

**ANALISIS SEGMENTASI PELANGGAN MENGGUNAKAN  
METODE *K-MEANS CLUSTERING* DENGAN MODEL  
*RECENCY, FREQUENCY, MONETARY* (RFM)  
(STUDI KASUS: CV. GARDEN CELL KEDIRI)**

**SKRIPSI**



**Twenty Oktavia Pujiliana**

**PROGRAM STUDI S1 SISTEM INFORMASI  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS AIRLANGGA**

**2022**

**ANALISIS SEGMENTASI PELANGGAN MENGGUNAKAN  
METODE *K-MEANS CLUSTERING* DENGAN MODEL  
*RECENCY, FREQUENCY, MONETARY* (RFM)  
(STUDI KASUS: CV. GARDEN CELL KEDIRI)**

**SKRIPSI**



**Twenty Oktavia Pujiliana**

**NIM: 081811633024**

**PROGRAM STUDI S1 SISTEM INFORMASI  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS AIRLANGGA**

**2022**

**ANALISIS SEGMENTASI PELANGGAN MENGGUNAKAN METODE K-  
MEANS CLUSTERING DENGAN MODEL *RECENCY, FREQUENCY,*  
*MONETARY* (RFM)**

**(STUDI KASUS: CV. GARDEN CELL KEDIRI)**

**SKRIPSI**

**Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Komputer**

**Bidang Sistem Informasi Pada Fakultas Sains dan Teknologi**

**Universitas Airlangga**

**Oleh:**

**(Twenty Oktavia Puiliana)**

**NIM: 081811633024**

**Tanggal Lulus:**

**22 Agustus 2022**

**Disetujui Oleh:**

**Pembimbing I,**



**Purbandini, S.Si., M.Kom.**

**NIP. 197107122008122001**

**Pembimbing II,**



**Endah Purwanti, S.Si., M.Kom**

**NIP. 197812172005012001**

## LEMBAR PENGESAHAN SKRIPSI

**Judul** : ANALISIS SEGMENTASI PELANGGAN  
MENGUNAKAN METODE *K-MEANS*  
*CLUSTERING* DENGAN MODEL *RECENCY*,  
*FREQUENCY*, *MONETARY* (RFM) (STUDI  
KASUS: CV. GARDEN CELL KEDIRI)

**Penyusun** : Twenty Oktavia Pujiliana

**NIM** : 081811633024

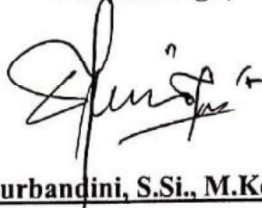
**Pembimbing I** : Purbandini, S.Si., M.Kom.

**Pembimbing II** : Endah Purwanti, S.Si., M.Kom.

**Tanggal Seminar** : 22 Agustus 2022

Disetujui oleh:

Pembimbing I,



Purbandini, S.Si., M.Kom.

NIP. 197107122008122001

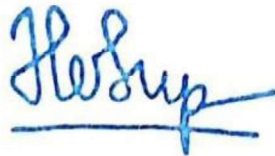
Pembimbing II,



Endah Purwanti, S.Si., M.Kom

NIP. 197812172005012001

Ketua Departemen Matematika  
Fakultas Sains dan Teknologi  
Universitas Airlangga



Dr. Herry Suprajitno, S.Si., M.Si.

NIP. 196804041994031020

Koordinator Program Studi  
S1 Sistem Informasi  
Fakultas Sains dan Teknologi  
Universitas Airlangga



Dr. Rimuljo Hendradi, S.Si., M.Si.

NIP. 197102111997021001

## **SURAT PERNYATAAN TENTANG ORISINALITAS**

Yang bertanda tangan di bawah ini:

Nama : Twenty Oktavia Pujiliana

NIM : 081811633024

Program Studi : S1 Sistem Informasi

Fakultas : Sains dan Teknologi

Jenjang : Sarjana (S1)

Menyatakan bahwa tidak melakukan tindakan plagiarisasi dalam penulisan skripsi yang berjudul:

*ANALISIS SEGMENTASI PELANGGAN MENGGUNAKAN METODE K-MEANS CLUSTERING DENGAN MODEL RECENCY, FREQUENCY, MONETARY (RFM) (STUDI KASUS: CV. GARDEN CELL KEDIRI).*

Apabila suatu saat terbukti melakukan tindakan plagiarisasi, maka saya akan menerima sanksi yang telah ditetapkan.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya.

