 1    | 0    | 1... | NULL  | 0    | 0    | 1     | 0    | 1      |
| 94  | 6... 6  | bank transfer | b... | 8... |       | 10   | 9... | 1... | NULL  | 9... | 0    | 1     | 1    | 1      |
| 95  | 5... 5  | bank transfer | b... | 8... |       | 2    | 1... | 1... | NULL  | 1... | 0    | 1     | 1    | 1      |
| 141 | 9... 9  | Waiting       |      |      |       | 12   | 2... | 1... | NULL  | 6... | 1... | 0     | -5   | 0      |
| 142 | 6... 6  | bank transfer | bni  | 9... |       | 9    | 2... | 1... | NULL  | 2... | 0    | 1     | -2   | 1      |
| 143 | 6... 6  | Free          |      |      |       | 1    | 0    | 1... | NULL  | 0    | 0    | 1     | 0    | 1      |

Gambar 3.8 Record Tabel Payment

Berdasarkan Gambar 3.8 Record Tabel Payment menampilkan sebagian record pada tabel payment, tabel payment terdiri atas 17 atribut diantaranya adalah id, id\_payment, user-id, date\_added, last\_modified, status, dan amount.

| id_j | id_u | pt | pb_payment       | pb   | pb_ | pb_ | pb_ | pb_j | pb_  | pb_  | pb_  | pb_date | pb_d    | pb_token |
|------|------|----|------------------|------|-----|-----|-----|------|------|------|------|---------|---------|----------|
| 1    | 21   |    | Penjualan kursus | 1... | 1   | 0   | 0   | 0    |      | 1    | 1... | 2020... | 00...   |          |
| 3    | 21   |    | Penarikan Saldo  | 5... | 2   | 0   | 3   | 1... | b... | 2    | 9... | 2020... | 00...   |          |
| 4    | 21   |    | Penarikan Saldo  | 5... | 2   | 0   | 3   | 1... | b... | 2    | 4... | 2020... | 20...   |          |
| 5    | 21   |    | Penjualan kursus | 1... | 1   | 0   |     | NULL | NULL | NULL | 1    | 1...    | 2020... | NULL     |
| 6    | 17   |    | Penjualan kursus | 1... | 1   | 0   |     | NULL | NULL | NULL | 1    | 1...    | 2020... | NULL     |
| 7    | 21   |    | Penarikan Saldo  | 5... | 2   | 0   | 3   | 1... | b... | 2    | 1... | 2020... | 20...   |          |
| 8    | 22   |    | Penjualan kursus | 2... | 1   | 0   |     | NULL | NULL | NULL | 1    | 2...    | 2020... | NULL     |

Gambar 3.9 Record Tabel Payment\_balance

Berdasarkan Gambar 3.9 Record Tabel Payment\_balance menampilkan sebagian record pada tabel payment\_balance, tabel payment\_balance terdiri atas 15 atribut diantaranya id\_pb, pb\_bank, pb\_status, pb\_affiliate, pb\_date, dan pb\_norek.

| id | rating | user_id | ratat | rate | date_added | last_modified | review |
|----|--------|---------|-------|------|------------|---------------|--------|
| 1  | 4      | 6       | 1     | c... | 1580688000 | NULL          |        |
| 2  | 5      | 1       | 7     | c... | 1581033600 | NULL          | Cocok  |
| 3  | NULL   | NULL    | NULL  | c... | 1637193600 | NULL          | NULL   |
| 4  | 5      | 6       | 13    | c... | 1585267200 | NULL          |        |
| 5  | 5      | 0       | 21    | c... | 1585267200 | NULL          |        |
| 6  | 5      | 1       | 17    | c... | 1585612800 | NULL          |        |

Gambar 3. 10 Record Tabel Rating

Berdasarkan Gambar 3.10 *Record* Tabel Rating menampilkan sebagian record pada tabel rating, tabel rating terdiri atas 8 atribut diantaranya adalah id, rating, user\_id, date\_added, last\_modified dan review.

