

VISUALISASI SEGMENTASI FITUR ATM MENGGUNAKAN ANALISIS RFM DAN ALGORITMA K-MEANS UNTUK MENDUKUNG EVALUASI FITUR ATM DI PT. BANK

Eka Ayu Kurniati, Rully A. Hendrawan, Retno A. Vinarti

Jurusan Sistem Informasi, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

Email : eraha@is.its.ac.id, zahra_17@is.its.ac.id

Abstrak - ATM merupakan salah satu layanan utama yang diberikan oleh PT. Bank BRISyariah. Untuk mempermudah layanan transaksi perbankan nasabah, PT. Bank BRISyariah telah mengoperasikan 413 mesin ATM di seluruh Indonesia dan dilengkapi dengan 261 fitur ATM. Jumlah transaksi ATM bisa mencapai ribuan tiap harinya. Dengan banyaknya transaksi ATM yang terjadi setiap hari, perusahaan merasa perlu adanya sebuah evaluasi untuk mengetahui fitur ATM mana saja yang paling banyak digunakan dan memberikan *fee-based income* kepada perusahaan. Pendekatan yang sesuai untuk membantu evaluasi fitur ATM yaitu dengan mengelompokkan fitur ATM berdasarkan karakteristik penggunaannya. Untuk itu diperlukan sebuah segmentasi terhadap fitur ATM untuk mengetahui fitur ATM mana saja yang berharga bagi perusahaan. Penelitian ini bertujuan untuk membuat segmentasi fitur ATM yang ada pada PT. Bank BRISyariah dengan teknik Data Mining yaitu *clustering*. Metode yang digunakan dalam proses *clustering* adalah metode Ward dan algoritma K-Means, sedangkan atribut *clustering* didapatkan dari analisis model RFM. Validasi untuk hasil *clustering* menggunakan indeks Dunn dan uji ANOVA yaitu indeks R-Squared & uji F. Selain itu, dibangun juga visualisasi dari hasil *clustering* untuk mempermudah pemahaman akan segmen fitur ATM yang terbentuk.

Kata Kunci : Algoritma K-Means, Analisis RFM, Data Mining, Indeks Dunn, Indeks R-Squared, Segmentasi, Ward.

I. PENDAHULUAN

Banyak perusahaan yang beranggapan bahwa untuk segmentasi pelanggan merupakan salah satu kunci utama untuk memahami keadaan pasar mereka. Dengan membangun hubungan yang baik dengan pelanggan, mereka lebih berpeluang untuk memenangkan persaingan pasar. Namun segmentasi pelanggan bukan merupakan satu-satunya cara untuk lebih memahami pasar. Bentuk lain dari segmentasi yaitu segmentasi produk atau layanan ikut berperan penting untuk mencapai kesuksesan dalam pemasaran [1].

Segmentasi adalah salah satu cara dalam mengetahui bagaimana tanggapan pasar terhadap produk, apakah produk tersebut laku dan memberikan keuntungan atau justru tanggapan akan produk tersebut tidak sebanding dengan usaha marketing yang telah dilakukan [2]. Teknik-teknik dalam *Data Mining* sering digunakan dalam mencari segmentasi suatu data. Teknik yang paling populer dalam segmentasi data adalah *clustering*. *Clustering* merupakan teknik untuk mengelompokkan objek bersama berdasarkan perbedaan dan kesamaan pada setiap objek dan membuat homogenitas tinggi pada *cluster* yang sama atau heterogenitas tinggi setiap kelompok [3].

Dalam penelitian ini, sebelum memasuki proses *clustering* akan dilakukan analisis model RFM (*Recency*,

Frequency, & *Monetary*) terlebih dahulu. Data yang digunakan dalam penelitian ini merupakan data transaksi ATM yang berisi catatan waktu transaksi dilakukan, jenis fitur ATM yang digunakan, besar uang dalam transaksi serta beberapa keterangan lainnya. Dari data transaksi ATM tersebut dapat dilakukan analisis rentang transaksi terakhir dan transaksi selanjutnya, berapa kali fitur ATM telah digunakan dan berapa jumlah uang yang ditransaksikan, sehingga penggunaan model RFM dalam penelitian ini sangat mendukung analisis terhadap data transaksi ATM.

Sedangkan untuk proses *clustering*, pada *clustering* di layanan bank [4] menggunakan algoritma K-Means dimana sebelumnya dilakukan analisis RFM untuk mendapatkan atribut-atribut *clustering*, selain itu perhitungan indeks R-Squared dipergunakan untuk menguji validitas dari hasil *cluster* algoritma K-Means.

PT. Bank BRISyariah merupakan bank syariah yang mengutamakan kenyamanan layanan nasabah melalui pemanfaatan perangkat teknologi informasi. Salah satu layanan berbasis teknologi informasi yang diberikan PT. Bank BRISyariah adalah ATM yang dilengkapi dengan berbagai fitur transaksi perbankan seperti penarikan tunai, transfer ke rekening BRISyariah maupun bank lainnya, pembayaran tagihan (telepon, listrik, internet dll), pembelian pulsa (telepon seluler atau token listrik) serta pembayaran zakat, infaq, shadaqah, wakaf dan qurban [5].

Terhitung sejak akhir 2013, PT. Bank BRISyariah telah mengoperasikan 413 mesin ATM di seluruh Indonesia [6]. Selain itu layanan ATM dilengkapi dengan 261 fitur dengan berbagai fungsi yang memudahkan aktifitas perbankan nasabah. Banyaknya data transaksi ATM tiap hari dari berbagai lokasi dan fitur ATM yang dimiliki, perusahaan merasa perlu adanya evaluasi fitur ATM. Evaluasi tersebut ditujukan untuk mengetahui fitur ATM mana saja sering digunakan nasabah dan memberikan *fee based income* yang tinggi bagi perusahaan. Pendekatan yang sesuai dan efektif untuk mengevaluasi fitur ATM adalah segmentasi produk [1].

Segmentasi produk dilakukan dengan cara membuat segmentasi terhadap fitur ATM berdasarkan data-data transaksi ATM dan data fitur ATM sendiri. Segmentasi fitur ATM berguna untuk mengidentifikasi fitur ATM yang berharga bagi perusahaan. Selain itu segmentasi fitur ATM dapat membantu dalam inovasi dan perbaikan layanan ATM yang diberikan.

Tugas akhir ini bertujuan untuk melakukan segmentasi fitur ATM berdasarkan data pada studi kasus nyata PT. Bank BRISyariah yang terdiri dari data transaksi ATM dan data fitur ATM. Diharapkan tugas akhir ini dapat memberikan kontribusi terhadap peningkatan layanan ATM yang diberikan oleh PT. Bank BRISyariah.