Surabaya, Agustus 2022



Twenty Oktavia Pujiliana

Twenty Oktavia Pujiliana, 2022. Analisis Segmentasi Pelanggan Menggunakan Metode *K-Means Clustering* dengan Model *Recency, Frequency, Monetary* (RFM) (Studi Kasus: CV. Garden Cell Kediri). Skripsi ini di bawah bimbingan Purbandini, S.Si., M.Kom. dan Endah Purwanti, S.Si., M.Kom. Program Studi S1 Sistem Informasi. Fakultas Sains dan Teknologi, Universitas Airlangga.

---

## ABSTRAK

*Customer Relationship Management* (CRM) menjadi salah satu komponen penting dalam membangun perusahaan. Karena dengan CRM, perusahaan dapat mengelola hubungan dengan pelanggan secara lebih terstruktur dan dapat mengikat pelanggan yang berpotensi memberikan keuntungan bagi kelangsungan bisnis. Di CV. Garden Cell Kediri, penerapan CRM hanya terfokus pada pemberian diskon, promo, dan bonus kepada pelanggan tanpa membedakan karakteristiknya. Sehingga perusahaan masih kesulitan untuk membedakan dan mengikat pelanggan potensial.

Oleh karena itu, penelitian ini berusaha membentuk segmentasi pelanggan di CV. Garden Cell Kediri dengan teknik *K-Means clustering* dan model RFM. Tahapan penelitian meliputi *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. *Business understanding* dilakukan untuk memahami dan mencari tahu hal-hal yang dibutuhkan dalam penelitian mengenai perusahaan dengan menggunakan metode wawancara. *Data understanding* bertujuan untuk memahami data yang diperoleh dari hasil wawancara dan mendefinisikan data sesuai dengan kebutuhan penelitian. *Data preparation* bertujuan untuk mempersiapkan data sebelum proses segmentasi dilakukan. Dalam tahap ini dilakukan pembersihan data dan pemilihan atribut yang berguna untuk proses utama. Kemudian dilakukan proses pembobotan RFM dengan *recency* merupakan nilai keterbaruan pelanggan melakukan transaksi, *frequency* adalah jumlah transaksi pelanggan, dan *monetary* adalah total biaya yang dikeluarkan pelanggan. Setelah pembobotan RFM berhasil dilakukan, dilanjutkan dengan normalisasi dan pengelompokan pelanggan berdasarkan nilai RFM. *Modeling* dilakukan untuk mengelompokkan pelanggan menggunakan metode *K-Means clustering*. Hasil akhir *clustering* menunjukkan bahwa data terbagi menjadi 4 *cluster*, yaitu C1, C2, C3, dan C4. *Evaluation* bertujuan untuk menguji seberapa baik hasil *clustering* dengan menggunakan metode *Silhouette Score* dan *Davies-Bouldin Index*. Sedangkan, *Deployment* digunakan untuk membuat pemprofilan perilaku pelanggan berdasarkan pembagian *cluster* dan membaginya dalam beberapa kelompok, yaitu *Most Valuable Customer* (MVC), *Most Growable Customer* (MGC), *Below Zero* (BZ), dan *Migrator*. Dari pembagian kelompok tersebut diperoleh persentase MVC sebesar 25,3%, MGC sebesar 26,32%, BZ sebesar 22,45%, dan *migrator* sebesar 25,93%. Hal tersebut menandakan lebih dari setengah pelanggan yang melakukan transaksi di tahun 2021 adalah pelanggan bernilai tinggi (MVC dan MGC). Oleh karena itu, perusahaan perlu memberikan perhatian lebih kepada kelompok pelanggan tersebut agar dapat mengikat pelanggan dan memperoleh keuntungan maksimal.

**Kata Kunci** : Segmentasi Pelanggan, Data Mining, Analisis RFM, Clustering K-Means, Perlakuan Kelompok Pelanggan.

*Twenty Oktavia Pujiliana, 2022. Customer Segmentation Analysis Using K-Means Clustering Method with Recency, Frequency, Monetary (RFM) Model (Case Study: CV. Garden Cell Kediri). This undergraduate thesis was under guidance of Purbandini, S.Si., M.Kom. and Endah Purwanti, S.Si., M.Kom. Majors of SI Information System. Faculty of Science and Technology, Airlangga University.*

