Implementasi Algoritma *Clustering* Untuk Pengelompokan Pelanggan Retail Berdasarkan Skor *Recency*, *Frequency*, Dan *Monetary*

Moch.Irfan Chanafi[1)], Dian Puspita Hapsari[2)],
Rinci Kembang Hapsari[3)], Tutuk Indriyani[4)]
[1,2,3,4]Teknik Informatika, Institut Teknologi Adhi Tama Surabaya
email: irfanijoel@gmail.com, hapsaridp2015@gmail.com

ABSTRACT

Retail is a technique to trade products from sellers to customers. Most retail business doers usually concern with general problems such as customer loyalty. Consequently, retailers start to be interested in applying data mining to classify customers based on the similar characteristics and to identify differences among groups. For this reason, the researcher aimed at investigating the customer mapping in relation to the customer relationship management. RFM method was employed to make assessment to customers based on the scores of recency, frequency, and monetary. Next, clustering method of DBSCAN and Fuzzy C-Means were also carried out to categorize data on the retail customers. DBSCAN is a clustering method used for making groups of data based on the density levels between high and low. Meanwhile, Fuzzy C-Means will make groups of data based on the degree of membership for every datum. The data of customers were classified by both methods into 5 groups for each method. After that, the results of every clustering process by either DBSCAN or Fuzzy C-Means were assessed to know their validity of cluster by using Silhouette Index method. DBSCAN on clusters 2 – 5 got SI value 1, while Fuzzy C-Means had cluster 4 with the highest validity by *0.5584288019243665.*

Keywords: Retail, RFM Analysis, DBSCAN, Fuzzy C-Means, Clustering, Silhouette Index

ABSTRAK

Retail atau dalam Bahasa Indonesia disebut ecerean merupakan sebuah teknik untuk memasarkan produk yang dilakukan oleh penjual ke pelanggan. Para pembisnis retail biasanya memperhatikan masalah umum yang terdapat pada bisnis ini, seperti pelanggan mana yang setia. Berdasarkan kasus tersebut, pembisnis retail mulai tertarik untuk menerapkan proses penambangan data untuk mengelompokan pelanggan berdasarkan karakteristik pelanggan yang serupa dan mengidentifikasi perbedaan antar kelompok.Tujuan dari penelitian ini adalah untuk menyelidiki masalah pemetaan pelanggan yang kaitannya dengan *customer relationship management. Pada penelitian ini akan diterapkan metode RFM analysis untuk melakukan penilaian pelanggan berdasarkan skor recency, frequency, dan monetary. Selanjutnya diterapkan metode klustering DBSCAN dan Fuzzy C-Means untuk melakukan pengelompokan data pelanggan retail. DBSCAN merupakan sebuah metode klustering yang mengelompokan data berdasarkan tingkat kerapatan data yang tinggi dengan yang rendah, sedangkan Fuzzy C-Means mengelompokan data berdasarkan derajat keanggotaan setiap data. Data pelanggan akan dikelompokan dengan 2 metode tersebut menjadi masing – masing metode 5 kelompok. Selanjutnya hasil setiap proses klustering baik DBSCAN dan Fuzzy C-Means akan dinilai validitas klusternya dengan menggunakan metode Silhouette Index. Dimana untuk DBSCAN pada kluster 2 sampai 5 memiliki nilai SI 1 sedangkan untuk Fuzzy C-Means memiliki kluster dengan nilai validitas tertinggi yaitu kluster 4 dengan nilai* 0.5584288019243665.

Kata Kunci: Retail, RFM Analysis, DBSCAN, Fuzzy C-Means, Clustering, Silhouette Index.

PENDAHULUAN

Retail atau dalam Bahasa Indonesia disebut ecerean merupakan sebuah teknik untuk memasarkan produk yang dilakukan oleh penjual ke pelanggan, baik untuk penggunaan secara individu maupun secara kelompok. Para pembisnis *retail* biasanya memperhatikan masalah umum

dalam dunia retail antara lain, pelanggan mana yang sering berbelanja, produk yang sering dibeli oleh pelanggan, dan pelanggan mana yang paling setia.

Recency, frequency, monetary atau yang biasa disingkat dengan RFM merupakan sebuah metode yang digunakan untuk mengukur nilai dari seorang pelanggan berdasarkan riwayat transaksi yang pernah dilakukan. *Recency* merupakan *interval* transaksi awal dan transaksi akhir dari pelanggan. *Frequency* merupakan total dari transaksi pelanggan dalam kurun waktu tertentu. Dan *monetary* merupakan hasil kumulatif pengeluaran dana pelanggan dalam kurun waktu tertentu.

