



**SEGMENTASI PELANGGAN MENGGUNAKAN ALGORITMA *FUZZY C-MEANS* (FCM) DAN ANALISIS RFM (*RECENCY, FREQUENCY, AND MONETERY*) PADA DATA PELANGGAN KEDAI**

**SKRIPSI**

**UUS RUSDIANA**

**1610511019**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI INFORMATIKA**

**2020**



**SEGMENTASI PELANGGAN MENGGUNAKAN ALGORITMA *FUZZY C-MEANS* (FCM) DAN ANALISIS RFM (*RECENCY, FREQUENCY, AND MONETERY*) PADA DATA PELANGGAN KEDAI**

**SKRIPSI**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana  
Komputer**

**UUS RUSDIANA**

**1610511019**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI INFORMATIKA**

2020

### **PERNYATAAN ORISINALITAS**

Skripsi ini adalah hasil karya saya sendiri, dan semua sumber yang dikutip maupun yang dirujuk saya nyatakan dengan benar.

Nama : Uus Rusdiana

NIM : 1610511019

Tanggal : 18 Juni 2020

Apabila dikemudian hari ditemukan ketidak sesuaian terhadap pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 18 Juni 2020

Yang menyatakan,



(Uus Rusdiana)

## PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Uus Rusdiana  
NIM : 1610511019  
Fakultas : Ilmu Komputer  
Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

**Segmentasi Pelanggan Menggunakan Algoritma *Fuzzy C-Means* (FCM) dan Analisis RFM (*Recency, Frequency, And Monetary*) pada Data Pelanggan Kedai**

Beserta perangkat yang ada (bila diperlukan). Dengan hak bebas royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Skripsi selama masih mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenar-benarnya.

Dibuat di : Jakarta  
Pada tanggal : 18 Juni 2020

Yang menyatakan,



(Uus Rusdiana)

## LEMBAR PENGESESAHAN

Dengan ini dinyatakan bahwa Skripsi berikut:

Nama : Uus Rusdiana  
NIM : 1610511019  
Program Studi : Informatika  
Judul Tugas Akhir : Segmentasi Pelanggan Menggunakan Algoritma *Fuzzy C-Means* (FCM) dan Analisis RFM (*Recency, Frequency, and Monetary*) pada Data Pelanggan Kedai

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Yuni Widiastiwi, S. Kom, M.Si.

Ketua Penguji



Bambang Tri Wahyono, S.Kom, M.Si.

Anggota Penguji



Iin Ernawati, S. Kom., M. Si

Pembimbing I




Noor Falih, S. Kom., M.T.

Pembimbing II



Dr. Ermanita, M. Kom.

Dekan



Anita Muliawati, S. Kom., MTI.

Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Persetujuan : 18 Juni 2020



**SEGMENTASI PELANGGAN MENGGUNAKAN ALGORITMA *FUZZY C-MEANS* (FCM) DAN ANALISIS RFM (*RECENCY, FREQUENCY, AND MONETERY*) PADA DATA PELANGGAN KEDAI**

