

## PENGELOMPOKAN LOYALITAS PELANGGAN DENGAN MENGGUNAKAN KOMBINASI RFM DAN ALGORITMA K-MEANS

Rita Hamdani

Teknik Informatika  
STMIK Pelita Nusantara, Jl. Iskandar Muda No. Medan

e-mail: r1t4.hamdani@gmail.com

### Abstrak

Dengan banyaknya perusahaan yang bergerak dalam bidang yang sama, maka tingkat persaingan sangat tinggi, sehingga akan memberikan ancaman bagi perusahaan untuk kehilangan pelanggan. Untuk mengatasi masalah ini perusahaan dapat menggunakan teknologi informasi untuk mengolah pengelompokan data pelanggan yang mempunyai loyalitas tinggi terhadap perusahaan. Hal tersebut dapat dilakukan secara cepat dengan akurasi tinggi meskipun jumlah data transaksi pelanggan sangat besar serta tingkat kompleksitas yang tinggi. Salah satu cara yang dapat digunakan untuk mendeteksi loyalitas pelanggan, di antaranya adalah dengan menggunakan analisis Recency Frequency Monetary (RFM), dan teknik untuk melakukan pengelompokan dapat menggunakan clustering dengan algoritma. K-Means.

**Kata Kunci:** Loyalitas pelanggan, RFM, Clustering, Algoritma K-Means

### Abstract

*With so many companies engaged in the same field, the level of competition is very high, so that it will provide a threat for companies to lose customers. To overcome this problem companies can use information technology to process the grouping of customer data that has high loyalty to the company. This can be done quickly with high accuracy even though the amount of customer transaction data is very large and the level of complexity is high. One way that can be used to detect customer loyalty, among them is to use a Monetary Frequency Recency (RFM) analysis, and techniques for grouping can use clustering with the K-Means algorithm.*

**Keywords:** Customer loyalty, RFM, Clustering, C-Means Algorithm

### 1. PENDAHULUAN

Dalam perusahaan yang bergerak dalam bidang penjualan pelanggan sangat dibutuhkan untuk menjamin keberlangsungan dan juga keuntungan sebuah bisnis. Pelanggan atau customer adalah individu atau kelompok yang terbiasa membeli sebuah produk atau jasa berdasarkan keputusan mereka atas pertimbangan manfaat maupun harga yang kemudian melakukan hubungan dengan perusahaan melalui telepon, surat, dan fasilitas lainnya untuk mendapatkan suatu penawaran baru dari perusahaan (Greenberg, 2010:8). Pelanggan yang puas akan sebuah produk memberikan banyak keuntungan bagi perusahaan dimana salah satu yang penting yaitu memungkinkan tercapainya loyalitas pelanggan (Lovelock et al 2005:395).

Loyalitas pelanggan adalah tempat diperolehnya laba nyata bisnis ini, bukan pada usaha menarik orang asing kedalam toko anda untuk pembelian satu kali berdasarkan penjualan berbiaya tinggi atau iklan yang mahal. Loyalitas muncul karena adanya suatu rasa kepercayaan dari konsumen karena kepuasan setelah menggunakan

suatu produk, kepuasan pelanggan terjadi apabila harapan pelanggan sesuai dengan apa yang dialami dan dirasakan bahkan melebihi harapannya. Kepuasan pelanggan juga akan terpenuhi apabila pemberi jasa mampu meneliti kriteria-kriteria spesifik masing-masing jasa yang akan diberikan pada konsumen (Sugiharti, 2012).