II. TINJAUAN PUSTAKA

A. Profil PT. Bank BRISyariah

PT. Bank BRISyariah adalah akuisisi PT. Bank Rakyat Indonesia (Persero), Tbk. terhadap Bank Jasa Arta pada 19 Desember 2007 dan pada tanggal 17 November 2008 PT. Bank BRISyariah secara resmi beroperasi. Tahun 2013, PT. Bank BRISyariah telah memiliki berbagai jenis jaringan layanan yang tersebar diseluruh Indonesia. Jaringan layanan tersebut terdiri dari 52 kantor cabang, 199 Kantor Cabang Pembantu, 8 Kantor Kas, 573 Kantor Layanan Syariah [6]. Selain itu jumlah mesin ATM di seluruh Indonesia telah mencapai 418 unit mesin. Saat ini jumlah fitur ATM yang dimiliki PT. Bank BRISyariah telah mencapai 261 macam fitur yang didukung dengan kerjasama pihak ketiga seperti ATM Prima, ATM Bersama, ATM BRI, Telkom, PLN dsb.

B. Data Mining

Data Mining adalah sebuah proses yang menggunakan statistika, matematika, dan kecerdasan buatan untuk mengekstrak dan mengidentifikasi informasi dan *knowledge* (pola-pola) yang berasal dari sekumpulan data yang sangat besar [7]. Data bisa berupa data spasial, data multimedia, data *time series*, data teks, dan data web [8].

C. Clustering

Clustering merupakan sebuah teknik yang berguna untuk menemukan *knowledge* dari kumpulan data [3]. Teknik ini adalah sebuah metode penjelajah guna membantu pemecahan masalah klasifikasi. *Clustering* digunakan saat dalam data struktur kategori sedikit atau bahkan tidak diketahui sama sekali. Tujuan dari teknik ini yaitu mengurutkan sampel kasus kedalam kelompok seperti anggota yang memiliki tingkat asosiasi tinggi berada dalam kelompok yang sama, sedangkan anggota yang memiliki tingkat asosiasi rendah berada dalam kelompok lain.

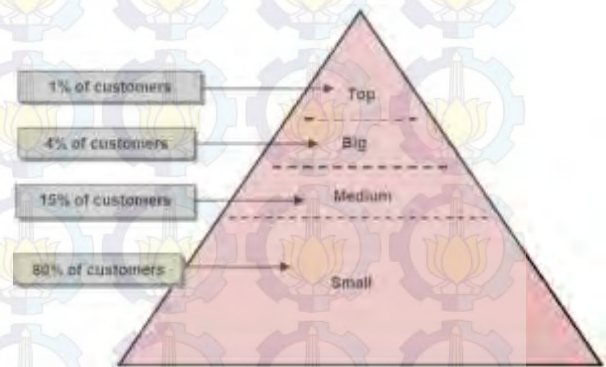
D. Analisis Model RFM

Analisis RFM merupakan teknik pemasaran yang digunakan analisis kebiasaan pelanggan seperti kapan terakhir kali transaksi (*recency*), seberapa sering transaksi dilakukan (*frequency*), dan berapa biaya yang untuk transaksi (*monetary*) [15]. Dari deskripsi nilai RFM diatas, nilai RFM yang dipakai dalam penelitian ini disesuaikan dengan jenis data yang dimiliki.

Berikut detail definisi model RFM yang dipakai dalam penelitian ini:

- *Recency* dari transaksi terakhir (R)
R mewakili dari *recency*, dimana menunjukkan rentang hari dari tanggal transaksi terakhir terjadi hingga tanggal evaluasi dilakukan.
- *Frequency* dari transaksi (F)
F mewakili dari *frequency*, dimana menunjukkan berapa kali fitur ATM digunakan untuk transaksi dalam periode tertentu.
- *Monetary* dari transaksi (M)
M mewakili nilai *monetary*, dimana nilai tersebut didapatkan dari rata-rata besar (uang) transaksi fitur ATM dalam satuan IDR.

Selain itu Jay Curry memperkenalkan sistem pengelompokan nilai RFM berdasarkan masing-masing nilai atribut RFM yang dikenal dengan piramida pelanggan. Berikut merupakan piramida pelanggan yang diperkenalkan oleh Jay Curry, Gambar 2:



Gambar 1 RFM Pyramid Model

Piramida tersebut terdiri dari 4 kategori yaitu :

1. "Top" Customers, merupakan 1% pelanggan aktif teratas dalam nilai RFM.
2. "Big" Customers, merupakan 4% selanjutnya dari pelanggan paling aktif dalam nilai RFM.
3. "Medium" Customers, merupakan 15% selanjutnya dari pelanggan aktif dalam nilai RFM.
4. "Small" Customers, merupakan 80% sisanya dari pelanggan aktif dalam nilai RFM. [16].

E. Metode Ward

Metode Ward adalah kriteria yang diaplikasikan pada analisis *clustering* hirarki. Metode ini diperkenalkan oleh Joe H. Ward, Jr sebagai sebuah kasus khusus dari pendekatan fungsi tujuan [17]. Tujuan dari metode Ward adalah untuk menghasilkan *cluster* yang homogen dan hirarki simetris dan definisi dari pusat *cluster* gravitasi yang merupakan cara mewakili sebuah *cluster*.

F. Algoritma K-Means

K-Means adalah salah satu algoritma yang paling populer untuk *clustering*, awalnya algoritma ini dikenal sebagai metode Forgy [18].

Proses *clustering* dalam algoritma K-Means dimulai dengan mengidentifikasi data yang akan dikelompokkan, $X_{ij} (i = 1, \dots, n; j = 1, \dots, m)$ dengan n adalah jumlah data yang akan dikelompokkan dan m adalah jumlah variable. Pada awal iterasi. Pusat setiap *cluster* yaitu k , ditentukan secara bebas atau sembarang, $C_{kj} (k = 1, \dots, k; j = 1, \dots, m)$. Kemudian dihitung jarak antara setiap data dengan setiap pusat *cluster*. Perhitungan jarak tersebut dapat menggunakan formula *Euclidean* seperti pada persamaan (1), yaitu:

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \quad (1)$$

Dimana :

d_{ik} = jarak antara data ke- i (x_{ij}) dan pusat *cluster* ke- k (c_{kj})

Jarak ke pusat *cluster* yang paling minimal dapat dihitung dengan persamaan (2), yaitu:

$$\text{Min } \sum_{k=1}^k d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \quad (2)$$

Selanjutnya nilai pusat *cluster* yang baru dapat dihitung dengan persamaan (3), yaitu:

$$c_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p} \quad (3)$$

Dimana :

$x_{ij} \in \text{cluster ke } k$

p = banyaknya anggota *cluster* ke - k

G. Indeks Dunn (DI)

Nilai indeks Dunn efektif untuk menghitung indeks validitas sebuah *cluster* [18]. Untuk menghitung indeks Dunn dapat menggunakan persamaan (4) seperti dibawah ini :

$$D = \min_{i=1 \dots n_c} \left\{ \min_{j=i+1 \dots n_c} \left(\frac{d(c_i, c_j)}{\min_{k=1 \dots n_c} (\text{diam}(c_k))} \right) \right\} \quad (4)$$

Dimana $d(c_i, c_j)$ dihitung dari persamaan (5) dan $\text{diam}(c_k)$ dihitung dari persamaan (6).

$$d(c_i, c_j) = \min_{x \in c_i} \{d(x, y)\} \quad (5)$$

$$\text{diam}(c_i) = \max_{x, y \in c_i} \{d(x, y)\} \quad (6)$$

Keterangan :