**UUS RUSDIANA**

**ABSTRAK**

Segmentasi pelanggan bertujuan untuk mengelompokkan pelanggan berdasarkan kesamaan karakteristik berupa ketertarikan atau permintaan pelanggan yang serupa. Untuk melakukan segmentasi pelanggan dapat dicapai melalui *data mining* dengan mengimplementasikan teknik *clustering*. Algoritma yang sering digunakan untuk melakukan *clustering* yaitu algoritma *fuzzy c-means* (berbasis *fuzzy*) dan algoritma *k-means* (berbasis klasik). Pemilihan metode pengukuran jarak data yang optimal untuk algoritma *clustering* menjadi masalah utama pada penelitian ini. Karena penggunaan metode pengukuran jarak data berpengaruh signifikan terhadap kualitas *cluster* yang dihasilkan. Algoritma yang akan digunakan pada penelitian ini yaitu *fuzzy c-means clustering* dengan *k-means clustering* sebagai pembanding yang kemudian akan coba diterapkan metode pengukuran jarak data *Euclidean distance*, *manhattan distance*, *Chebyshev distance*, dan *minkowski distance*. *Cluster* yang dihasilkan menggunakan metode pengukuran jarak data yang berbeda-beda akan dievaluasi menggunakan indeks validitas diantaranya *partition coefficient index* (PC), *modified partition coefficient index* (MPC), dan RMSE. Hasil penelitian menunjukkan algoritma *fuzzy c-means* lebih unggul daripada *k-means* dengan metrik jarak yang optimal yaitu *manhattan distance* (PC = 0.95, MPC = 0.9, dan RMSE = 0.7745) untuk pengujian pada *cluster* sebesar 2 dan *minkowski distance* (PC = 0.9338, MPC = 0.9007, dan RMSE = 0.8366) untuk pengujian pada *cluster* sebesar 3. Analisis RFM terhadap hasil segmentasi pelanggan menggunakan algoritma *fuzzy c-means* menunjukkan jumlah *cluster* yang optimal yaitu sebanyak 3 *cluster* yang memisahkan pelanggan ke dalam tiga karakteristik yaitu retensi pelanggan tinggi (nilai RFM = 3 hingga 5), retensi pelanggan sedang (nilai RFM = 2 hingga 3) dan retensi pelanggan rendah (nilai RFM=1).

**Kata kunci:** Segmentasi pelanggan, Analisis RFM, *Fuzzy C-Means*, *K-Means*, Metrik jarak.

**CUSTOMER SEGMENTATION USING FUZZY C-MEANS (FCM)  
ALGORITHM AND RFM (RECENCY, FREQUENCY, AND MONETERY)  
ANALYSIS ON CUSTOMER DATA**

**UUS RUSDIANA**

**ABSTRACT**

Customer segmentation aims to classify customers based on the similarity of characteristics in the form of similar customer interests or demands. To perform customer segmentation, it can be achieved through data mining by implementing clustering techniques. The algorithms that are often used for clustering are the fuzzy c-means algorithm (fuzzy based) and the k-means algorithm (classical based). The selection of the optimal distance metric measurement method for the clustering algorithm is the main problem in this study. Because the use of the distance metric measurement method has a significant effect on the quality of the resulting clusters. The algorithm that will be used in this study is fuzzy c-means clustering with k-means clustering as a comparison which will then try to apply the Euclidean distance, Mahnattan distance, Chebyshev distance, and Minkowski distance measurement methods. The resulting clusters using different distance metrics measurement methods will be evaluated using validity indices including Partition Coefficient Index (PC), Modified Partition Coefficient Index (MPC), and RMSE. The results show that the fuzzy c-means algorithm is superior to k-means with the optimal distance metric, namely Manhattan distance (PC = 0.95, MPC = 0.9, and RMSE = 0.7745) for testing on clusters of 2 and minkowski distance (PC = 0.9338, MPC = 0.9007, and RMSE = 0.8366) for testing on clusters of 3. RFM analysis of the results of customer segmentation using the fuzzy c-means algorithm shows the optimal number of clusters, namely 3 clusters which separate customers into three characteristics, namely high customer retention (value RFM = 3 to 5), moderate customer retention (RFM value = 2 to 3) and low customer retention (RFM value = 1).



**Keywords:** Customer Segementation, RFM Analysis, Fuzzy C-Means, K-Means, Distance Metric.

## KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Allah SWT atas rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan penulisan skripsi yang berjudul “**Segmentasi Pelanggan Menggunakan Algoritma *Fuzzy C-Means* (FCM) Dan Analisis RFM (*Recency, Frequency, And Monetary*) Pada Data Pelanggan Kedai**” dengan tepat waktu.

Penulisan skripsi ini merupakan salah satu syarat untuk memperoleh gelar sarjana Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta. Untuk itu, ungkapan terimakasih penulis sampaikan kepada:

1. Ibu Dr. Ermatita Zuhairi Sattar, M. Kom, selaku Dekan Fakultas Ilmu Komputer.
2. Ibu Anita Muliawati, S. Kom, MTI, selaku kepala program studi Informatika.
3. Ibu Iin Ernawati, S. Kom., M. Si, selaku dosen Pembimbing I
4. Bapak Noor Falih, S. Kom., M.T, selaku dosen Pembimbing II
5. Dan kepada rekan penulis angkatan 2016 yang selalu memberikan dukungan kepada penulis dalam proses penulisan laporan skripsi.