Dengan banyaknya perusahaan yang bergerak dalam bidang yang sama, maka tingkat persaingan sangat tinggi, sehingga akan memberikan ancaman bagi perusahaan untuk kehilangan pelanggan. Untuk mengatasi masalah ini perusahaan dapat menggunakan teknologi informasi untuk mengolah pengelompokan data pelanggan yang mempunyai loyalitas tinggi terhadap perusahaan. Hal tersebut dapat dilakukan secara cepat dengan akurasi tinggi meskipun jumlah data transaksi pelanggan sangat besar serta tingkat kompleksitas yang tinggi. Ada beberapa cara yang dapat digunakan untuk mendeteksi loyalitas pelanggan, di antaranya adalah dengan menggunakan analisis Recency Frequency Monetary (RFM), yaitu sebuah teknik marketing yang digunakan untuk melihat perilaku pelanggan, seperti waktu terakhir pelanggan



pelanggan melakukan transaksi (frequency) dan seberapa besar nilai uang dari pelanggan atas transaksinya (monetary) (D. Birant, 2013).

Segmentasi pelanggan dapat dilakukan untuk mengetahui kelompok pelanggan dengan kesamaan perilaku. Teknik data mining dapat memberikan peran yang signifikan untuk menopang peran ahli marketing dalam mengetahui perilaku pelanggan berdasarkan sekumpulan data yang terkait dengan pelanggan (O. Dogan, 2018), namun hal tersebut masih tergantung pada ahli marketing dalam hal memanfaatkan teknik tersebut.

Salah satu teknik dalam data mining yang dapat digunakan untuk melakukan segmentasi adalah clustering, antara lain dengan algoritma K-Means. Teknik clustering merupakan teknik untuk mengelompokkan sekumpulan objek data yang memiliki kesamaan karakter akan masuk dalam kelompok tertentu, serta sekumpulan objek data yang tidak memiliki kesamaan karakter akan dikelompokkan pada cluster yang lain (R. A. Soeini, 2012). Melalui algoritma pengelompokan (clustering) yang dikombinasikan dengan analisis RFM dapat digunakan untuk mengetahui segmentasi perilaku pelanggan terutama terkait dengan loyalitas pelanggan.

## 2. Metode

### a. Algoritma Clustering

Menurut Asroni, Ronald Adrian Jurnal Ilmiah Semesta Teknika Vol. 18, No. 1, 76-82, Mei 2015, menyatakan Algoritma K-Means merupakan algoritma klasterisasi yang mengelompokkan data berdasarkan titik pusat klaster (centroid) terdekat dengan data. Tujuan dari K-Means adalah pengelompokkan data dengan memaksimalkan kemiripan data dalam satu klaster dan meminimalkan kemiripan data antar klaster. Ukuran kemiripan yang digunakan dalam klaster adalah fungsi jarak. Sehingga pemaksimalan kemiripan data didapatkan berdasarkan jarak terpendek antara data terhadap titik

Menurut fajar astuti herawati (2018:127), *K-Means Clustering* menggunakan pendekatan partitional clustering. Tiap cluster dihubungkan dengan sebuah centroid (titik pusat). Tiap titik di tempatkan ke dalam cluster dengan centroid terdekat. Jumlah cluster, K, Harus ditentukan. Algoritma dasarnya sangat sederhana, yaitu:

1. Pilih K titik sebagai centroid awal
2. Ulangi
3. Bentuk K cluster dengan menempatkan semua titik yang terdekat,
4. Ulangi perhitungan centroid dari tiap cluster,
5. Sampai centroid tidak berubah.

menjelaskan bahwa *K-Means Clustering* adalah, K dimaksudkan sebagai konstanta jumlah cluster yang diinginkan, *Means* dalam hal ini berarti nilai suatu rata-rata dari suatu grup data yang dalam hal ini didefinisikan sebagai cluster, sehingga *K-Means Clustering* adalah suatu metode penganalisaan data atau metode data mining yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode *K-Means* berusaha mengelompokkan data yang ada kedalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada didalam kelompok yang lain.