Density-based spatial clustering of applications with noise (DBSCAN)adalah algoritma pengelompokan yang didasarkan pada kepadatan (*density*) data. Metode ini ditemukan oleh Martin Ester, Hans-Peter Kriegel, Jörg Sander and Xiaowei Xu (1996). Konsep kepadatan yang dimaksud dalam DBSCAN adalah banyaknya data (minPts) yang berada dalam radius Eps (ε) dari setiap data. Konsep kepadatan seperti ini menghasilkan tiga macam status dari setiap data, yaitu inti (*core*), batas (*border*), dan *noise*. DBSCAN memiliki kelebihan, diantaranya metode ini dapat menemukan kluster yang secara total dikelilingi oleh kluster yang lain, dan juga dapat menemukan kluster yang memiliki bentuk yang tidak tentu (Ahmad & Dang, 2015)

Fuzzy C-Means atau FCM. *Fuzzy C–means* atau biasa juga disebut dengan *fuzzy K–means* merupakan suatu metode klustering yang mana keberadaan setiap titik data dalam sebuah kluster ditentukan oleh derajat keanggotaannya. Metode ini memungkinkan sebuah objek data untuk memiliki lebih dari satu kluster didalam derajat yang sama. *Output* dari algoritma FCM adalah deretan pusat kluster dan beberapa derajat keanggotaan untuk tiap – tiap titik data. . Metode *fuzzy C-Means* memiliki kelebihan dibandingkan dengan metode *C-Means*. Diantaranya, algoritma *fuzzy C-Means* memiliki akurasi dan kualitas cluster yang lebih baik dari algoritma *K-means* (Sivarathri, 2014). Selain itu, algoritma *fuzzy K-Means* cocok untuk menangani isu yang berkaitan untuk memahami suatu pola, data yang tidak lengkap, data yang tercampur, interaksi manusia dan dapat memberikan perkiraan solusi yang lebih cepat (Singh, 2014)

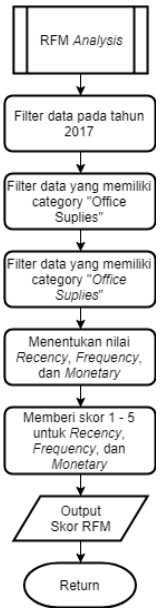
Silhouette Index merupakan salah satu teknik untuk menentukan cluster optimal yang mengacu pada metode intrepetasi dan validasi konsisten didalam suatu kelompok data (Rousseeuw, 1987). Metode ini merupakan salah satu algoritma optimasi kluster yang mengacu pada metode interpretasi dan validasi konsistensi dalam sebuah kelompok data. *Silhouette Index* memberikan representasi grafik ringkas tentang seberapa baik sebuah objek didalam klusternya. *Silhouette Index* memiliki kelebihan, diantaranya metode ini memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan *Davies-bouldin index* (prasetyo, 2012).

METODE

Pengelompokan data pelanggan merupakan sebuah proses untuk menempatkan pelanggan dalam sebuah kelompok tertentu berdasarkan karakteristik dari pelanggan tersebut, hal ini dapat diketahui dari riwayat transaksi yang dilakukan oleh pelanggan. Pada penelitian ini data yang digunakan adalah data *sample-supertore* pada negara Amerika Serikat, data ini memiliki 19 fitur data, namun dari fitur – fitur tersebut hanya akan digunakan 3 fitur yaitu *customersid, orderdate*, dan *sales* untuk menentukan skor RFM dari pelanggan. Pada penelitian ini, peneliti tidak melakukan proses *cleaning* data karena diasumsikan data tidak memiliki nilai yang kosong. Berikut merupakan tahapan dari system yang dibangun:

a. Input data, fitur data yang digunakan adalah *customersid, orderdate*, dan *sales*
b. Implementasi metode RFM *Analysis*, menentukan skor RFM dari pelanggan berdasarkan *input* data