---

## ABSTRACT

*Customer Relationship Management (CRM) is an important component in building a company. Because with CRM, companies can manage relationships with customers in a more structured manner and can bind customers that have the potential to provide benefits for business continuity. In CV. Garden Cell Kediri, CRM implementation only focuses on giving discounts, promos, and bonuses to customers without distinguishing their characteristics. So the company is still difficult to distinguish and bind potential customers.*

*Therefore, this study seeks to establish customer segmentation in CV. Garden Cell Kediri with K-Means clustering technique and RFM model. The research stages include business understanding, understanding data, data preparation, modeling, evaluation, and deployment. Business understanding is carried out to understand and find out what is needed in research about companies using the interview method. Data understanding aims to understand the data obtained from interviews and define the data according to research needs. Data preparation aims to prepare data before the segmentation process is carried out. In this stage, data cleaning and selection of attributes that are useful for the main process is carried out. Then the RFM weighting process is carried out with recency being the customer's recent transaction value, frequency is the number of customer transactions, and monetary is the total cost incurred by the customer. After the RFM weighting has been successfully carried out, it is continued with normalization and grouping of customers based on the RFM value. Modeling is done to group customers using the K-Means clustering method. The end results of clustering shows that the data can be divided into 4 clusters, which is C1, C2, C3, and C4. Evaluation is done for the purpose of testing how good the end results of clusterization are with both Silhouette Score and Davies-Bouldin Index methods. On the other hand, Deployment is used to profile the customer and group them according to their appropriate cluster division; Most Valuable Customer (MVC), Most Growable Customer (MGC), Below Zero (BZ), and Migrator. After dividing them, we know that the percentage of MVC is 25,3%, MGC is 26,32%, BZ is 22,45%, and migrators are 25,93%. These percentages show that more than half of customers that did their transactions in 2021 are high valued customers (MVC and MGC), based on that, the company needs to give their more of their attention to previously mentioned customer group so that they can retain them and get maximized profits.*

**Keywords** : *Customer Segmentation, Data Mining, RFM Analysis, K-Means Clustering, Customer Group Treatment.*

## KATA PENGANTAR

Puji Syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa yang telah memberikan berkat dan karunia-Nya, sehingga penulis mampu menyelesaikan penyusunan skripsi yang berjudul **“Analisis Segmentasi Pelanggan Menggunakan Metode *K-Means Clustering* dengan Model *Recency, Frequency, Monetary* (RFM) (Studi Kasus: CV. Garden Cell Kediri)”** dengan baik.

Penulisan skripsi ini tidak akan terwujud tanpa bantuan dan dukungan dari berbagai pihak yang telah memberikan waktu, tenaga, dan pikirannya kepada penulis selama proses penyusunan skripsi. Pada kesempatan ini penulis ingin mengucapkan terima kasih kepada beberapa pihak berikut ini:

1. Tuhan Yang Maha Esa yang telah memberikan segala anugrah-Nya berupa kesehatan, kekuatan, kemudahan, dan kesempatan untuk penulis dapat menyelesaikan skripsi tepat waktu.
2. Kedua orang tua dan seluruh keluarga besar Edy Sampoerna yang telah memberikan dukungan penuh berupa doa, semangat, kasih sayang, dan dukungan finansial sejak awal kehidupan penulis, awal perkuliahan, serta awal hingga akhir penulisan skripsi.
3. Purbandini, S.Si., M.Kom selaku dosen pembimbing I yang telah membimbing, memahami, membantu, dan memberikan ilmu selama masa penulisan skripsi dengan kesabaran yang luar biasa.
4. Endah Purwanti, S.Si., M.Kom selaku dosen pembimbing II yang telah membimbing, memahami, membantu, dan memberikan ilmu selama masa penulisan skripsi dengan kesabaran yang luar biasa.

5. Faried Effendy, S.Si., M.Kom selaku dosen wali yang dengan sabar memberikan pengarahan dan membimbing sejak awal hingga akhir perkuliahan.
6. Bapak Samsul selaku pihak CV. Garden Cell Kediri yang telah menjadi jembatan bagi penulis dan manajemen tempat penelitian, serta beberapa pihak yang telah membantu menyediakan informasi dan sumber data bagi penulis.
7. Fery Andrian selaku teman dekat penulis yang selalu memberikan dukungan, semangat, pengertian, dan waktu sebagai teman *sharing* kepada penulis dalam menyelesaikan skripsi.
8. Aini Latifah selaku kakak tingkat dan sahabat yang telah memberikan waktu serta tenaga untuk membantu penulis menyelesaikan masalah yang ditemui dalam penyusunan skripsi.
9. Keluarga besar Sistem Informasi 2018 yang telah memberikan banyak ilmu dan pengalaman selama masa perkuliahan, serta motivasi untuk penulis menyelesaikan skripsi.