| id            | ip_addi | timestamp | data |
|---------------|---------|-----------|------|
| 7eaab5e29...  | 178.... | 15876...  | BLOB |
| 045cd6008...  | 178.... | 15876...  | BLOB |
| 13f024760...  | 202.... | 15876...  | BLOB |
| 18dbfef77f... | 202.... | 15876...  | BLOB |
| a30612f9c...  | 178.... | 15876...  | BLOB |
| 3048631ea...  | 202.... | 15876...  | BLOB |
| 1f699ad7a...  | 178.... | 15876...  | BLOB |
| 13838e42e...  | 36.9... | 15876...  | BLOB |
| 46a479f1f...  | 64.2... | 15876...  | BLOB |

Gambar 3. 11 Record Tabel Ci\_sessions

Berdasarkan Gambar 3.11 Record Tabel Ci\_sessions menampilkan sebagian record pada tabel ci\_sessions, tabel ci\_sessions terdiri atas 4 atribut yaitu id, ip\_address, timestamp, dan data.

### 3.3.2.3 Fase Persiapan Data (*Data Preparation Phase*)

Tahap persiapan data ini melakukan seleksi data yang digunakan, mengintegrasikan data yang telah diseleksi, membersihkan data-data yang tidak digunakan, dan memformat data agar dapat diproses menggunakan teknik data mining. Pada fase persiapan ini menggunakan bahasa pemrograman SQL dan Python. Bahasa pemrograman SQL digunakan untuk mengintegrasikan tabel dan atribut dari database yang akan dibutuhkan dalam melakukan analisis. Sedangkan bahasa pemrograman Python dengan library Pandas digunakan untuk melakukan pembersihan data seperti menghilangkan data yang tidak diinginkan, mengolah dan

menganalisa data-data terstruktur. Berikut tabel dan field yang telah diseleksi dan akan digunakan dalam penelitian :

Tabel 3.4 Tabel dan Field yang digunakan

| No | Nama Tabel      | Field   |
|----|-----------------|---|
| 1. | Users           | id, first_name, last_name, email, ho_hp, date_added                                   |
| 2. | Users_detail    | id, jns_kel, usia   |
| 3. | Course          | id, title   |
| 4. | Category        | id, name  |
| 5. | Payment         | id, user_id, amount, date_added, status_payment, status, status_payment, payment_type |
| 6. | Payment_balance | id_users, pb_payment, pb_nominal  |
| 7. | Rating          | rating, review, retable_id, user_id   |
| 8. | Ci_sessions     | timestamp   |

#### 3.3.2.4 Fase Pemodelan (Modelling Phase)

Tahap pemodelan ini merupakan fase yang secara langsung melibatkan teknik data mining yaitu dengan memilih teknik data mining dan menentukan algoritma apa yang akan digunakan. Dalam fase ini menggunakan algoritma *K-Means* berdasarkan model RFM (*Recency*, *Frequency* dan *Monetary*) untuk mengklusterkan konsumen yang telah bergabung dengan perusahaan. Pada penelitian ini menggunakan *software* RapidMiner 9.9 untuk mengolah dan mengklusterkan data tersebut. Selain itu, pada penelitian ini juga akan dilakukan perhitungan menggunakan Microsoft Excel dan membandingkan kedua nilai dari

nilai *centroid* yang dihasilkan RapidMiner dengan perhitungan manual menggunakan Microsoft Excel.

### **3.3.2.5 Fase Evaluasi (*Evaluation*)**

Tahap evaluasi merupakan tahap lanjutan yang dilakukan untuk mengevaluasi tujuan data mining secara mendalam sehingga menghasilkan pemodelan yang sesuai dengan yang diinginkan. Beberapa hal yang dilakukan dalam fase ini yaitu, mengevaluasi hasil sejauh mana pemodelan yang dibuat telah memenuhi tujuan yang diinginkan, melakukan proses *review* atau pengecekan ulang dengan memastikan seluruh tahapan yang dilakukan tidak terlewatkan, dan menentukan langkah yang dilakukan selanjutnya yaitu melanjutkan ke tahap *deployment* atau kembali ke tahap awal yaitu *business understanding*. Pada tahapan fase evaluasi, hasil *cluster* pada tahapan pemodelan dilakukan evaluasi menggunakan *Davies Bouldin Index* untuk mendapatkan jumlah cluster yang optimal.