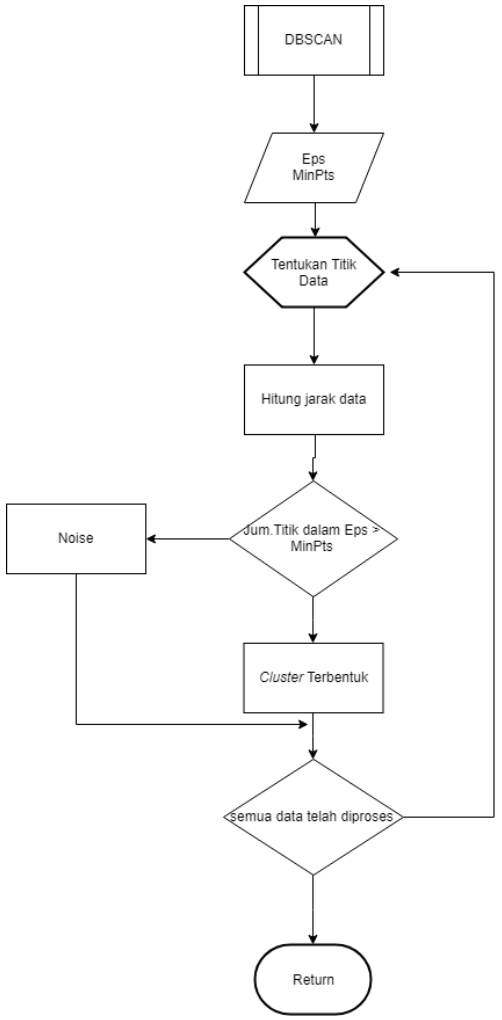
c. Pengelompokan pelanggan, melakukan pengelompokan pelanggan berdasarkan skor RFM dengan menggunakan DBSCAN dan *Fuzzy C-Means*
d. Penilaian validitas *cluster*, melakukan penilaian validitas *cluster* yang dihasilkan dengan menggunakan metode *Silhouette Index*



Gambar 2. 1 Flowchart RFM Analysis

Gambar 2.1 merupakan alur dari implementasi metode RFM *Analysis*. Dimana langkah awal ialah melakukan pemilihan data pelanggan yang melakukan transaksi pada tahun 2017. Selanjutnya memilih data pelanggan yang memiliki transaksi dengan category *office supplies*. Selanjutnya menentukan nilai *recency, frequency*, dan *monetary*. Nilai dari *recency* suatu *customer id* diambil dari waktu terakhir pelanggan melakukan transaksi dan waktu pertama pelanggan melakukan transaksi, atau bisa disebut juga dengan selisih waktu transaksi pelanggan. *Frequency* merupakan ukuran seberapa sering pelanggan melakukan transaksi, *frequency* diambil dari banyaknya data yang didasarkan pada *customer id*. Sedangkan *monetary* merupakan jumlah kumulatif biaya yang dikeluarkan oleh pelanggan, *monetary* didapatkan dari total *sales* yang didasarkan dari *customer id*.

Setelah didapatnya nilai *recency, frequency*, dan *monetary* selanjutnya akan ditentukan skor antara 1 sampai dengan 5. Dimana dari ketiga *variable* tersebut akan dibagi menjadi lima kelompok, dimana 20% nilai teratas akan diberikan nilai 5, 20% selanjutnya akan diberikan nilai 4 dan seterusnya. Setelah itu hasil yang didapat akan digunakan sebagai fitur untuk melakukan pengelomppokan data pelanggan. Selanjutnya, hasil dari skor RFM pelanggan tersebut akan digunakan untuk melakukan pengelompokan pelanggan. Dalam proses pengelompokan pelanggan akan diimplementasikan dua metode, yaitu DBSCAN dan *Fuzzy C-Means*.



Gambar 2. 2 Flowchart DBSCAN

Pada gambar 2.2 merupakan flowchart atau alur pengelompokan data dengan menggunakan metode DBSCAN, pada tahap awal akan tentukan *Eps* dan *MinPts*, lalu akan dipilih point awal (p) adalah data *index* ke-0, setelah itu akan diambil semua *point* yang *Density-reachable* dari point (p), lalu akan diuji apakah data termasuk *core point*, jika iya maka akan dibentuk sebuah *cluster*. namun apabila (p) merupakan *border point* dan tidak ada titik data yang *Density-reachable* dari titik (p) maka akan dilanjutkan ke data selanjutnya, proses ini dilakukan sampai semua titik data telah diproses.

Proses awal dari FCM ialah melakukan input data yang akan di*cluster (x)*. data tersebut berupa matriks *n x m* (*n* = jumlah sample data, *m* = atribut setiap data). x_{ij} = data sample ke *i* (*i* = 1,2,3,4,…,n), atribut ke-*j* (*j*= 1,2,3,…,m). Setelah itu tentukan jumlah cluster (*c*), pangkat (*w*), proses iterasi maksimum (*MaxIter*), *error* terkecil yang diharapkan (ξ), fungsi objektif awal (p_0) = 0, dan iterasi awal (*t=1*). Setelah itu bangkitkan nilai *random* (μ_{ik}), *i* = 1,2,3,…,n; *k* = 1,2,3,…,*c;* sebagai elemen – elemen matriks awal (*U*) dimana nilai dari setiap elemen ini adalah 0 sampai 1. Setelah itu hitung jumlah setiap kolom dengan persamaan berikut:

$$\text{[illegible]}_i = \sum_{k=1}^{c} \mu_{ik} \quad (2.1)$$

Dengan j = 1,2,3,…,n.