### b. Penerapan Metode K-Means Clustering

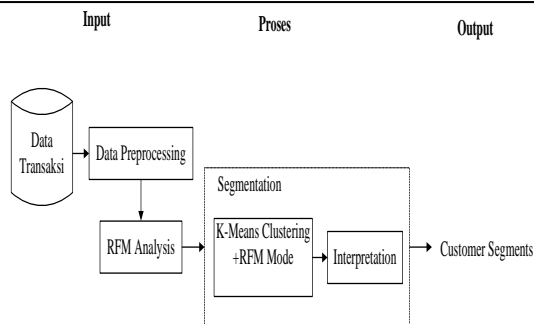
Pada penelitian ini dilakukan tahapan-tahapan yang dimulai dari penyiapan dataset transaksi. Data Transaksi diperoleh dari sistem transaksi yang ada sebelumnya berupa data spread sheet. Sesuai dengan konsep data mining, maka dilakukan cleaning data sesuai dengan kebutuhan proses data mining. Gambar 1 memperlihatkan proses utama yang dilakukan untuk menghasilkan segmentasi loyalitas pelanggan berbasis RFM.

Pada penelitian yang dilakukan pada kali ini akan difokuskan untuk menyelesaikan 2 (dua) masalah secara bersamaan, yaitu:

1. Masalah pertama yang ingin diselesaikan pada penelitian kali ini adalah bagaimana mencari nilai dari RFM (Recency, Frequency, Monetary) dari banyak yang transaksi yang dilakukan oleh pelanggan.
2. Masalah kedua yang ingin diselesaikan pada penelitian kali ini adalah bagaimana mengkombinasikan nilai RFM (Recency, Frequency, Monetary) yang di dapat di permasalahan yang pertama dengan menggunakan metode clustering k-means. Untuk melakukan pengelompokkan dari loyalitas pelanggan.

Dari bahasan dapat di gambarkan seperti digambar dibawah ini:





**Gambar 1. Proses Pengelompokan RFM dengan Clustering K-Means**

Data yang digunakan dalam penelitian ini ada bersumber dari database UCI dataset Ritel Online II yang berasal dari web <https://archive.ics.uci.edu/ml/datasets.php>. Data yang tersedia di UCI dataset Ritel Online II adalah tahun 2009-2010 dan 2010-2011. Pada tahun 2009-2010 Data yang digunakan adalah.

1. Data customer yang melakukan transaksi.
2. Data riwayat transaksi dengan atribut yang dibutuhkan oleh RFM, yang terdiri dari:
  - a. Waktu transaksi terakhir
  - b. Jumlah nominal transaksi
  - c. Jumlah frekuensi transaksi

#### RFM Analysis

Sebelum proses segmentasi dengan clusterisasi, maka perlu ditetapkan atribut-atribut sebagai nilai R, F dan M pada dataset yang ada. Perhitungan nilai R, F dan M diperoleh melalui perhitungan selisih waktu sekarang dengan tanggal transaksi pada setiap data penjualan untuk atribut R (recency), sedangkan untuk nilai atribut F (frequency) diperoleh dari banyaknya invoice berdasarkan pelanggan yang sama. Dan terakhir penentuan nilai M (monetary) dengan menggunakan perkalian dari kuantitas barang yang dibeli dengan harga satuan sesuai dengan harga barang.

**Tabel 1.**  
Tabel Recency Frequency dan Monetary Transaksi Pelanggan

Customer id	R	F	M
13313	17-11-2011	5	1555.32
18097	02-12-2011	4	2510.33
16656	17-11-2011	13	8197.04
16875	03-10-2011	4	2095.53
13094	18-11-2011	11	1703.64
17315	08-12-2011	10	6181.39
16255	23-11-2011	6	2245.66
13113	09-12-2011	20	11872.99
13458	02-12-2011	13	8460.71
13220	22-09-2011	4	819.29

Pada recency kurang begitu bagus, jadi diubah dengan aturan berikut 2011 menjadi 1, Bulan dan tanggal tidak dirubah