Dalam penulisan skripsi ini, penulis menyadari bahwa masih terdapat kekurangan dan ketidaksempurnaan. Oleh karena itu, penulis bersedia menerima secara terbuka segala kritik dan saran yang bersifat membangun untuk kesempurnaan skripsi ini. Penulis berharap agar skripsi ini dapat memberikan manfaat dan pengetahuan yang berguna bagi seluruh akademisi dan praktisi.

Surabaya, Agustus 2022

Penulis

## DAFTAR ISI

|                                                           |            |
|-----------------------------------------------------------|------------|
| <b>LEMBAR PENGESAHAN SKRIPSI .....</b>                    | <b>iii</b> |
| <b>SURAT PERNYATAAN TENTANG ORISINALITAS .....</b>        | <b>iv</b>  |
| <b>ABSTRAK .....</b>                                      | <b>v</b>   |
| <b>ABSTRACT .....</b>                                     | <b>vi</b>  |
| <b>KATA PENGANTAR.....</b>                                | <b>vii</b> |
| <b>DAFTAR ISI.....</b>                                    | <b>ix</b>  |
| <b>DAFTAR GAMBAR.....</b>                                 | <b>xi</b>  |
| <b>DAFTAR TABEL .....</b>                                 | <b>xii</b> |
| <b>BAB 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.....                                  | 5          |
| <b>BAB II TINJAUAN PUSTAKA.....</b>                       | <b>6</b>   |
| 2.1 CV. Garden Cell Kediri.....                           | 6          |
| 2.2 Penelitian Sebelumnya .....                           | 10         |
| 2.3 Segmentasi Pelanggan.....                             | 12         |
| 2.4 <i>Customer Relationship Management (CRM)</i> .....   | 12         |
| 2.5 <i>Data Mining</i> .....                              | 13         |
| 2.6 <i>Model Recency, Frequency, Monetary (RFM)</i> ..... | 16         |
| 2.7 <i>Clustering</i> .....                               | 21         |
| 2.8 <i>K-Means Clustering</i> .....                       | 23         |
| 2.9 <i>Profiling</i> .....                                | 28         |
| <b>BAB III METODOLOGI PENELITIAN .....</b>                | <b>32</b>  |
| 3.1 Tempat dan Waktu Penelitian .....                     | 32         |
| 3.2 Objek Penelitian .....                                | 32         |
| 3.3 Langkah-Langkah Penelitian.....                       | 32         |
| <b>BAB IV HASIL DAN PEMBAHASAN .....</b>                  | <b>42</b>  |
| 4.1 <i>Business Understanding</i> .....                   | 42         |

|                                                                                                                          |                                 |              |
|--------------------------------------------------------------------------------------------------------------------------|---------------------------------|--------------|
| 4.2                                                                                                                      | <i>Data Understanding</i> ..... | 44           |
| 4.3                                                                                                                      | <i>Data Preparation</i> .....   | 45           |
| 4.4                                                                                                                      | <i>Modeling</i> .....           | 55           |
| 4.5                                                                                                                      | <i>Evaluation</i> .....         | 63           |
| 4.6                                                                                                                      | <i>Deployment</i> .....         | 71           |
| <b>BAB V KESIMPULAN DAN SARAN</b> .....                                                                                  |                                 | <b>77</b>    |
| 5.1                                                                                                                      | Kesimpulan.....                 | 77           |
| 5.2                                                                                                                      | Saran .....                     | 78           |
| <b>DAFTAR PUSTAKA</b> .....                                                                                              |                                 | <b>xiii</b>  |
| <b>Lampiran 1. Hasil Wawancara Awal</b> .....                                                                            |                                 | <b>xvi</b>   |
| <b>Lampiran 2. Contoh Data Bobot <i>Recency</i>, <i>Frequency</i>, <i>Monetary</i> Sebelum Transformasi</b> .....        |                                 | <b>xvii</b>  |
| <b>Lampiran 3. Contoh Data Bobot <i>Recency</i>, <i>Frequency</i>, <i>Monetary</i> Setelah Transformasi</b> .....        |                                 | <b>xxii</b>  |
| <b>Lampiran 4. Contoh Data Label <i>Recency</i>, <i>Frequency</i>, <i>Monetary</i> dan Hasil <i>Clustering</i></b> ..... |                                 | <b>xxvii</b> |
| <b>Lampiran 4. <i>Source Code</i> Penelitian dalam Bahasa Python</b> .....                                               |                                 | <b>xxxii</b> |