$$\mu_{ik} = \frac{\mu_{ik}}{\text{[illegible]}_i} \quad (2.2)$$

Setelah itu hitung pusat *cluster* ke-*k*: V_{kj}, dengan k = 1,2,3,…,c; dan j = 1,2,3,…,m dengan menggunakan persamaan berikut:

$$V_{kj} = \frac{\sum_{i=1}^{n}((\mu_{ik})^{w} * X_{ij})}{\sum_{i=1}^{n}(\mu_{ik})^{w})} \quad (2.3)$$

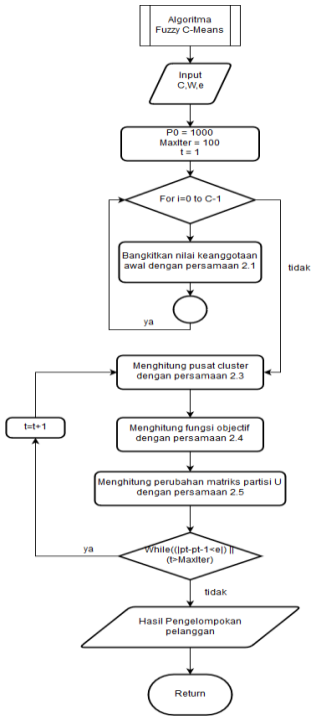
setelah ditentukannya pusat *cluster*, selanjutnya akan dilakukan perhitungan untuk fungsi objektif dengan menggunakan persamaan berikut:

$$P_t = \sum_{i=1}^{n}\sum_{k=1}^{c}([\sum_{i=1}^{m}(X_{ij} - V_{kj})^2](\mu_{ik})^{w}) \quad (2.4)$$

Setelah itu dilakukan perhitungan untuk mengetahui perubahan dari matriks partisi dengan menggunakan persamaan berikut:

$$\mu_{ik} = \frac{\sum_{j=1}^{m}[(X_{ij} - V_{kj})^2]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c}[\sum_{i=1}^{m}(X_{ij} - V_{kj})^2]^{\frac{-1}{m-1}}} \quad (2.5)$$

Setelah semua proses tersebut selesai, selanjutnya akan dilakukan pengecekan kondisi berhenti. Jika (|Pt – Pt-i| < ξ) atau (*t* > *MaxIter*) maka proses iterasi akan dihentikan, apabila tidak, maka proses akan diulangi dari langkah ke-4 atau penghitungan kembali pusat *cluster*, pada gambar 2.3 merupakan alur proses *Fuzzy C-Means.*



Gambar 2. 3 Flowchart Fuzzy C-Means

Setelah proses pengelompokan data pelanggan selesai, selanjutnya *cluster* yang dihasilkan dari metode DBSCAN dan *Fuzzy C-Means* akan dinilai tingkat validitas *cluster* yang dihasilkan dengan metode *Silhouette Index*. *Silhouette index* memiliki dua *variable* utama, yang pertama yaitu (*a*), merupakan jarak rata – rata antara datum ke-*i* dengan semua data didalam sebuah *cluster* yang

sama. Lalu terdapat (b), variable b(i) merupakan nilai rata – rata jarak terkecil terhadap data ke-i terhadap semua titik dari *cluster* yang lain

Persamaan dari silhouette index dapat didefinisikan seperti dibawah ini:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Pengelompokan data dapat di cirikan dengan lebar rata – rata dari nilai *silhouette* dari setiap individu. Rata – rata lebar nilai *silhouette* yang terbesar diatas K yang berbeda menunjukkan bahwa itu merupakan jumlah *cluster* yang terbaik (Kodinariya, 2013).

HASIL DAN PEMBAHASAN

Pada langkah awal implementasi RFM *Analysis*, dilakukan *input* data pelanggan berupa *file* .CSV. Dimana data ini memiliki 17 fitur, namun fitur yang akan digunakan hanya 3, yaitu *order date*, *customer id*, dan *sales*. Data yang digunakan memiliki 2003 *record* dalam skala waktu 1 tahun (01/01/2017 sampai dengan 30/12/2017) untuk *retail* di United States, dimana data yang digunakan memiliki kategori *office supplies*. Pada tabel 3.1 Merupakan 3 fitur utama yang akan digunakan dalam proses RFM.