**Tabel 2.**  
Tabel Recency Frequency dan Monetary Transaksi Pelanggan

Customer id	R	F	M
13313	29	5	1555.32
18097	15	4	2510.33
16656	29	13	8197.04
16875	14	4	2095.53
13094	30	11	1703.64
17315	21	10	6181.39
16255	34	6	2245.66
13113	22	20	11872.99
13458	15	13	8460.71
13220	22	4	819.29

#### Proses Clustering menggunakan algoritma

K-Means Tahapan dalam pemakaian algoritma K-Means adalah menentukan banyaknya cluster, cluster yang akan dibuat sebanyak 3 cluster. Banyaknya cluster harus lebih kecil dari pada banyaknya data. Lalu, menentukan secara acak nilai centroid awal dari data yang sudah dimiliki.

1. Tentukan jumlah cluster  
Berdasarkan data Cluster yang akan dijadikan centroid awal dengan jumlah 3 cluster
2. Alokasikan data ke dalam kelompok secara acak  
Data penjualan dialokasikan berdasarkan Kode Produk yang ada di tabel 1
3. Hitung pusat cluster (centroid) menggunakan K-Means untuk masing-masing kelompok  
Pusat cluster yang diambil secara acak adalah

**Tabel 3.**  
Pusat Cluster

Pusat cluster C1	29	5	1555.32
Pusat cluster C2	15	4	2510.33
Pusat cluster C3	21	10	6181.39

Rumus menghitung jarak pusat cluster euclidian adalah sebagai berikut:

$$D(X_1, X_2) = \sqrt{\sum_{i=1}^p |X_{1i} - X_{2i}|^2}$$

1. Jarak Objek A

a. Pusat cluster 1

$$D = \sqrt{(1556-1556)^2 + (2130-2130)^2 + (2064-2064)^2 + (1639-1639)^2}$$

$$= 0$$

b. Pusat cluster 2

$$D = \sqrt{(1556-442)^2 + (2130-394)^2 + (2064-302)^2 + (1639-162)^2}$$

$$= 3088,829$$

c. Pusat cluster 3

$$D = \sqrt{(1556-230)^2 + (2130-283)^2 + (2064-185)^2 + (1639-233)^2}$$

$$= 3267,593$$

Pada tabel cluster dilakukan pengelompokkan sesuai dengan cluster dengan menggunakan rumus =IF(AND(C4<D4,C4<E4),1,IF(AND(D4<C4,D4<E4),2,3)

Proses perhitungan diatas dilakukan untuk masing-masing item terjual sehingga di dapat hasil perhitungan pada tabel literasi 1

**Tabel 4**  
Literasi 1

Data Ke-i	C1	C2	C3	Cluster
1	0.00	955.11	4626.08	1
2	955.11	0.00	3671.07	2
3	6641.72	5686.73	2015.67	3
4	540.42	414.80	4085.87	2
5	148.44	806.86	4477.76	1
6	4626.08	3671.07	0.00	3
7	690.36	265.36	3935.75	2
8	10317.68	9362.68	5691.61	3
9	6905.41	5950.39	2279.33	3
10	736.06	1691.05	5362.10	1

Setelah di dapat literasi 1, maka dilakukan pengelompokkan cluster untuk mendapat nilai pusat cluster yang baru dengan menggunakan cara

1. Pengelompokkan nilai 1

- a. Jika cluster = 1, maka menuliskan bilang pertama (2015), jika tidak akan bernilai 0, jadi berdasarkan bilangan pertama, maka hasilnya adalah 1556

kedua (2016), jika tidak akan bernilai 0, jadi berdasarkan bilangan kedua, maka hasilnya adalah 2130

c. Jika cluster = 1, maka menuliskan bilang ketiga (2017), jika tidak akan bernilai 0, jadi berdasarkan bilangan ketiga, maka hasilnya adalah 2064

d. Jika cluster = 1, maka menuliskan bilang keempat (2018), jika tidak akan bernilai 0, jadi berdasarkan bilangan keempat, maka hasilnya adalah 1639