### **3.3.2.6 Fase Penyebaran (*Deployment*)**

Pada tahap penyebaran ini, informasi dan pengetahuan yang telah didapatkan kemudian direpresentasikan menjadi bentuk yang lebih mudah dipahami oleh orang awam. Tahapan ini juga menerapkan Google Data Studio untuk memvisualisasikan hasil yang didapat. Dengan penerapan Google Data Studio akan memberikan kemudahan dalam membaca data, menyebarkan data, berkolaborasi bahkan memungkinkan dalam mengekspor data.

Hasil penelitian ini berupa *clustering* konsumen, data potensial konsumen, data kategori, kursus, tipe pembayaran, data *gender* dan usia, data rating kursus teratas, data pendapatan *affiliate marketing* terbesar dan data trend akses website serta memberikan rekomendasi strategi pemasaran yang dapat dilakukan perusahaan berdasarkan data yang dihasilkan agar meningkatkan pendapatan perusahaan.

## 5. KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Berdasarkan hasil penelitian yang dilakukan, maka didapat beberapa kesimpulan :

1. Berdasarkan penelitian yang dilakukan menggunakan *software* RapidMiner dengan metode K-Means dan model RFM dapat membantu mengetahui segmentasi konsumen yang dimiliki dan menghasilkan bahwa 3 *cluster* memiliki nilai 0.197 menggunakan uji validitas *Davies Bouldin Index*. Dimana *cluster low priority* berjumlah 1090 konsumen, *cluster medium priority* 73 konsumen dan *cluster high priority* 969 konsumen.
2. Karakteristik yang menjadi pertimbangan yaitu *Recency*, *Frequency*, dan *Monetary* dapat mendeskripsikan karakteristik yang ada pada setiap *cluster*. Sehingga berdasarkan karakteristik berbeda yang dimiliki setiap *cluster* dapat menghasilkan strategi pemasaran untuk meningkatkan pendapatan perusahaan.
3. Pemanfaatan pengetahuan data lainnya yaitu data potensial konsumen, *gender* dan usia, rating kursus terbaik, pendapatan *affiliate marketing* terbesar dan *trend access* website juga dapat dimanfaatkan dalam menentukan strategi pemasaran perusahaan dan keputusan bisnis, diantaranya pembuatan kategori

kursus yang paling banyak terjual, pemberian bonus tambahan dan E-Certificate untuk *affiliate marketing* yang memiliki pendapatan terbesar.

4. Hasil analisis yang divisualisasikan dengan *column chart*, *donut chart*, *time series chart*, tabel dan *control drop down list* menggunakan penerapan Google Data Studio dapat mempermudah pengguna dalam membaca grafik, mengekspor data, dan proses penyebaran data.

## 5.2 Saran

Adapun saran dari penelitian ini adalah :

1. Diharapkan pada penelitian selanjutnya menggunakan *software* selain RapidMiner dan menggunakan metode *clustering* yang lain agar dapat membandingkan dan memberikan hasil yang lebih baik.
2. Data perusahaan sebaiknya lebih diperkaya dengan mewajibkan user dalam menginputkan *gender* dan tanggal lahir ketika mendaftar sehingga mendapatkan jumlah yang sesuai maka akan menghasilkan pengetahuan data yang maksimal.

## DAFTAR PUSTAKA

- [1] B. R. T. Putri, *Manajemen Pemasaran*, Denpasar: Universitas Udayana, 2017.
- [2] H. M. Ritonga, M. E. Fikri, N. Siregar, R. R. Agustin and R. Hidayat, *Manajemen Pemasaran Konsep dan Strategi*, Medan: Universitas Pembangunan Panca Budi, 2018.
- [3] S. Z. ZA, *Manajemen Pemasaran Teori & Strategi*, Samarinda: Forum Pemuda Aswaja, 2020.
- [4] O. F. Sitorus and N. Utami, *Strategi Promosi Pemasaran*, Jakarta: Universitas Muhammadiyah Prof. Dr. HAMKA, 2017.
- [5] H. Wijoyo, A. Ariyanto, F. Wongso, I. Indrawan, M. M. F. Akbar, N. Anggraini, S. S. and W. S. , *Strategi Pemasaran UMKM di Masa Pandemi*, Solok: Insan Cendekia Mandiri, 2021.
- [6] H. Kurniawan, S. Defit and S. , "Data Mining Menggunakan Metode K-Means Clustering untuk Menentukan Besaran Uang Kuliah Tunggal," *Journal of Applied Computer Science and Technology*, vol. 1, no. 2, pp. 80 - 89, 2020.
- [7] Noviyanto, "Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia," *Paradigma - Jurnal Informatika dan Komputer*, vol. 22, no. 2, pp. 183 - 188, 2020.
- [8] D. A. C, D. A. Baskoro, L. Ambarwati and I. W. Simri, *Belajar Data Mining dengan RapidMiner*, Jakarta, 2013.
- [9] I. A. Nikmatun and I. Waspada, "Implementasi Data Mining Untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Simetris*, vol. 10, no. 2, pp. 421 - 432, 2019.