Tabel 3. 1 Fitur yang digunakan

	OrderDate	CustomersId	Sales
0	5/4/2017	12220	243.160
1	2/26/2017	13060	29.600
2	8/3/2017	21055	16.520
3	12/11/2017	19435	22.750
4	12/11/2017	19435	735.980

Untuk menentukan nilai recency dari pelanggan, langkah awal adalah dengan menentukan tanggal terakhir transaksi pelanggan. Dalam kasus ini tanggal transaksi terakhir yang dilakukan pelanggan adalah tanggal 30/12/2017, selanjutnya akan ditentukan tanggal terbaru, dimana tanggal terbaru ini nanti akan dikurangi tanggal terakhir transaksi pelanggan, dalam kasus ini tanggal terbaru ditentukan pada 1/1/2018. Selanjutnya dibentuk sebuah tabel bernama recency, dimana tabel recency ini memiliki sebuah kolom TglPesanAkhir, dimana kolom ini berisi tanggal transaksi terakhir pelanggan, tabel recency. Selanjutnya ditambahkan kolom NilaiRecency, dimana nilai ini berasal dari proses pengurangan tanggal terbaru dengan tanggal terakhir transaksi, pada tabel 2.5 akan disajikan hasil dari tabel recency.

Tabel 3. 2 nilai *recency*

	CustomersId	NilaiRecency
0	10030	14
1	10045	119
2	10060	57
3	10075	36
4	10090	30

Langkah selanjutnya adalah menentukan nilai *frequency* pelanggan. dimana *frequency* berisi banyaknya CustomerId yang melakukan transaksi Pada tabel 3.3 akan disajikan tabel *frequency*

Tabel 3. 3 nilai *frequency*

	CustomersId	NilaiFrequency
0	10030	1
1	10045	3
2	10060	2
3	10075	4
4	10090	3

Selanjutnya adalah menentukan nilai *monetary*. Dimana nilai ini didapatkan dengan cara menjumlah sales dari setiap pelanggan. Pada tabel 3.4 akan disajikan nilai *monetary* dari setiap pelanggan.

Tabel 3. 4 nilai *monetary*

	CustomersId	NilaiMonetary
0	10030	18.704
1	10045	153.216
2	10060	88.536
3	10075	233.844
4	10090	601.244

Selanjutnya, dari ketiga tabel diatas akan digabungkan menjadi satu tabel yang akan disajikan dalam tabel 3.5. Selanjutnya berdasarkan tabel nilai RFM tersebut akan ditentukan skor RFM setiap pelanggan. Dimana berdasarkan nilai RFM akan dibagi menjadi 5 *quantile*, dimana untuk nilai *recency*, semakin kecil maka akan mendapatkan skor yang tinggi, sedangkan untuk *frequency* dan *monetary*, semakin tinggi nilainya maka akan mendapat nilai yang tinggi juga. Untuk nilai *quantile* yang ditentukan adalah 0.20, 0.40, 0.60, dan 0.80. pada tabel 3.6 akan disajikan data skor RFM pelanggan. Dimana, skor RFM ini yang digunakan sebagai fitur pengelompokan pelanggan.

Tabel 3. 5 tabel nilai RFM

	CustomersId	NilaiRecency	NilaiFrequency	NilaiMonetary
0	10030	14	1	18.704
1	10045	119	3	153.216
2	10060	57	2	88.536
3	10075	36	4	233.844
4	10090	30	3	601.244

Tabel 3. 6 skor RFM pelanggan

	CustomersId	SkorRecency	SkorFrequency	SkorMonetary
0	10030	5	1	1
1	10045	2	4	4
2	10060	3	2	3
3	10075	4	5	4
4	10090	4	4	5

Langkah awal implementasi metode DBSCAN adalah menentukan Eps dan MinPts. Pada penelitian ini digunakan Eps = 1 MinPts = 15, dan data ditambahkan kolom 'jarak' dengan nilai 0, dan juga kolom label dengan data berisi '*Unlabeled'* untuk seluruh data. Dalam gambar 3.7 akan disajikan kolom yang ditambahkan.

Tabel 3. 7 penambahan kolom

jarak	labels
0	unlabeled
0	unlabeled
0	unlabeled
0	unlabeled
0	unlabeled

Corepoint awal ditentukan pada data dengan *index* ke-0. Langkah awal adalah menghitung jarak setiap data terhadap data yang ditetapkan sebagai *core point*, setelah itu akan dihitung jumlah data yang memiliki jarak <= eps apabila jumlah data lebih dari MinPts akan dilabeli sebagai border,namun apabila jumlah data kurang dari MinPts maka data akan diberi label *noise*. Pada gambar 3.8 akan diberikan contoh perhitungan jarak semua data terhadap data ke-0 dan pemberian label.