Akhir kata, semoga skripsi ini dapat memberikan manfaat kepada para pembaca yang budiman.

Jakarta, 28 Mei 2020



Uus Rusdiana

## DAFTAR ISI

PERNYATAAN ORISINALITAS .....	ii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI .....	iii
LEMBAR PENGESESAHAN .....	iv
ABSTRAK .....	v
ABSTRACT .....	vii
KATA PENGANTAR.....	ix
DAFTAR ISI .....	x
DAFTAR GAMBAR .....	xiii
DAFTAR TABEL.....	xv
DAFTAR SIMBOL.....	xvi
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Ruang Lingkup .....	2
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Luaran Penelitian.....	4
1.7 Sistematika Penulisan .....	4
BAB 2 TINJAUAN PUSTAKA .....	6
2.1 Segmentasi Pelanggan ( <i>Customer Segmentation</i> ).....	6
2.2 Model Analisis RFM ( <i>Recency, Frequency, dan Monetary</i> ).....	6
2.3 Penambangan Data ( <i>Data Mining</i> ).....	8
2.3.1 <i>Cross Industry Standard Process for Data Mining (CRISP-DM)</i> .....	9
2.3.2 Tugas Pada <i>Data Mining</i> .....	11
2.4 <i>Clustering</i> (Pengelompokan) .....	12
2.5 Logika <i>Fuzzy</i> .....	13
2.5.1 <i>Membership Function</i> (Fungsi Keanggotaan) .....	13
2.5.2 <i>Crisp Sets</i> (Himpunan klasik).....	14
2.5.3 <i>Fuzzy Sets</i> (Himpunan <i>Fuzzy</i> ) .....	14
2.5.4 Partisi .....	15
2.6 <i>K-Means Clustering</i> .....	16
2.6.1 Algoritma <i>K-Means</i> .....	17
2.7 <i>Fuzzy C-Means (FCM) Clustering</i> .....	17

2.7.1 Algoritma Fuzzy C-Means .....	18
2.8 Metrik Jarak .....	20
2.9 Indeks Validitas.....	22
2.10 Studi Literatur .....	23
BAB 3 METODOLOGI PENELITIAN .....	25
3.1 Kerangka Berpikir .....	25
3.1.1 Identifikasi Masalah.....	26
3.1.2 Studi Literatur.....	26
3.1.3 Pengumpulan Data.....	26
3.1.4 Persiapan Data .....	26
3.1.5 Pemodelan .....	27
3.1.6 Evaluasi.....	27
3.1.7 Penulisan Laporan .....	27
3.2 Perangkat Penelitian .....	27
3.2.1 Perangkat Lunak .....	27
3.2.2 Perangkat Keras.....	27
3.3 Waktu dan Tempat Penelitian .....	28
3.4 Jadwal Penelitian.....	28
BAB 4 HASIL DAN PEMBAHASAN.....	29
4.1 Data .....	29
4.1.1 Deskripsi Data .....	29
4.1.2 Praproses Data.....	31
4.2 Pemodelan.....	37
4.2.1 Membangun Model <i>Fuzzy C-Means Clustering</i> .....	37
4.2.2 Membangun Model <i>K-Means Clustering</i> .....	38
4.3 Pengujian Model .....	39
4.3.1 Menggunakan <i>Euclidean Distance</i> .....	40
4.3.2 Menggunakan <i>Manhattan Distance</i> .....	42
4.3.3 Menggunakan <i>Chebyshev Distance</i> .....	45
4.3.4 Menggunakan <i>Minkowski Distance</i> .....	48
4.4 Pembahasan Hasil Pengujian .....	50
4.4.1 Pengujian Menggunakan <i>Fuzzy C-Means</i> .....	51
4.4.2 Pengujian Menggunakan <i>K-Means</i> .....	52
4.4.3 Kesimpulan Pengujian .....	53
4.5 Melakukan Segmentasi Pelanggan .....	54