- [10] B. E. Adiana, I. Soesanti and A. E. Permanasari, "Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering," *JUTEI*, vol. 2, no. 1, pp. 23 - 32, 2018.
- [11] A. N. Khomarudin, *Teknik Data Mining : Algoritma K-Means Clustering*, Komunitas eLearning Ilmu Komputer, 2016.
- [12] F. Liantoni and L. Cahyani, "Pemanfaatan Hierarchical Clustering untuk Pengelompokan Daun Berdasarkan Fitur Moment Invariant," *Jurnal Ilmiah Edutic*, vol. 3, no. 2, pp. 91 - 98, 2017.
- [13] Y. Agusta, "K-Means - Penerapan, Permasalahan dan Metode Terkait," *Jurnal Sistem dan Informatika*, vol. 3, pp. 47-60, 2007.
- [14] M. Wahyudi, M. R. Saragih and S. , *Data Mining : Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering*, Yayasan Kita Menulils, 2020.
- [15] A. Bastian, H. Sujadi and G. Febrianto, "Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)," *Jurnal Sistem Informasi*, vol. 14, no. 1, pp. 26-32, 2018.
- [16] Lathifaturrahmah, "Perbandingan Hasil Penggerombolan K-Means, Fuzzy K-Means, dan Two Step Clustering," *JPM IAIN Antasari*, vol. 2, no. 1, pp. 39-62, 2014.
- [17] U. D. Arni, "Kelebihan dan Kekurangan K-Means Clustering," *Garuda Cyber Indonesia*, 07 01 2019. [Online]. Available: <https://news.garudacyber.co.id/>. [Accessed 31 12 2021].
- [18] V. Kumar, *Managing Customers for Profit : Strategies To Increase Profits And Build Loyalty*, 2008: Pearson Education, India.
- [19] D. L. Olson and G. Lauhoff, *Descriptive Data Mining*, Germany: Springer Singapore, 2019.
- [20] R. C. Blattberg, B.-D. Kim and S. A. Neslin, *Database Marketing: Analyzing and Managing Customers*, Netherlands: Springer New York, 2008.
- [21] J.-T. Wei, S.-Y. Lin and H.-H. Wu, "A Review of The Application of RFM Model," *African Journal of Business Management*, vol. 4, no. 19, pp. 4199 - 4026, 2010.
- [22] RapidMiner, *Data Mining Use Cases and Business Analytics Applications*, United Kingdom: Taylor & Francis, 2013.
- [23] B. Santoso, A. I. S. Azis and Z. , *Machine Learning & Reasoning Fuzzy Logic Algoritma, Manual, Matlab, & Rapid Miner*, Penerbit Deepublish, 2020.