D = indeks Dunn

n_c = jumlah *cluster*

$d(c_i, c_j)$ = jarak antar *cluster* i dan *cluster* j

$\text{diam}(c_i)$ = diameter *cluster* i

H. Uji ANOVA

1) Indeks R-Squared

R-Squared digunakan untuk menghitung perbedaan pada *cluster*, seperti yang dideskripsikan pada persamaan (7), yaitu:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{SS_{res}}{SS_{res} + SS_{reg}} \quad (7)$$

Dimana:

SS_{res} = kuadrat total residual (*residual sum of squares*)

SS_{reg} = kuadrat total karena regresi (*explained sum of squares*)

SS_{tot} = jumlah seluruh kuadrat total pada data

Untuk mencari kuadrat total dalam kelompok dapat menggunakan persamaan (8), yaitu:

$$SS_{res} = \sum_{i=1}^n (y_i - f(x_i))^2 \quad (8)$$

Dimana :

y_i = nilai ke- i dari variabel yang diprediksi

x_i = nilai ke- i dari variable penjelas

$f(x_i)$ = nilai yang diprediksi y_i

Sedangkan untuk mencari jumlah seluruh kuadrat total pada data menggunakan persamaan (9), yaitu :

$$SS_{reg} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (10)$$

Dimana :

$\hat{y}_i$ = nilai yang diprediksi ke- i dari variabel respon

$\bar{y}$ = rata-rata variable respon

[4]

I. Uji F

Uji F sering digunakan untuk membandingkan model statistik yang telah disesuaikan dengan kumpulan data (*dataset*) yang tujuannya untuk mengidentifikasi model mana yang paling sesuai dengan populasi sampel data. Berikut merupakan persamaan. Uji F biasanya disertakan dalam tabel ANOVA yang dimana diformulasikan sebagai berikut;

$$F = \frac{\text{perbedaan antar kelompok}}{\text{perbedaan dalam kelompok}} \quad (11)$$

$$\text{Perbedaan antar kelompok} = \frac{\sum_i n_i (\bar{Y}_i - \bar{Y})^2}{(K-1)} \quad (12)$$

$$\text{Perbedaan dalam kelompok} = \frac{\sum_i n_i (Y_{ij} - \bar{Y}_i)^2}{(N-K)} \quad (13)$$

Dimana :

$\bar{Y}_i$ = rata-rata kelompok ke- i

$\bar{Y}$ = rata-rata seluruh kelompok

K = jumlah kelompok

Y_{ij} = observasi ke- j

i = kelompok ke- i dari K

N = ukuran sampel data

III. METODE PENELITIAN

Berikut ini adalah tahapan dalam proses pengerjaan tugas akhir. Alur pengerjaan penelitian ini dapat dilihat pada Gambar 3.

A. Data Mentah

Data mentah yang dipakai dalam proses segmentasi fitur ATM ini merupakan sampel data transaksi ATM yang diambil dari tanggal 11 Oktober 2014 hingga 04 Desember 2014.

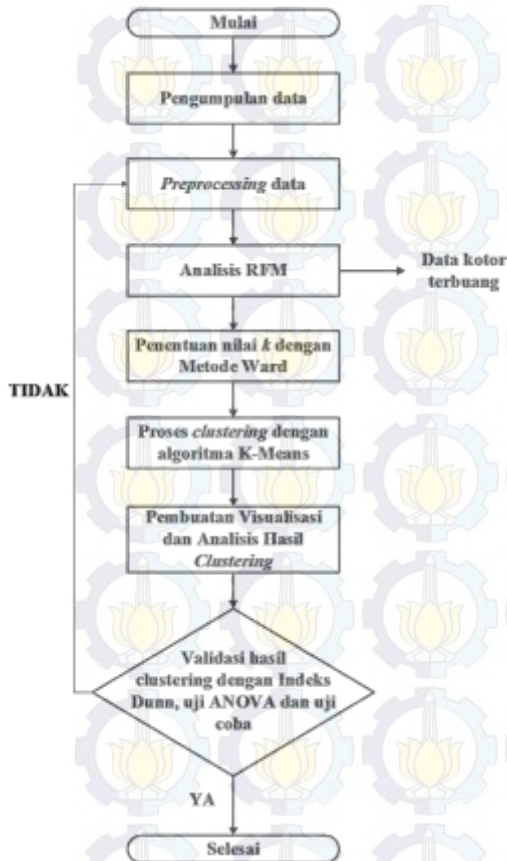
B. Data Preprocessing

Data *preprocessing* merupakan tahapan penyesuaian data sebelum memasuki proses *clustering*. Beberapa

proses yang dilakukan pada data *preprocessing* yaitu data *extraction*, data *cleaning*, dan data *transformation*.

1) Data Integration

Integrasi data diperlukan untuk mempermudah pembacaan dan pengelolaan data nantinya dalam proses *clustering*. Pada tahap ini data akan disimpan kedalam database MySQL.



Gambar 2 Alur Pengerjaan Tugas Akhir

2) Data Cleaning

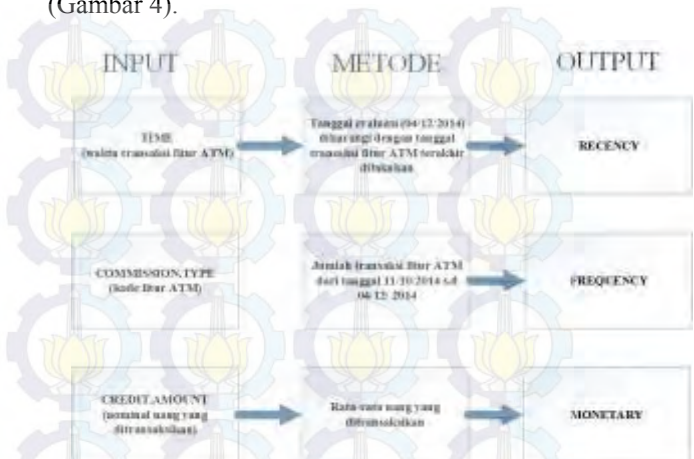
Data *cleaning* merupakan proses yang bertujuan untuk membersihkan tabel data dari item data yang mempunyai nilai kosong atau *NULL*.

Tabel 1 Potongan Hasil Analisis RFM

KODE FITUR	RECENCY	FREQUENCY	MONETARY
STREDCMIKRO	1	124	1397322.581
APAYUNISBA	52	2	3200000
ATMBDELIMA	1	141	2645804.865
ATMBERSTRF	1	1835	1308310.649
ATMBERSWDL	1	4538	395013.2217
ATMBEXMON	1	53	3706286.66
ATMBRISTRF	1	376	804383.2952
ATMBRISWDL	1	3480	430545.977
ATMBRITRF	1	648	909871.7145
ATMBRIWDL	1	6994	450986.5599
.....			

3) Data Transformation

Berikut merupakan alur proses transformasi data (Gambar 4).



Gambar 3 Alur Proses Transformasi Data

C. Analisis Model RFM

Pada tahap ini dilakukan analisis nilai masing-masing atribut RFM dari rentang transaksi terakhir dan transaksi selanjutnya, berapa kali fitur ATM telah digunakan dan berapa jumlah uang yang ditransaksikan pada data transaksi ATM yang berisi catatan waktu transaksi dilakukan, jenis fitur ATM yang digunakan, besar uang dalam transaksi. Potongan hasil akhir analisis model RFM merupakan hasil transformasi data yang dilakukan pada bab sebelumnya, Tabel 1.