2. Pengelompokkan cluster 2

a. Jika cluster = 2, maka menuliskan bilang pertama (2015), jika tidak akan bernilai 0, jadi berdasarkan bilangan pertama, maka hasilnya adalah 0

b. Jika cluster = 1, maka menuliskan bilang kedua (2016), jika tidak akan bernilai 0, jadi berdasarkan bilangan kedua, maka hasilnya adalah 0

c. Jika cluster = 1, maka menuliskan bilang ketiga (2017), jika tidak akan bernilai 0, jadi berdasarkan bilangan ketiga, maka hasilnya adalah 0

d. Jika cluster = 1, maka menuliskan bilang keempat (2018), jika tidak akan bernilai 0, jadi berdasarkan bilangan keempat, maka hasilnya adalah 0

3. Pengelompokkan cluster 3

a. Jika cluster = 3, maka menuliskan bilang pertama (2015), jika tidak akan bernilai 0, jadi berdasarkan bilangan pertama, maka hasilnya adalah 0

b. Jika cluster = 3, maka menuliskan bilang kedua (2016), jika tidak akan bernilai 0, jadi berdasarkan bilangan kedua, maka hasilnya adalah 0

c. Jika cluster = 3, maka menuliskan bilang ketiga (2017), jika tidak akan bernilai 0, jadi berdasarkan bilangan ketiga, maka hasilnya adalah 0

d. Jika cluster = 3, maka menuliskan bilang keempat (2018), jika tidak akan bernilai 0, jadi berdasarkan bilangan keempat, maka hasilnya adalah 0

Nilai rata-rata dari tabel di dapat dari mencari nilai berdasarkan tahun masing-masing.

Maka hasil dari pengelompokkan cluster pengelompokkan cluster untuk mendapat nilai pusat cluster yang baru dapat di lihat pada tabel 5.

**Tabel 5.**  
Pengelompokkan Berdasarkan Cluster Pada Literasi 1

Data ke-i	Cluster 1			Cluster 2			Cluster 3		
	R	F	M	R	F	M	R	F	M
1	29	5	1555.32	0	0	0	0	0	0
2	0	0	0	15	4	2510.33	0	0	0
3	0	0	0	0	0	0	29	13	8197.04
4	0	0	0	14	4	2095.53	0	0	0
5	30	11	1703.64	0	0	0	0	0	0
6	0	0	0	0	0	0	21	10	6181.39
7	0	0	0	34	6	2245.66	0	0	0
8	0	0	0	0	0	0	22	20	11873
9	0	0	0	0	0	0	15	13	8460.71
10	22	4	819.29	0	0	0	0	0	0
Total	81	20	4078.25	63	14	6851.52	87	56	34712.1
Jumlah data	3	3	3	3	3	3	4	4	4
Rata-rata	27	6.67	1359.42	21	4.67	2283.84	21.75	14	8678.03

4. Alokasikan masing-masing data ke centroid terdekat

Dari tabel 4 diatas dapat dilihat alokasi nilai data ke centroid terdekat dengan menentukan nilai minimum dari masing-masing baris. Setelah diketahui nilai ke centroid terdekat, maka nilai centroid baru dapat diketahui dengan cara menggunakan rumus rata-rata pada data awal di tabel diatas menggunakan nilai centroid terdekat di dapat nilai pusat cluster baru

**Tabel 6.**  
Pusat Cluster Baru Pada Literasi 1

Pusat cluster C1	27	6.67	1359.42
Pusat cluster C2	21	4.67	2283.84
Pusat cluster C3	21.75	14	8678.03

5. Kembali kelangkah 3, jika masih ada data yang berpindah cluster atau jika nilai centroid diatas nilai ambang, atau jika nilai pada fungsi obyektif yang digunakan masih diatas ambang batas