- [24] S. Dua and P. Chowriappa, *Data Mining for Bioinformatics*, United Kingdom: Taylor & Francis, 2013.
- [25] A. A. Az-zahra, A. F. Marsaoly, I. P. Lestyani, R. Salsabila and W. O. Z. Madjida, "Penerapan Algoritma K-Modes Clustering dengan Validasi Davies Bouldin Index pada Pengelompokan Tingkat Minat Belanja Online di Provinsi Daerah Istimewa Yogyakarta," *Jurnal Matematika dan Statistika serta Aplikasinya*, vol. 9, no. 1, pp. 24 -36, 2021.
- [26] I. D. Id, T. R. Angraini and Sukamto, "Penentuan Daerah Rawan Titik Api di Provinsi Riau Menggunakan Clustering Algoritma K-Means," *JUITA*, vol. 7, no. 2, pp. 137 - 148, 2018.
- [27] S. Butsianto and N. Saepudin, "Penerapan Data Mining Terhadap Minat Siswa Dalam Mata Pelajaran Matematika Dengan Metode K-Means," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 3, no. 1, pp. 51-59, 2020.
- [28] S. Hendrian, "Algoritma Klasifikasi Data Mining untuk Memprediksi Siswa dalam Memperoleh Bantuan Dana Pendidikan," *Faktor Exacta*, vol. 11, no. 3, pp. 266 - 274, 2018.
- [29] Z. Muhammad, R. Rahmadhani, H. Rizqifaluthi and M. A. Yaqin, "Process Mining Akademik Sekolah Menggunakan RapidMiner," *Matics : Jurnal Ilmu Komputer dan Teknologi Informasi*, vol. 10, no. 2, pp. 47-51, 2018.
- [30] W. R. Paczkowski, *Deep Data Analytics for New Product Development*, United Kingdom: Taylor & Francis, 2020.
- [31] G. Snipes, "Google Data Studio," *Journal of Librarianship and Scholarly Communication*, vol. 6, no. 1, 2018.
- [32] L. Hurst, *Hands On With Google Data Studio: A Data Citizen's Survival Guide*, United States: Wiley, 2020.
- [33] Y. Kurniawan, *Cara Mudah Menguasai Google Docs*, Jakarta: Elex Media Komputindo, 2010.
- [34] C. Rafter, *Guide To Cleaning Data with Excel and Ggoogle Sheets*, DSM Media, 2019.
- [35] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer and R. Wirth, *CRISP-DM 1.0, The CRISP-DM consortium*, 2020.
- [36] S. Ssonang, A. T. Purba and F. O. I. Pardede, "Pengelompokan Jumlah Penduduk Berdasarkan Kategori Usia dengan Metode K-Means," *Tekinkom*, vol. 2, no. 2, pp. 166 - 172, 2019.

- [37] W. A. Taqvim, N. Y. Setiawan and F. A. Bachtiar, "Analisis Segmentasi Pelanggan Dengan RFM Model Pada Pt. Arthamas," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 2, pp. 1986-1993, 2019.
- [38] Y. H. Chrisnanto and A. Kanianingsih, "Pengelompokan Ekuitas Pelanggan Berbasis *Recency Frequency Monetary* (RFM) Menggunakan K-Means Clustering," *Seminar Nasional Teknologi Informasi dan Komunikasi*, pp. 45-53, 2019.
- [39] C. D. Rumiarti and I. Budi, "Segmentasi Pelanggan pada Customer Relationship Management di Perusahaan Ritel : Studi Kasus PT Gramedia Asri Media," *Jurnal Sistem Informasi*, vol. 13, no. 1, pp. 1-10, 2017.
- [40] Setiawan, H. Amani and W. Tripiawan, "Perancangan Segmentasai Pelanggan dengan Metode Clustering K-Means dan Model RFM pada Klinik Kecantikan Seoul Secret," *E-Proceeding of Engineering*, vol. 8, no. 2, pp. 2286 - 2293, 2021.
- [41] F. M. Aditya, "Perancangan Dashboard Sebagai Pendukung Keputusan Berdasarkan Data Antrean Pada Aplikasi Antribos (Studi Kasus : Medical Center ITS)," Institut Teknologi Sepuluh November, Surabaya, 2018.
- [42] D. Fernando, "Visualisasi Data Menggunakan Google Data Studio," *SNARTISI*, pp. 71-77, 2018.
- [43] P. Anitha and M. M. Patil, "RFM Model for Customer Purchase Behavior using K-Means Algorithm," *Jurnal of King Saud University*, pp. 1 - 8, 2019.
- [44] K. Archana and K. G. Saranya, "Mall Customer Segmentation Using Clustering Algorithm," *International Journal of Multidisciplinary Educational Research*, vol. 8, no. 12(6), pp. 94 - 99, 2019.
- [45] O. Dogan, Z. A. Bulut and E. Aycin, "Customer Segmentation by Using RFM Model and Clustering Methods : A Case Study in Retail Industry," *International Journal of Contemporary Economics and Administrative Sciences*, vol. 8, no. 1, pp. 1-19, 2018.