1) RFM Scoring

RFM *Scoring* dilakukan untuk menentukan *score* yang mewakili keanggotaan fitur ATM pada masing-masing faktor RFM. Dimana dihasilkan nilai RFM yang didapatkan dari total penjumlahan perkalian tiap *score* atribut RFM dengan bobot masing-masing. Pada Tabel 2 menyajikan potongan hasil akhir dari RFM *Scoring*. Nilai RFM yang dihasilkan pada tahap ini digunakan dalam membuat kategori piramida RFM. Rumus nilai RFM adalah sebagai berikut :

$$RFM = (recency * bobot\ Recency) + (frequency * bobot\ Frequency) + (monetary * bobot\ Monetary)$$

Tabel 2 Potongan Hasil RFM Scoring

COMMISSION.TYPE	Recency_score	Frequency_score	Monetary_score	RFM_score
STREDCMIKRO	5	1	3	4.4
APAYUNISBA	1	2	4	3.7
ATMBDELIMA	5	2	2	5.1
ATMBERSTRF	5	4	4	7.7
ATMBERSWDRL	5	5	2	8.1
ATMBEXMON	3	4	4	6.7
ATMBRISTRF	4	5	2	7.6
ATMBRISWDRL	5	4	2	7.1
ATMBRITRF	5	3	4	6.7
ATMBRIWDL	5	5	4	8.7

D. Penentuan Jumlah Cluster dengan Metode Ward

Dari hasil analisa SPSS, dihasilkan tabel *Agglomeration Schedule* yang menunjukkan solusi untuk setiap jumlah *cluster* (k) sesuai dengan jumlah data (*case*) yang diinputkan. Jumlah *cluster* yang optimal dapat ditentukan dengan melihat loncatan nilai koefisien yang besar. Pada Tabel 3, perbedaan koefisien pada *stage* 41 dan *stage* 42 adalah sebesar 35384155635295.80. Nilai perbedaan koefisien tersebut cukup besar dibanding dengan nilai perbedaan koefisien sebelumnya, maka dapat diperoleh kesimpulan bahwa jumlah *cluster* (k)

yang sesuai untuk proses *clustering* yaitu berjumlah 3 *cluster*.

E. Proses Clustering dengan Algoritma K-Means

Dari proses sebelumnya yaitu penentuan jumlah *cluster* dengan metode Ward telah didapatkan bahwa jumlah *cluster* yang sesuai adalah 3 *cluster*. Jumlah *cluster* tersebut akan dijadikan sebagai nilai k pada proses *clustering* dengan algoritma K-Means.

Pada Tabel 4 menyajikan potongan hasil *clustering* menggunakan algoritma K-Means, dari tabel tersebut dapat dilihat anggota dari *cluster* 1, 2 dan 3 selain itu terdapat juga kolom jarak *cluster* (*Distance*).

Tabel 3 Potongan Tabel Agglomeration Schedule SPSS

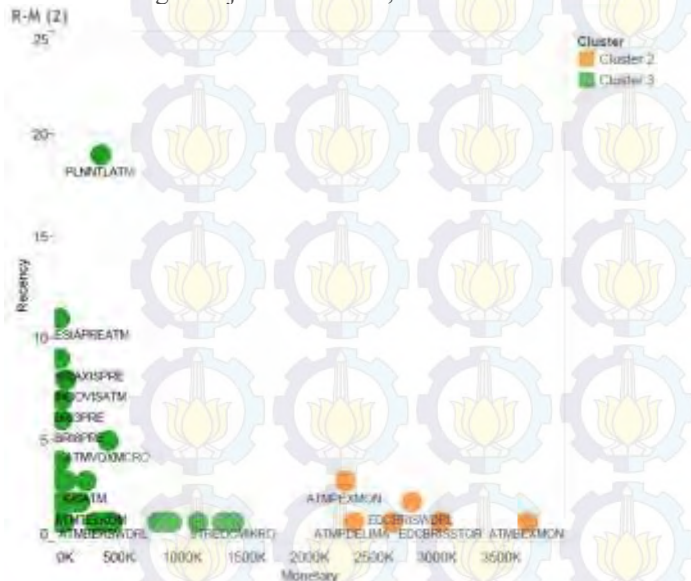
Stage	Number of Cluster	Coefficients	Coefficients Difference
37	7	1059126299058.38	495542860992.17
38	6	1615540838775.45	556414539717.07
39	5	2460249997902.93	844709159127.48
40	4	6656933891236.03	419668389333.10
41	3	10904995594393.80	4248061703157.80
42	2	46289151229689.60	35384155635295.80
43	1	135327897251708.00	89038746022018.20

Tabel 4 Potongan Hasil Clustering Algoritma K-Means

Case Number	KODE FITUR	Cluster	Distance
1	STREDCMIKRO	3	1092218.421
2	APAYUNISBA	2	69292.812
3	ATMBDELIMA	2	484902.383
4	ATMBERSTRF	3	1003207.058
5	ATMBERSWDRL	3	89992.867
6	ATMBEXMON	2	575579.417
7	ATMBRISTRF	3	499279.084
8	ATMBRISWDRL	3	125473.559
9	ATMBRITRF	3	604767.427

F. Visualisasi Hasil Clustering & Analisis Hasil Clustering

Berikut merupakan diagram *scatter plot* dengan *Monetary* sebagai sumbu x dan *Recency* sebagai sumbu y setelah dibagi menjadi 2 kwadran, Gambar 5 :

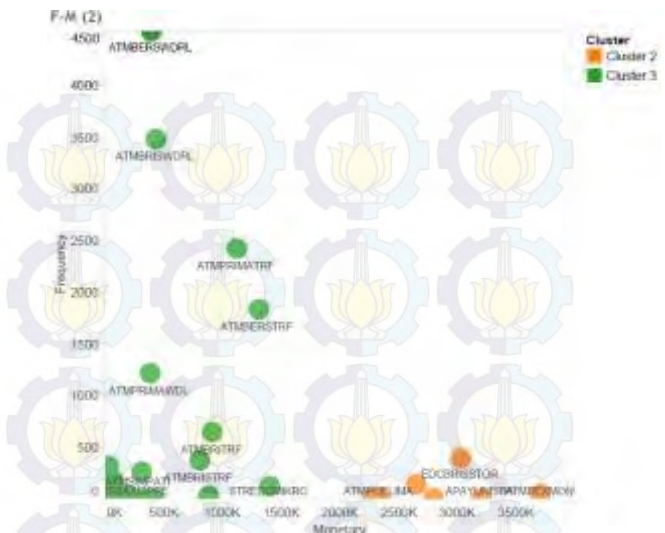


Gambar 4 Analisis Diagram Scatter Plot Monetary dan Recency

Berikut merupakan hasil analisis *cluster* dilihat dari sisi *Monetary* dan *Recency* berdasarkan Gambar 5, Tabel 8:

Tabel 5 Hasil Analisis Diagram Scatter Plot Monetary dan Recency

No.	Hasil Analisis
1.	<i>Cluster 3</i> mempunyai rentang nilai <i>Monetary</i> 0 s.d 1.400.000 dimana tergolong tingkat Monetary sedang yang menunjukkan bahwa fitur ATM yang menjadi anggota <i>cluster 3</i> memberikan <i>fee-based income</i> yang cukup untuk perusahaan.
2.	<i>Cluster 3</i> mempunyai rentang nilai <i>Recency</i> 0 s.d 19 dimana tergolong tingkat Recency tinggi yang menunjukkan bahwa fitur ATM yang menjadi anggota <i>cluster 3</i> digunakan terakhir kali tidak lebih dari 1 bulan terhitung sejak tanggal evaluasi yaitu 04 Desember 2014.
3.	<i>Cluster 2</i> mempunyai rentang nilai <i>Monetary</i> 2.200.000 s.d 3.800.000 dimana tergolong tingkat Monetary tinggi yang menunjukkan bahwa fitur ATM yang menjadi anggota <i>cluster 2</i> memberikan <i>fee-based income</i> tinggi untuk perusahaan.
4.	<i>Cluster 2</i> mempunyai rentang nilai <i>Recency</i> 0 s.d 3 dimana tergolong tingkat Recency sangat tinggi yang menunjukkan bahwa fitur ATM yang menjadi anggota <i>cluster 2</i> digunakan terakhir kali tidak lebih dari 3 hari terhitung sejak tanggal evaluasi yaitu 04 Desember 2014.



Gambar 5 Analisis Diagram Scatter Plot Monetary dan Frequency

Selanjutnya merupakan hasil analisis untuk diagram *scatter plot* dengan *Monetary* sebagai sumbu x dan *Frequency* sebagai sumbu y. Berikut hasil analisis dari segi *Monetary* & *Frequency* yang didapatkan dari Gambar 6, Tabel 9 :

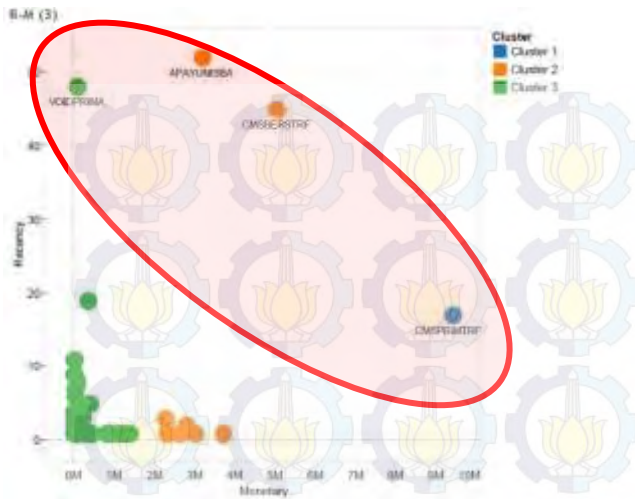
Tabel 6 Hasil Analisis Diagram Scatter Plot Monetary dan Frequency

No.	Hasil Analisis
1.	<i>Cluster 3</i> mempunyai rentang nilai <i>Monetary</i> 0 s.d 1.400.000 dimana tergolong tingkat Monetary sedang yang menunjukkan bahwa fitur ATM yang menjadi anggota <i>cluster 3</i> memberikan <i>fee-based income</i> yang cukup untuk perusahaan.
2.	<i>Cluster 3</i> mempunyai rentang nilai <i>Frequency</i> 0 s.d 4500 dimana tergolong tingkat Frequency sangat tinggi yang menunjukkan bahwa fitur ATM yang menjadi anggota <i>cluster 3</i> sangat sering digunakan untuk transaksi ATM.
3.	<i>Cluster 2</i> mempunyai rentang nilai <i>Monetary</i> 2.200.000 s.d 3.800.000 dimana tergolong tingkat Monetary tinggi yang menunjukkan bahwa fitur ATM yang menjadi anggota <i>cluster 2</i> memberikan <i>fee-based income</i> yang tinggi pada perusahaan.
4.	<i>Cluster 2</i> mempunyai rentang nilai <i>Frequency</i> 0 s.d 400 dimana tergolong tingkat Frequency sedang yang menunjukkan bahwa fitur ATM yang menjadi anggota <i>cluster 2</i> cukup sering digunakan untuk transaksi ATM.

Untuk diagram *scatter plot* dengan *Frequency* sebagai sumbu x dan *Recency* sebagai sumbu y tidak dilakukan analisis karena pola persebaran *cluster* yang tidak bisa memberikan informasi.

1) Analisis Data Outliers Hasil Clustering

Data *outliers* pada diagram ditandai dengan lingkaran merah. Analisis pertama akan dilakukan untuk diagram *scatter plot* (Gambar 7) dimana *Monetary* sebagai sumbu x dan *Recency* sebagai sumbu y.



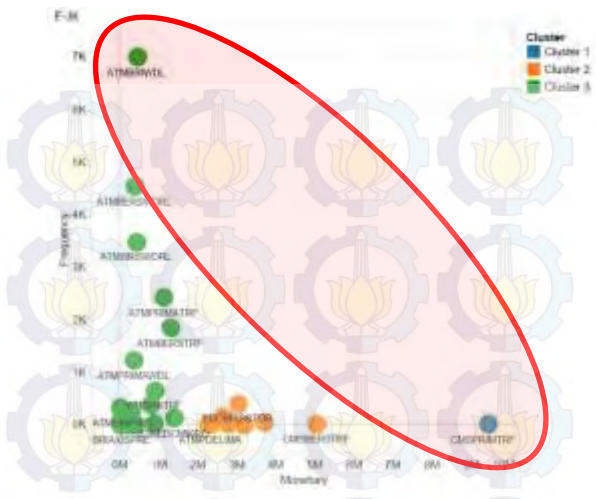
Gambar 6 Analisis Data Outliers Diagram Scatter Plot Monetary & Recency

Dari Gambar 7 diperoleh beberapa hasil analisis yang dijelaskan pada Tabel yaitu :