Tabel 3. 8 hasil perhitungan jarak pada data ke-0

SkorRFM	jarak	labels
111	0.0	core
111	0.0	border
111	0.0	border
111	0.0	border
111	0.0	border

core point selanjutnya dipilih data yang memiliki jarak terkecil dari *core point* namun bukan termasuk border, langkah ini dilakukan sampai dengan semua data telah selesai diproses. Selanjutnya adalah penentuan kluster, hal ini dapat dilakukan dengan melihat apakah data tersebut *Density-connected,* hal ini dapat dilihat apabila ada sebuah *core point* yang berhubungan dengan *core point* lain melalui sebuah titik data dengan menghitung jarak antara *core point* terhadap data tersebut, apabila jarak kurang dari Eps maka data dan *core point* dinyatakan *Density-connected*. *Core point* yang dihasilkan terdapat pada data *index* ke 0, 39, 87, 103, 127, 250, 406, 422, dan 503 sedangkan untuk *noise* didapatkan sebanyak 400 baris data. Selanjutnya akan dibuat sebuah kolom baru bernama kluster untuk menampung kluster setiap data. Proses *clustering* dengan

menggunakan metode DBSCAN ini membutuhkan waktu kurang lebih 4,05229 detik. Pada tabel berikut akan disajikan hasil yang diperoleh dari implementasi metode DBSCAN.

Tabel 3. 9 DBSCAN kluster 1

Customers Id	Skor Recency	Skor Frequency	Skor Monetary	Skor RFM	*Cluster*
11770	1	1	1	111	1
15805	1	1	1	111	1
15910	1	1	1	111	1
10480	1	1	1	111	1

Pada table 3.9 merupakan contoh data pada *cluster* 1 yang memiliki data pelanggan dengan skor RFM 111, dimana pada *cluster* ini didapatkan pelanggan dengan rata – rata skor RFM terendah.

Tabel 3. 10 DBSCAN kluster 2

Customers Id	Skor Recency	Skor Frequency	Skor Monetary	Skor RFM	*Cluster*
20605	2	1	1	211	2
14950	2	1	1	211	2
15235	2	1	1	211	2
10300	2	1	1	211	2

Pada table 3.10 merupakan contoh data pada *cluster* 2 yang memiliki data pelanggan dengan skor RFM 211, pada cluster ini, berisi pelanggan dengan skor recency yang lebih tinggi dari cluster 1, namun skor frequency dan monetary sama.

Tabel 3. 11 DBSCAN kluster 3

CustomersId	Skor Recency	Skor Frequency	Skor Monetary	Skor RFM	*Cluster*
10645	3	1	1	311	3
10765	3	1	1	311	3
20320	3	1	1	311	3
17275	3	1	1	311	3

Pada table 3.11 merupakan contoh data pada *cluster* 3 yang memiliki data pelanggan dengan skor RFM 311, dimana data pada cluster ini memiliki skor recency yang lebih tinggi dari *cluster* satu dan dua namun dengan skor frequency dan monetary yang sama. Pada table 3.12 merupakan contoh data pada *cluster* 4 yang memiliki data pelanggan dengan skor RFM 422, dimana data pada cluster ini memiliki skor recency, frequency, dan monetary yang lebih tinggi dari cluster sebelumnya.

Tabel 3. 12 DBSCAn kluster 4

CustomersId	Skor Recency	Skor Frequency	Skor Monetary	Skor RFM	*Cluster*
20770	4	2	2	422	4
20305	4	2	2	422	4
19855	4	2	2	422	4
16525	4	2	2	422	4

Tabel 3. 13 DBSCAN kluster 5

CustomersId	Skor Recency	Skor Frequency	Skor Monetary	Skor RFM	*Cluster*
19795	4	2	3	423	5
18880	4	2	3	423	5
19390	4	2	3	423	5
15715	4	2	3	423	5

Pada table 3.13 merupakan contoh data pada *cluster* 5 yang memiliki data pelanggan dengan skor RFM 211, dimana data pada cluster ini memiliki skor monetary yang lebih tinggi dari cluster empat.

Tabel 3. 14 DBSCAN noise

CustomersId	Skor Recency	Skor Frequency	Skor Monetary	Skor RFM	*Cluster*
13225	5	2	5	525	0
18655	5	4	1	541	0
15325	1	1	3	113	0
21565	1	1	4	114	0

Pada table 3.14 merupakan contoh data yang termasuk kedalam *noise* data. dimana data – data ini tidak termasuk kedalam *cluster* yang telah ditemukan. Untuk data yang termasuk kedalam noise, maka data tersebut akan diberi nilai 0 pada kolom *cluster*. Sedangkan *cluster* yang berisi data pelanggan terbaik, dapat dilihat dari rata – rata skor RFM yang dimiliki pelanggan, dalam kasus ini *cluster* dengan pelanggan terbaik terdapat pada kluster 5. Metode DBSCAN memerlukan waktu eksekusi 4,05229 Detik. Selain DBSCAN, diterapkan juga metode Fuzzy C-Means.