## DAFTAR GAMBAR

|                                                                               |    |
|-------------------------------------------------------------------------------|----|
| <b>Gambar 2. 1</b> Diagram Proses CRISP-DM (Venkat, 2017).....                | 14 |
| <b>Gambar 3. 1</b> Flowchart Metodologi Penelitian .....                      | 33 |
| <b>Gambar 3. 2</b> Block Diagram Tahapan Pembobotan RFM.....                  | 35 |
| <b>Gambar 3. 3</b> Flowchart Algoritma K-Means (Bustamam, Alhadi, 2017) ..... | 38 |
| <b>Gambar 4. 1</b> Langkah Perhitungan Bobot Recency .....                    | 47 |
| <b>Gambar 4. 2</b> Langkah Perhitungan Bobot Frequency .....                  | 48 |
| <b>Gambar 4. 3</b> Langkah Perhitungan Bobot Monetary.....                    | 49 |
| <b>Gambar 4. 4</b> Nilai k Optimal dengan Metode Elbow .....                  | 57 |
| <b>Gambar 4. 5</b> Plot Hasil Clustering.....                                 | 62 |
| <b>Gambar 4. 6</b> Rata-Rata Centroid di Setiap Cluster dan Atribut RFM ..... | 63 |
| <b>Gambar 4. 7</b> Grafik Nilai Silhouette Score pada Setiap Cluster .....    | 67 |
| <b>Gambar 4. 8</b> Plot Silhouette Score untuk 3 Cluster Tertinggi .....      | 67 |
| <b>Gambar 4. 9</b> Plot Jarak Intercluster pada Setiap Cluster .....          | 68 |
| <b>Gambar 4. 10</b> Grafik Hasil Davies-Bouldin Index .....                   | 70 |

## DAFTAR TABEL

|                                                                                   |    |
|-----------------------------------------------------------------------------------|----|
| <b>Tabel 2. 1</b> Ringkasan Penelitian Sebelumnya .....                           | 10 |
| <b>Tabel 2. 2</b> Kelompok Kategori Pelanggan .....                               | 30 |
| <b>Tabel 4. 1</b> Deskripsi Atribut Awal Data Transaksi .....                     | 44 |
| <b>Tabel 4. 2</b> Jumlah Missing Value di Setiap Atribut .....                    | 45 |
| <b>Tabel 4. 3</b> Atribut Terpilih untuk Proses Pengolahan Data .....             | 46 |
| <b>Tabel 4. 4</b> Hasil Pembobotan Recency, Frequency, Monetary (RFM) .....       | 50 |
| <b>Tabel 4. 5</b> Hasil Transformasi Data Atribut RFM .....                       | 52 |
| <b>Tabel 4. 6</b> Nilai Kuartil Atribut RFM .....                                 | 53 |
| <b>Tabel 4. 7</b> Pelabelan Skor Berdasarkan Bobot RFM .....                      | 54 |
| <b>Tabel 4. 8</b> Kategori Pelabelan RFM .....                                    | 54 |
| <b>Tabel 4. 9</b> Hasil Pengelompokan Data Berdasarkan Kategori .....             | 55 |
| <b>Tabel 4. 10</b> Nilai Sum of Squared Error (SSE) .....                         | 56 |
| <b>Tabel 4. 11</b> Hasil Iterasi Pertama Proses K-Means Clustering .....          | 59 |
| <b>Tabel 4. 12</b> Pusat Centroid untuk Iterasi Kedua .....                       | 61 |
| <b>Tabel 4. 13</b> Hasil Iterasi Kedua Proses K-Means Clustering .....            | 61 |
| <b>Tabel 4. 14</b> Hasil Akhir Proses K-Means Clustering .....                    | 62 |
| <b>Tabel 4. 15</b> Pusat Centroid untuk Iterasi Akhir .....                       | 63 |
| <b>Tabel 4. 16</b> Data untuk Perhitungan Silhouette Score .....                  | 64 |
| <b>Tabel 4. 17</b> Nilai Silhouette Score pada Setiap Cluster .....               | 66 |
| <b>Tabel 4. 18</b> Nilai DBI pada Setiap Cluster .....                            | 70 |
| <b>Tabel 4. 19</b> Pembagian Kelompok Pelanggan Berdasarkan Rata-Rata Cluster ... | 71 |