Tabel 7 Hasil Analisis Data Outliers Diagram Scatter Plot Monetary & Recency

No.	Hasil Analisis
1.	Cluster 1 mempunyai nilai <i>Monetary</i> sebesar 9.400.000 dimana tergolong tingkat Monetary sangat tinggi yang menunjukkan bahwa fitur ATM yang menjadi anggota cluster 1 memberikan <i>fee-based income</i> yang tinggi pada perusahaan.
2.	Fitur ATM anggota cluster 2 yaitu APAYUNISBA & CMSBERSTRF mempunyai nilai rentang <i>Monetary</i> 3.200.000 s.d 5.100.000 dimana tergolong tingkat Monetary tinggi yang menunjukkan bahwa fitur ATM tersebut memberikan <i>fee-based income</i> yang cukup tinggi pada perusahaan
3.	Fitur ATM anggota cluster 3 yaitu VOIDPRIMA mempunyai nilai <i>Monetary</i> sebesar 100.000 dimana tergolong tingkat Monetary rendah yang menunjukkan bahwa fitur ATM tersebut memberikan <i>fee-based income</i> yang rendah pada perusahaan.
4.	Cluster 1 mempunyai nilai <i>Recency</i> sebesar 17 dimana tergolong tingkat Recency tinggi yang menunjukkan bahwa anggota cluster 1 yaitu CMSPRIMTRF digunakan tidak lebih dari 1 bulan terhitung dari tanggal evaluasi yaitu 04 Desember 2014.
5.	Fitur ATM anggota cluster 2 yaitu APAYUNISBA & CMSBERSTRF mempunyai nilai rentang <i>Recency</i> 45 s.d 52 dimana tergolong tingkat Recency sedang yang menunjukkan bahwa fitur ATM tersebut digunakan tidak lebih dari 2 bulan terhitung dari tanggal evaluasi yaitu 04 Desember 2014
6.	Fitur ATM anggota cluster 3 yaitu VOIDPRIMA mempunyai nilai <i>Recency</i> sebesar 48 dimana tergolong tingkat Recency sedang yang menunjukkan bahwa fitur ATM tersebut digunakan tidak lebih dari 2 bulan terhitung dari tanggal evaluasi yaitu 04 Desember 2014.

Selanjutnya merupakan analisis kedua yang dilakukan pada diagram *scatter plot* (Gambar 8) dengan *Monetary* sebagai sumbu x dan *Frequency* sebagai sumbu y.



Gambar 7 Analisis Data Outliers Diagram Scatter Plot Monetary & Frequency

Dari diagram diatas diperoleh beberapa hasil analisis yang dijelaskan pada Tabel 11 yaitu :

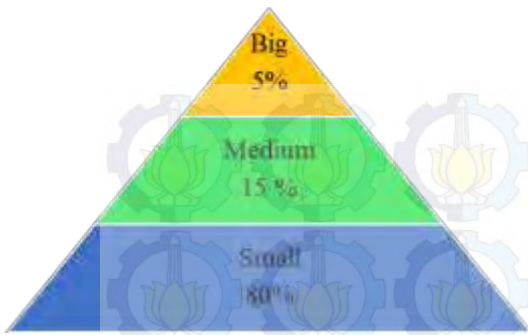
Tabel 8 Analisis Data Outliers Scatter Plot Monetary & Frequency

No.	Hasil Analisis
1.	Cluster 1 mempunyai nilai <i>Monetary</i> sebesar 9.400.000 dimana tergolong tingkat Monetary sangat tinggi yang menunjukkan bahwa fitur ATM yang menjadi anggota cluster 1 memberikan <i>fee-based income</i> yang tinggi pada perusahaan.
2.	Fitur ATM anggota cluster 3 yaitu ATMBRIWDL mempunyai nilai <i>Monetary</i> sebesar 451.000 dimana tergolong tingkat Monetary rendah yang menunjukkan bahwa fitur ATM tersebut memberikan <i>fee-based income</i> yang kecil pada perusahaan.
3.	Cluster 1 mempunyai nilai <i>Frequency</i> sebesar 6 dimana tergolong tingkat Frequency sangat rendah yang menunjukkan bahwa fitur ATM yang menjadi anggota cluster 1 sangat jarang digunakan untuk transaksi ATM.
4.	Fitur ATM anggota cluster 3 yaitu ATMBRIWDL mempunyai nilai <i>Frequency</i> sebesar 7.000 dimana tergolong tingkat Frequency sangat tinggi yang menunjukkan bahwa fitur tersebut sangat sering digunakan untuk transaksi ATM.

2) Analisis Piramida RFM

Dari hasil nilai RFM sebelumnya yang didapatkan pada sub-bab RFM *Scoring* diketahui bahwa nilai RFM terbesar adalah 8.7 dan nilai RFM terkecil 3.1. Setelah dilakukan pembagian persentase keanggotaan piramida menurut Jay Curry didapatkan piramida untuk fitur ATM seperti gambar berikut, Gambar 9.

Piramida pada Gambar 9 menunjukkan persebaran fitur ATM pada masing-masing kategori piramida fitur ATM. Semakin tinggi nilai RFM yang dipunyai, semakin kecil pula jumlah persentase anggota pada piramida. Dan sebaliknya semakin kecil nilai RFM maka semakin besar pula jumlah persentase anggota pada piramida.



Gambar 8 Piramida Fitur ATM Berdasarkan Nilai RFM

Daftar fitur ATM yang menjadi anggota masing-masing kategori piramida ATM dapat dilihat pada Tabel 12. Pada Tabel 12, diketahui bahwa terdapat 2 fitur ATM yang

menjadi anggota kategori “Big” dengan nilai RFM diatas 8 yang menunjukkan bahwa kedua fitur ATM tersebut yaitu ATMBRIWDL dan ATMBERSWDRL banyak memberikan keuntungan bagi perusahaan melalui layanan transaksi ATM. Sedangkan untuk kategori “Medium” dengan nilai RFM antara 7 s.d < 8 terdapat 7 fitur ATM yang menjadi anggotanya yaitu ATMPRIMATRF, ATMBRISTRF, PLNPRESMSB, XLPREATM, ATMPEXMON, PLNPSTSMSB, dan CMSBERSTRF.

Ketujuh fitur tersebut menunjukkan bahwa cukup memberikan keuntungan besar untuk perusahaan melalui layanan transaksi ATM. Dan 32 fitur ATM lainnya dengan nilai RFM < 7 menjadi anggota kategori piramida “Small” dimana fitur-fitur ATM tersebut memberikan keuntungan kecil bagi perusahaan melalaui layanan transaksi ATM.

Tabel 9 Potongan Daftar Fitur ATM Anggota Kategori Piramida

COMMISSION.TYPE	RFM_score	Pyramid_category
ATMBRIWDL	8.7	Big
ATMBERSWDRL	8.2	Big
ATMPRIMATRF	7.9	Medium
ATMBRISTRF	7.7	Medium
PLNPRESMSB	7.7	Medium
XLPREATM	7.6	Medium
ATMPEXMON	7.2	Medium
PLNPSTSMSB	7.2	Medium
CMSBERSTRF	7.1	Medium
ATMBERSTRF	6.8	Small
ATMBRISWDRL	6.7	Small
DEBITPRIMA	6.7	Small

3) USULAN STRATEGI

Berikut merupakan usulan strategi untuk penanganan masing-masing segmen, Tabel 6.6. Usulan strategi yang Tabel 6.10 Usulan Strategi Fitur ATM

diberikan berdasarkan hasil wawancara dengan pihak bank.