4.5.1 Segmentasi Ke dalam Dua <i>Cluster</i> .....	54
4.5.2 Segmentasi Ke dalam Tiga <i>Cluster</i> .....	59
BAB 5 KESIMPULAN DAN SARAN.....	66
5.1 Kesimpulan .....	66
6.2 Saran .....	67
DAFTAR PUSTAKA .....	68
DAFTAR RIWAYAT HIDUP .....	71
LAMPIRAN .....	72

## DAFTAR GAMBAR

Gambar 2.1 Fase CRISP-DM.....	10
Gambar 2.2 Contoh himpunan klasik.....	14
Gambar 2.3 Contoh himpunan fuzzy .....	15
Gambar 3.1 Diagram metode penelitian.....	25
Gambar 4.1 Sampel file rekaman transaksi (01-10-2019).....	29
Gambar 4.2 Skema analisis model RFM.....	30
Gambar 4.3 Gambaran data transaksi .....	31
Gambar 4.4 Missing values pada file record transaksi.....	31
Gambar 4.5 Total data keseluruhan .....	32
Gambar 4.6 Duplikat customr_id.....	33
Gambar 4.7 Hasil grouping data (berdasarkan customer_id) .....	34
Gambar 4.8 Data setelah grouping.....	34
Gambar 4.9 Konversi invoice_no ke interval hari .....	35
Gambar 4.10 Hasil normalisasi ke rentang 0 - 100.....	36
Gambar 4.11 Data setelah dilakukan scoring ke dalam RFM .....	36
Gambar 4.12 Hasil FCM dengan (k=2, metric=euclidean) .....	38
Gambar 4.13 Hasil K-Means dengan (k=2, metric=euclidean) .....	39
Gambar 4.14 Plotting FCM (kiri) dan K-Means (kanan) (k/c = 2, metric=euclidean).....	40
Gambar 4.15 Plotting FCM (kiri) dan K-Means (kanan) (k/c = 3, metric=euclidean).....	41
Gambar 4.16 Plotting FCM (kiri) dan K-Means (kanan) (k/c = 4, metric=euclidean).....	41
Gambar 4.17 Plotting FCM (kiri) dan K-Means (kanan) (k/c = 5, metric=euclidean).....	42
Gambar 4.18 Plotting FCM (kiri) dan K-Means (kanan) (k/c = 2, metric=manhattan) ....	43
Gambar 4.19 Plotting FCM (kiri) dan K-Means (kanan) (k/c = 3, metric=manhattan) ....	43
Gambar 4.20 Plotting FCM (kiri) dan K-Means (kanan) (k/c = 4, metric=manhattan) ....	44
Gambar 4.21 Plotting FCM (kiri) dan K-Means (kanan) (k/c = 5, metric=manhattan) ....	45
Gambar 4.22 Plotting FCM (kiri) dan K-Means (kanan) (k/c = 2, metric=chebyshev) ....	45
Gambar 4.23 Plotting FCM (kiri) dan K-Means (kanan) (k/c = 3, metric=chebyshev) ....	46
Gambar 4.24 Plotting FCM (kiri) dan K-Means (kanan) (k/c = 4, metric=chebyshev) ....	47
Gambar 4.25 Plotting FCM (kiri) dan K-Means (kanan) (k/c = 5, metric=chebyshev) ....	47
Gambar 4.26 Plotting FCM (kiri) dan K-Means (kanan) (k/c = 2, metric=minkowski) ...	48
Gambar 4.27 Plotting FCM (kiri) dan K-Means (kanan) (k/c = 3, metric=minkowski) ...	49
Gambar 4.28 Plotting FCM (kiri) dan K-Means (kanan) (k/c = 4, metric=minkowski) ...	49