**Tabel 7.**  
Literasi 2

Data Ke-i	C1	C2	C3	Cluster
1	195.92	728.56	7122.72	1
2	1150.98	226.57	6167.71	2
3	6837.63	5913.21	481.05	3
4	736.23	188.44	6582.51	2
5	344.26	580.30	6974.40	1
6	4821.98	3897.55	2496.65	3
7	886.27	40.35	6432.39	2
8	10513.58	9589.16	3194.96	3
9	7101.31	6176.88	217.43	3
10	540.16	1464.55	7858.75	1

**Tabel 8.**  
Pengelompokkan Berdasarkan Cluster Pada Literasi 2

Data ke-i	Cluster 1			Cluster 2			Cluster 3		
	R	F	M	R	F	M	R	F	M
1	29	5	1555.3	0	0	0	0	0	0
2	0	0	0	15	4	2510.33	0	0	0
3	0	0	0	0	0	0	29	13	8197.04
4	0	0	0	14	4	2095.53	0	0	0
5	30	11	1703.6	0	0	0	0	0	0
6	0	0	0	0	0	0	21	10	6181.39
7	0	0	0	34	6	2245.66	0	0	0
8	0	0	0	0	0	0	22	20	11873
9	0	0	0	0	0	0	15	13	8460.71
10	22	4	819.29	0	0	0	0	0	0
Total	81	20	4078.3	63	14	6851.52	87	56	34712.1
Jumlah data	3	3	3	3	3	3	4	4	4
Rata-rata	27	6.67	1359.4	21	4.667	2283.84	21.75	14	8678.03

6. Alokasikan masing-masing data ke centroid terdekat

Dari tabel 4.7 diatas dapat dilihat alokasi nilai data ke centroid terdekat dengan menentukan nilai minimum dari masing-masing baris. Setelah diketahui nilai ke centroid terdekat, maka nilai centroid baru dapat diketahui dengan cara menggunakan rumus rata-rata pada data awal di tabel diatas menggunakan nilai centroid terdekat di dapat nilai pusat cluster baru

**Tabel 9.**  
Pusat Cluster Baru Pada Literasi 2

Pusat cluster C1	27	6.67	1359.4
Pusat cluster C2	21	4.67	2283.8
Pusat cluster C3	21.75	14	8678

7. Karena hasil dari pusat cluster yang baru dari literasi 1 dan 2 sama, maka perhitungan cluster dihentikan.

### 3. Hasil dan Pembahasan

#### a. Hasil

Hasil clustering dibawah ini merupakan hasil dari proses RFM dan Clustering K-Means

**Tabel 10.**  
Hasil Perhitungan Cluster

No	Customer id	Cluster
1	13313	1
2	18097	2
3	16656	3
4	16875	2
5	13094	1
6	17315	3
7	16255	2
8	13113	3
9	13458	3
10	13220	1

#### b. Pembahasan

Penelitian ini berdasarkan data 2011, dimana data pelanggan yang diuji hanya diambil sebanyak 10 data. Hal ini digunakan untuk memudahkan dalam melakukan uji coba secara manual dan dapat membandingkannya dengan menggunakan perangkat lunak WEKA (Waikato Environment for Knowledge Analysis). Adapun tampilan hasil proses dengan menggunakan perangkat lunak WEKA (Waikato Environment for Knowledge Analysis) dibagi menjadi 2 adalah sebagai berikut.

##### 1. Proses Clustering

Dari gambar dibawah dapat diketahui bahwa iterasi yang dihasilkan sebanyak 2 iterasi, cluster 0 ada sebanyak 3 data, cluster 1 ada sebanyak 3 data, serta cluster 4 ada sebanyak 4 data.

```

=== Clustering model (full training set) ===

kMeans
=====
Number of iterations: 2
Within cluster sum of squared errors: 0.8202564812596788

Initial starting points (random):

Cluster 0: 14,4,2095.53
Cluster 1: 34,6,2245.66
Cluster 2: 29,13,8197.04

Missing values globally replaced with mean/mode

```

Final cluster centroids:

Attribute	Cluster#		
	0	1	2
Full Data	(10.0)	(3.0)	(3.0)
Recency	23.1	17	31
Frequency	9	4	7.3333
Monetary	4564.19	1808.3833	1834.8733

Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

```

0    3 ( 30%)
1    3 ( 30%)
2    4 ( 40%)

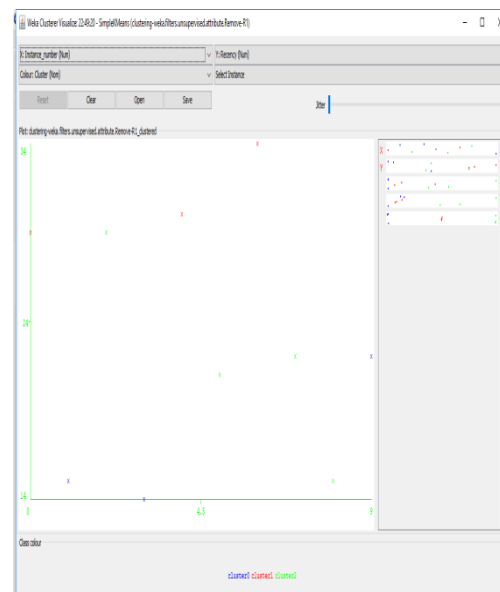
```

**Gambar 2.**

Hasil Proses Clustering K-Means Dengan Menggunakan WEKA.

##### 2. Tampilan Weka Cluster Visualize

Pada gambar dibawah ini bintang berwarna biru menunjukkan cluster 0, bintang berwarna merah menunjukkan cluster 1, dan bintang hijau menunjukkan cluster 2



**Gambar 3.**

Tampilan Weka Cluster Visualize

#### 4. Kesimpulan

Berdasarkan pengujian data penerapan RFM dengan Metode Clustering K-Means dapat mengelompokkan loyalitas pelanggan dan hasil dari uji coba menyimpulkan loyalitas pelanggan sangat baik dan dengan menggunakan aplikasi weka maka proses semakin efektif.

**5. Referensi**

- [1] Greenberg. Paul. 2010. CRM at the speed of light : Social CRM Strategies, Tool, and Techniques for Engaging Your Customer. (4th edition). New York: McGraw-Hill, Inc.
- [2] Cristoper, Lovelock. 2005. Manajemen Pemasaran Jasa, Kelompok Gramedia Indeks Indonesia.
- [3] Sugiharti. (2012),” Membangun Citra Merek yang Positif Dalam Rangka Menciptakan Kepuasan dan Kesetiaan Pelanggan,” Sekolah Tinggi Ilmu Ekonomi (STIE) Total Win JBMA, Vol 1, No 1, Juli, h.54-64
- [4] Birant, D. (2013). Data Mining Using RFM Analysis. In Knowledge-Oriented Applications in Data Mining (p. 6). Turkey: Dokuz Eylul University
- [5] Blattberg, R. a. (136-144). Manage marketing by the customer equity test. Harvard Business Review. DOĞAN1, O. (2018). Customer Segmentation by Using RFM Model and Clustering Methods: A Case Study in Retail Industry. International Journal of Contemporary Economics and Administrative Sciences, 8(1)
- [6] Soeini, R. A. (2012). Customer Segmentation based on Modified RFM Model in the Insurance Industry. International Conference on Machine Learning and Computing. Singapore
- [7] Asroni, Ronald Adrian, (Mei 2015), *Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang*, Jurnal Ilmiah Semesta Teknika, ditemukanali 11 April 2018 , dari <http://journal.umy.ac.id/index.php/st/article/view/708/858>
- [8] Hermawati, Fajar Astuti. (2013). *Konsep dan Teori Pengolahan Citra Digital*, Yogyakarta : Andi Offset
- [9] Fina Nasari, Surya Darma (2015) “Penerapan K-Means Clustering pada Data Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Potensi Utama)” Jurnal Seminar Nasional Teknologi Informasi dan Multimedia 2015