No.	Segmen	Karakteristik	Usulan Strategi	Detail Strategi
1	Segmen 1	- Recency tinggi - Frequency sangat rendah - Monetary sangat tinggi	- Menambah aktivitas promosi kepada nasabah yang bertujuan untuk meningkatkan frekuensi transaksi. - Mensosialisasikan kepada mitra tentang adanya sistem <i>reward</i> jika tingkat penggunaan fitur ATM mencapai angka tertentu. - Menjaga ketersediaan uang dalam mesin ATM selalu cukup untuk melayani transaksi ATM nasabah guna mengantisipasi adanya transaksi ATM dengan nominal yang besar.	Strategi 1: - Menggunakan media sosial dan media eletronik lainnya sebagai sarana promosi fitur ATM. - Memberikan tawaran menarik untuk nasabah jika menggunakan fitur ATM tersebut. - Sosialisasi kepada nasabah tentang manfaat penggunaan fitur ATM melalui <i>customer service</i>. Strategi 2: - Mencantumkan sistem <i>reward</i> pada dokumen MOU kerjasama dengan mitra. - Menjalin komunikasi yang erat dengan mitra guna mengetahui usaha mitra dalam memenuhi sistem <i>reward</i>. Strategi 3: - Memonitor secara rutin jumlah uang dalam mesin ATM dan

No.	Segmen	Karakteristik	Usulan Strategi	Detail Strategi
				nominal transaksi pelanggan. 2. Membuat peramalan tentang jadwal pengisian mesin ATM berdasarkan data waktu dan nominal transaksi ATM, serta jadwal pengisian mesin ATM terdahulu.
2	Segmen 2	• <i>Recency</i> sangat tinggi • <i>Frequency</i> sedang • <i>Monetary</i> tinggi	• Mengembangkan aktivitas promosi kepada nasabah yang bertujuan untuk meningkatkan frekuensi transaksi. • Menjaga ketersediaan uang dalam mesin ATM selalu cukup untuk melayani transaksi ATM nasabah guna mengantisipasi adanya transaksi ATM dengan nominal yang besar.	Strategi 1: 1. Menggunakan media sosial sebagai sarana promosi fitur ATM. Strategi 2: 1. Memonitor secara rutin jumlah uang dalam mesin ATM dan nominal transaksi pelanggan. 2. Membuat peramalan tentang jadwal pengisian mesin ATM berdasarkan data waktu dan nominal transaksi ATM, serta jadwal pengisian mesin ATM terdahulu.
3	Segmen 3	• <i>Recency</i> tinggi • <i>Frequency</i> sangat tinggi • <i>Monetary</i> sedang	• Mempertahankan aktivitas promosi fitur ATM yang sudah ada.	-

G. Validasi dengan Indeks Dunn (DI)

Validasi indeks Dunn bertujuan untuk mengetahui apakah jumlah *cluster* yang digunakan sebagai nilai *k* pada *clustering* dengan algoritma K-Means sudah sesuai.

Pada penelitian ini, program aplikasi R digunakan untuk menghitung nilai indeks Dunn. Untuk mengetahui bahwa 3 *cluster* merupakan jumlah *cluster* yang sesuai, dilakukan uji coba perhitungan indeks Dunn dengan menggunakan jumlah *cluster* 2 s.d 10. Berikut merupakan ringkasan hasil perhitungan indeks Dunn, Tabel 5:

Tabel 11 Uji Validitas dengan Indeks Dunn (DI)

Number of Cluster	Dunn's Index (DI)
2	0.1227234
3	0.3173171
4	0.1275784
5	0.2472033
6	0.04885966
7	0.06587666
8	0.1178215
9	0.1260685
10	0.2540008

Dari Tabel 5, dapat disimpulkan bahwa 3 *cluster* merupakan jumlah *cluster* yang paling optimal untuk digunakan sebagai inputan jumlah *cluster* (*k*) pada *clustering* dengan algoritma K-Means karena nilai indeks Dunn-nya yang lebih tinggi daripada jumlah *cluster* lainnya.

H. Uji Anova dengan Indeks R-Squared (R^2)

Indeks *R-Squared* digunakan untuk mengukur variasi pada *cluster* guna mengetahui tingkat homogenitas pada setiap *cluster*. Program aplikasi SPSS digunakan untuk menghitung indeks *R-Squared* hasil *clustering*, dimana

Distance of Case sebagai *dependent variable* dan *RECENCY*, *FREQUENCY*, & *MONETARY* sebagai *independent variable* (*predictors*).

Berikut merupakan hasil validasi indeks *R-Squared* (R^2) menggunakan SPSS, Tabel 12.

Tabel 12 Uji ANOVA dengan Indeks R-Squared

Model	R	R Square	Adjusted R Square
1	.310 ^a	.096	.028

Pada Tabel 6, didapat nilai *R* sebesar 0.310 > nilai tabel *R*, yang menunjukkan bahwa hubungan *dependent variable* dengan *predictors* mempunyai **korelasi positif dan lemah**. Korelasi positif dan lemah mengindikasikan bahwa nilai *RECENCY*, *FREQUENCY*, & *MONETARY* sedikit berpengaruh pada besarnya nilai *Distance of Case* atau jarak *cluster*. Sedangkan untuk nilai (R^2) hanya sebesar 0.096, dimana nilai tersebut tergolong lemah karena berada dibawah 0.2. Nilai (R^2) akan tergolong lemah jika dibawah 0.2, moderate jika berada diantara 0.2 s.d 0.4 dan kuat jika berada diatas 0.4. Acuan yang digunakan dalam mengetahui kategori nilai *R* dan (R^2) adalah buku Salkind " *Statistics for people who think they hate statistics*" yang diterbitkan pada tahun 2000.

Selain menghasilkan nilai *R* dan indeks (R^2), hasil uji indeks (R^2) juga menghasilkan Tabel 13 *Change Statistic*

Tabel 13 Change Statistic

F Change	df1	df2	Sig. F Change
1.414	3	40	0.253

*probabilitas 0.05

Dari Tabel 13, diketahui bahwa nilai *Sig. F Change* = 0.253 yang mengindikasikan bahwa *Sig. F Change* > 0.05. Sehingga dapat disimpulkan bahwa **perbedaan dalam cluster yang terbentuk tidak signifikan**.

I. Uji Coba Clustering

Dari visualisasi tersebut diketahui bahwa terdapat beberapa fitur ATM yang menjadi data *outliers* yaitu data yang menyimpang jauh dari kumpulan data yang ada. Fitur ATM yang tidak diikutsertakan dalam uji coba *clustering* karena merupakan data *outliers* yaitu APAYUNISBA, ATMBRIWDL, CMSPRIMTRF, CMSBERSTRF, DEBITPRIMA, dan VOIDPRIMA. Sehingga dari hanya 38 fitur ATM yang diikutsertakan dalam uji coba *clustering* ini. Dari hasil visualisasi terlihat bahwa terdapat 2 *cluster* yang berkumpul secara baik pada suatu titik. Untuk itu pada uji coba ini, dipilih **jumlah cluster** sebanyak 2 *cluster* sebagai inputan *clustering* dengan algoritma K-Means.