Langkah Awal dalam proses pengelompokan pelanggan dengan menggunakan metode *Fuzzy C-Means*, adalah dengan menentukan inisialisasi parameter awal yang digunakan. Pada penelitian ini, penulis menggunakan parameter awal sebagai berikut:

jumlah *cluster* (*c*) = 5
pangkat (*w*) = 2
iterasi maksimum (*MaxIter*) = 100
error terkecil yang diharapkan (ξ) = 10^{-5}
fungsi objektif awal (p_0) = 10000
iterasi awal (*t*) = 1

Dengan menggunakan parameter awal diatas, hasil yang didapatkan dari proses pengeelompokan data pelanggan dengan metode *Fuzzy C-Means* sebagai berikut

MaxIter = 63
Waktu Eksekusi = 291, 171415090 Detik

Dalam tabel 3.15 akan disajikan centroid akhir yang didapatkan dari proses pengelompokan pelanggan dengan menggunakan metode *Fuzzy C-Means*. Selanjutnya pada table 3.16 sampai dengan 3.20 akan disajikan hasil pengelompokan pelanggan dengan menggunakan metode Fuzzy C-Means. Dimana c*luster* yang berisi data pelanggan terbaik dapat dilihat dari rata – rata skor RFM yang dimiliki pelanggan. Dalam kasus ini cluster dengan pelanggan terbaik terdapat pada kluster 4.

Tabel 3. 15 *Centroid* Akhir Data

	Fitur Data			
Cluster		Skor Recency	Skor Frequency	Skor Monetary
	1	4.0732752377352	1.5158384588749	1.745548527151
	2	1.7720548469040	1.8087336418578	4.076403704317
	3	1.4367620265061	1.1886305113306	1.478109739304
	4	3.9980909731516	4.3643223724259	4.191931283989
	5	2.1182489	4.1411572681688	2.036870383684

Tabel 3. 26 FCM *Cluster* 1

CustomersId	Skor Recency	Skor Frequency	Skor Monetary	Skor RFM	*Cluster*
10030	5	1	1	511	1
10105	4	2	2	422	1

CustomersId	Skor Recency	Skor Frequency	Skor Monetary	Skor RFM	*Cluster*
10135	4	1	1	411	1
10195	3	2	1	321	1

Pada table 3.16 merupakan contoh data yang termasuk kedalam *cluster* 1, dimana data pada cluster ini memiliki skor recency tertinggi yaitu 5, namun skor frequency dan monetary sangat rendah yaitu 1.

Tabel 3. 37 FCM *Cluster* 2

CustomersId	Skor Recency	Skor Frequency	Skor Monetary	Skor RFM	*Cluster*
10255	1	2	4	124	2
10405	1	2	4	124	2
10660	1	1	4	114	2
10705	2	1	4	214	2

Pada table 3.17 merupakan contoh data yang termasuk kedalam *cluster* 2, dimana pada cluster ini memiliki data dengan skor recency dan frequency yang rendah yaitu satu dan dua, sedangkan skor monetarynya tinggi yaitu empat.

Tabel 3. 48 FCM *Cluster* 3

CustomersId	Skor Recency	Skor Frequency	Skor Monetary	Skor RFM	*Cluster*
10240	1	1	2	112	3
10300	2	1	1	211	3
10315	1	1	1	111	3
10345	2	1	1	211	3

Pada table 3.18 merupakan contoh data yang termasuk kedalam *cluster* 3, dimana data pada cluster ini memiliki skor recency, antara satu dan dua sedangkan frequency dan monetary memiliki skor satu. Pada table 3.19 merupakan contoh data yang termasuk kedalam *cluster* 4, dimana data pada cluster ini memiliki skor recency antara dua sampai empat sedangkan skor frequency dan monetary antara empat sampai lima.

Tabel 3. 19 FCM *Cluster* 4

CustomersId	Skor Recency	Skor Frequency	Skor Monetary	Skor RFM	*Cluster*
10045	2	4	4	244	4
10075	4	5	4	454	4

CustomersId	Skor Recency	Skor Frequency	Skor Monetary	Skor RFM	*Cluster*
10090	4	4	5	445	4
10120	3	5	4	354	4

Tabel 3. 50 FCM *Cluster* 5

CustomersId	Skor Recency	Skor Frequency	Skor Monetary	Skor RFM	*Cluster*
10060	3	2	3	323	5
10150	4	1	3	413	5
10225	5	2	3	523	5
10375	5	2	3	523	5