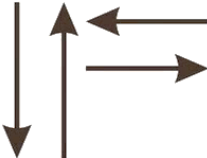


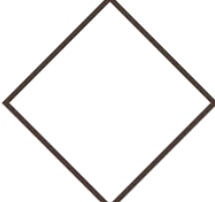


Gambar 4.29 Plotting FCM (kiri) dan K-Means (kanan) ( $k/c = 5$ , $\text{metric}=\text{minkowski}$ ) ...	50
Gambar 4.30 Plot hasil pengelompokan dua kelompok .....	54
Gambar 4.31 Grafik jumlah pelanggan cluster 0 dan 1 .....	55
Gambar 4.32 Data pelanggan cluster 0 .....	56
Gambar 4.33 Data pelanggan cluster 1 .....	57
Gambar 4.34 Grafik pelanggan pada cluster 0.....	58
Gambar 4.35 Grafik pelanggan pada cluster 1.....	58
Gambar 4.36 Plot hasil pengelompokan 3 kelompok.....	59
Gambar 4.37 Grafik jumlah pelanggan cluster 0, 1, dan 2 .....	60
Gambar 4.38 Data pelanggan cluster 0 .....	60
Gambar 4.39 Data pelanggan cluster 1 .....	61
Gambar 4.40 Data pelanggan cluster 2 .....	62
Gambar 4.41 Grafik pelanggan pada cluster 0.....	63
Gambar 4.42 Grafik pelanggan pada cluster 1.....	64
Gambar 4.43 Grafik pelanggan pada cluster 2.....	64

## DAFTAR TABEL

Tabel 2.1 Kuintil untuk pembagian data .....	8
Tabel 3.1 Jadwal penelitian .....	28
Tabel 4.1 Gambaran transformasi data ke model RFM .....	30
Tabel 4.2 Deskripsi atribut .....	30
Tabel 4.3 Nilai idex FCM (k=2, metric=euclidean).....	38
Tabel 4.4 Hasil K-Means dengan (k=2, metric=euclidean).....	39
Tabel 4.5 Nilai indeks validitas (k/c=2, metric=euclidean).....	40
Tabel 4.6 Nilai indeks validitas (k/c=3, metric=euclidean).....	41
Tabel 4.7 Nilai indeks validitas (k/c=4, metric=euclidean).....	42
Tabel 4.8 Nilai indeks validitas (k/c=5, metric=euclidean).....	42
Tabel 4.9 Nilai indeks validitas (k/c=2, metric=manhattan) .....	43
Tabel 4.10 Nilai indeks validitas (k/c=3, metric=manhattan.....	44
Tabel 4.11 Nilai indeks validitas (k/c=4, metric=manhattan).....	44
Tabel 4.12 Nilai indeks validitas (k/c=5, metric=manhattan).....	45
Tabel 4.13 Nilai indeks validitas (k/c=2, metric=chebyshev) .....	46
Tabel 4.14 Nilai indeks validitas (k/c=3, metric=chebyshev) .....	46
Tabel 4.15 Nilai indeks validitas (k/c=4, metric=chebyshev) .....	47
Tabel 4.16 Nilai indeks validitas (k/c=5, metric=chebyshev) .....	48
Tabel 4.17 Nilai indeks validitas (k/c=2, metric=minkowski) .....	48
Tabel 4.18 Nilai indeks validitas (k/c=3, metric=minkowski) .....	49
Tabel 4.19 Nilai indeks validitas (k/c=4, metric=minkowski) .....	50
Tabel 4.20 Nilai indeks validitas (k/c=5, metric=minkowski) .....	50
Tabel 4.21 Hasil pengujian Fuzzy C-Means.....	51
Tabel 4.22 Hasil pengujian K-Means.....	52
Tabel 4.23 Perbandingan hasil pengujian (kesimpulan).....	53



## DAFTAR SIMBOL

Simbol	Nama	Keterangan
	<i>Flow Direction Symbol</i>	Berfungsi sebagai penghubung antar simbol.
	<i>Terminator Symbol</i>	Simbol untuk mengindikasikan permulaan atau terminasi.
	<i>Processing Symbol</i>	Menunjukkan pemrosesan yang dilakukan oleh komputer.
	<i>Decision Symbol</i>	Pemilihan kondisi.
	<i>Output Symbol</i>	Menyatakan proses input atau output.
	<i>Document Symbol</i>	Menyatakan input berasal dari dokumen.