Pada table 3.20 merupakan contoh data yang termasuk kedalam *cluster* 5, dimana data pada cluster ini memiliki skor recency antara tiga sampai lima, namun memiliki skor skor frequency yang rendah yaitu satu sampai dua, dan memiliki skor monetary tiga. Selanjutnya dilakukan penilaian tingkat validitas kluster yang dihasilkan dengan menggunakan silhouette Index, berikut merupakan perbandingan tingkat validitas kluster antara DBSCAN dan *Fuzzy C-Means*. Berdasarkan tabel 3.21, metode DBSCAN memiliki nilai SI terkecil pada *cluster* 1 dengan nilai 0.5909237206373249, sedangkan *cluster* yang lain memiliki nilai 1. Untuk nilai SI total dari metode DBSCAN adalah 0,918184744127465. Sedangkan metode *Fuzzy C-Means* memiliki nilai SI terkecil pada cluster 5 dengan nilai 0.3019606172004608 dan nilai SI terbesar pada cluster 4 dengan nilai 0.5584288019243665. sedangkan untuk nilai SI total dari *Fuzzy C-Means* adalah 0,387801709898368.

Tabel 3. 61 Nilai SI per-Kluster

Cluster	Nilai SI DBSCAN	Nilai SI FCM
1	0.5909237206373249	0.3483927305714975
2	1.0	0.3619443734506899
3	1.0	0.3331743318021693
4	1.0	0.5584288019243665
5	1.0	0.3019606172004608

KESIMPULAN

Berdasarkan pengujian dan evaluasi yang telah dilakukan oleh penulis, dapat diambil kesimpulan, diantaranya:

1. Telah berhasil dilakukan implementasi metode klustering DBSCAN dan juga *Fuzzy C-Means*. dari kedua klustering yang diimplementasikan, masing – masing metode menghasilkan perbandingan untuk metode *clustering*, diantaranya:
2. metode DBSCAN memiliki tingkat validitas kluster (SI total) yang lebih tinggi sebesar 0,918184744127465, sedangkan *Fuzzy C-Means* memiliki tingkat validitas kluster (SI total) yang lebih rendah yaitu 0,387801709898368
3. Metode DBSCAN memerlukan waktu eksekusi yang lebih cepat yaitu 4,05229 Detik, sedangkan *Fuzzy C-means* memerlukan waktu eksekusi yang lebih lama yaitu 291, 17141 Detik
4. Metode DBSCAN memiliki nilai SI terendah pada *cluster* 1 dengan nilai 0.5909237206373249 namun pada *cluster* yang lain memiliki nilai SI 1, sedangkan *Fuzzy C-Means*, memiliki nilia SI *cluster* tertinggi pada *cluster* 4 dengan nilai 0.5584288019243665 dan SI terendah adalah *cluster* 3 dengan nilai 0.3019606172004608
5. Metode DBSCAN menghasilkan 400 *noise* data, sedangkan *Fuzzy C-Means* berhasil mengelompokkan data kedalam *cluster* yang ditentukan.

DAFTAR PUSTAKA

[1] Ahmad, H. P., & Dang, S. (2015). Performance Evaluation of Clustering Algorithm Using different dataset. *international Journal of Advance Research in Computer Science and Management Studies*, 8.

[2] Charu, A. C. (2015). *Data Mining The TextBook.* Springer.

[3] Dey, A. (2016). Machine Learning Algorithms: A Review. *International Journal of Computer Science and Information Technologies, Vol. 7 (3)*, 6.

[4] Essinger, S. D., & Rosen, G. L. (2011). AN INTRODUCTION TO MACHINE LEARNING FOR STUDENTS IN SECONDARY EDUCATION. 6.

[5] Kassambara, a. (2017). *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning.* STHDA.

[6] Kodinariya, T. M., & Makwana, P. R. (20136). Review on determining number of Cluster in K-Means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*.

[7] Kusumadewi, S., & Purnomo, H. (2010). *Aplikasi Logika Fuzzy untuk Mendukung Keputusan.* Yogyakarta: Graha Ilmu.

[8] prasetyo, e. (2012). *Data Mining Konsep dan aplikasi menggunakan matlab.* yogyakarta: Andi.

[9] Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 13.

[10] Singh, T., & Mahajan, M. (2014). Performance Comparison of Fuzzy C Means with Respect to Other Clustering Algorithm. *International Journal of Advanced Research in Computer Science and Software Engineering*, 5.

[11] Sivarathri, S., & Govardhan, A. (2014). EXPERIMENTS ON HYPOTHESIS "FUZZY K-MEANS IS BETTER THAN K-MEANS FOR CLUSTERING". *International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.4, No.5*, 14.

[12] Subbalakshmi, C., Krishna, G. R., Rao, S. K., & Rao, P. V. (2014). A Method to Find Optimum Number of Clusters Based on Fuzzy Silhouette on Dynamic Data Set. *International Conference on Information and Communication Technologies (ICICT 2014)*, 8.

[13] Wei, J.-T., Lin, S.-Y., & Wu, H.-H. (2010). A review of the application of RFM model. *African Journal of Business Management Vol. 4(19)*, 9.