Daftar potongan hasil uji coba dapat dilihat pada Tabel 14. Dari hasil uji coba *clustering* didapatkan beberapa analisis tentang tingkat *Recency*, *Frequency* dan *Monetary* untuk masing-masing *cluster* sebagai berikut ;

- Tingkat *Recency*
 - *Cluster 1* dengan rentang nilai *Recency* 1 s.d 19 tergolong tingkat **Recency tinggi** yang menunjukkan bahwa fitur ATM yang menjadi anggota *cluster 1* digunakan terakhir kali tidak lebih dari 1 bulan terhitung sejak tanggal evaluasi yaitu 04 Desember 2014.
- Tingkat *Frequency*
 - *Cluster 1* dengan rentang nilai *Frequency* 1 s.d 4500 dimana tergolong tingkat **Frequency sangat tinggi** yang menunjukkan bahwa fitur ATM yang menjadi anggota *cluster 1* sangat sering digunakan untuk transaksi ATM.
 - *Cluster 2* dengan rentang nilai *Frequency* 9 s.d 398 dimana tergolong tingkat **Frequency sedang** yang menunjukkan bahwa fitur ATM yang menjadi anggota *cluster 2* cukup sering digunakan untuk transaksi ATM.
- Tingkat *Monetary*
 - *Cluster 1* dengan rentang nilai *Monetary* 18500 s.d 1.400.000 dimana tergolong tingkat **Monetary sedang** yang menunjukkan bahwa fitur ATM yang menjadi anggota *cluster 1* memberikan *fee-based income* yang cukup untuk pihak bank.
 - *Cluster 2* dengan rentang nilai *Monetary* 2.200.000 s.d 3.700.000 dimana tergolong tingkat **Monetary tinggi** yang menunjukkan bahwa fitur ATM yang menjadi anggota *cluster 2* memberikan *fee-based income* yang tinggi pada pihak bank.

Tabel 14 Daftar Potongan Hasil Uji Coba Clustering 2 Cluster

KODE_FITUR	RECENCY	FREQUENCY	MONETARY	Cluster	Distance
ATMBDELIMA	1.00	141.00	2645804.87	2	475003.41612
ATMBEXMON	1.00	53.00	3706286.66	2	585478.38279
ATMPDELIMA	1.00	10.00	2338940.00	2	781868.28293
ATMPXEMON	3.00	9.00	2276456.78	2	844351.50469
CMSBERSTRF	45.00	20.00	5046979.25	2	1926170.97315
EDCBRISSTOR	1.00	398.00	3031590.40	2	89218.40560
EDCBRISWDRL	2.00	15.00	2799600.00	2	321208.28792
ATMBERSTRF	1.00	1835.00	1308310.65	1	1001497.41594
ATMBERSWDRL	1.00	4538.00	395013.22	1	88292.08910
ATMBRISTRF	1.00	376.00	804383.30	1	497569.16645
ATMBRISWDRL	1.00	3480.00	430545.98	1	123768.04377
ATMBRITRF	1.00	648.00	909871.71	1	603057.59513
ATMPLN	1.00	32.00	883981.69	1	577167.72539

Untuk uji ANOVA pada uji coba *clustering* ini didapatkan tabel ANOVA sebagai berikut, Tabel 15;

Tabel 15 Uji ANOVA Uji Coba Clustering

Model	R	R Square	Adjusted R Square	Change Statistics				
				R Square Change	F Change	df1	df2	Sig. F Change
1	.432 ^a	.187	.115	.187	2.606	3	34	.068

Dari tabel diatas didapatkan bahwa **indeks R-Square** uji coba *clustering* ini sebesar **0.187** yang dimana nilai tersebut tergolong **lemah** karena dibawah 0.2. . Hal menunjukkan bahwa **variasi yang terbentuk dalam cluster adalah kecil**. Sedangkan untuk nilai *Sig. F Change* sebesar 0.068 lebih besar dari probabilitasnya sebesar 0.05. Nilai *Sig. F Change* > 0.05 menunjukkan bahwa **perbedaan dalam cluster yang terbentuk tidak signifikan**. Dengan begitu dapat disimpulkan bahwa uji coba *clustering* dengan 2 *cluster* menghasilkan persebaran data yang lebih rapi dimana nilai *Sig. F Change* yang jauh lebih kecil dibanding dengan 3 *cluster*.

IV. KESIMPULAN

Dari hasil *clustering* yang dilakukan menggunakan metode Ward dan algoritma K-Means didapatkan kesimpulan sebagai berikut :

1. Parameter yang digunakan sebagai kriteria penentuan fitur ATM paling berharga adalah waktu dan jumlah uang yang digunakan dalam transaksi fitur ATM. Parameter waktu dibagi menjadi 2 yaitu tingkat kebaruan (*Recency*) dan tingkat penggunaan (*Frequency*) fitur ATM dalam transaksi.
2. Fitur ATM pada PT. Bank BRISyariah terbagi menjadi 3 *cluster* atau segmen dimana ketiga *cluster* atau segmen tersebut memiliki karakteristik sebagai berikut :
 1. *Cluster 1* : Fitur ATM dengan tingkat ***Recency tinggi*** dan ***Frequency sangat rendah***, serta tingkat ***Monetary sangat tinggi***.
 2. *Cluster 2* : Fitur ATM dengan tingkat ***Recency sangat tinggi*** dan ***Frequency sedang***, serta tingkat ***Monetary tinggi***.
 3. *Cluster 3* : Fitur ATM dengan tingkat ***Recency tinggi*** dan ***Frequency sangat tinggi***, serta tingkat ***Monetary sedang***.
3. Dari hasil analisis piramida RFM didapatkan bahwa fitur ATM yang berada pada *cluster* 3 yaitu ATMBRIWDL dan ATMBERSWDRL paling banyak memberikan keuntungan pada PT. Bank BRISyariah dengan tingkat ***Monetary sedang*** namun memiliki tingkat ***Recency tinggi*** dan ***Frequency sangat tinggi***.
4. Visualisasi dengan diagram *scatter plot* mempermudah pengamatan persebaran fitur ATM dan letak pusat tiap *cluster* pada atribut RFM. sedangkan piramida RFM mempermudah dalam mengurutkan peringkat fitur ATM berdasarkan nilai RFM yang dimiliki sehingga

diketahui fitur ATM mana yang paling memberikan keuntungan bagi pihak bank.

5. Hasil uji coba 2 *cluster* menghasilkan persebaran fitur ATM yang lebih baik dibanding dengan *clustering* dengan 3 *cluster*. Berikut merupakan karakteristik *cluster* hasil uji coba 2 *cluster*:

1. *Cluster 1* : Fitur ATM dengan tingkat ***Recency tinggi*** dan ***Frequency sangat tinggi***, serta tingkat ***Monetary sedang***.
2. *Cluster 2* : Fitur ATM dengan tingkat ***Recency sangat tinggi*** dan ***Frequency sedang***, serta tingkat ***Monetary tinggi***.

V. SARAN

Saran yang dapat dipertimbangkan untuk pengembangan tugas akhir ini yaitu sebaiknya atribut yang digunakan dalam *clustering* untuk menentukan segmentasi fitur ATM ditambah dengan variabel *Length* yaitu durasi transaksi fitur ATM dilakukan.

DAFTAR PUSTAKA

- [1] J. Gilmartin, "Product Segmentation - Using the Customer's Mind," *Coming of Age Baby Boomer Marketing & Senior Marketing*, p. 1, 2006.
- [2] Alstons, "Cluster Analysis for Segmentation," *UVA-M-0748*, p. 1, 2007.
- [3] Y.-F. Chen, C.-Y. Chiu, I.-T. Kuo and H. C. Ku, "An Intelligent market segmentation system using k-means and particle swarm optimization," *Expert Systems with Applications*, p. 1, 2009.
- [4] M. Alvandi, S. Fazli and R. S. Abdoli, "K-Mean Clustering Method For Analysis Customer Lifetime Value With LFRM Relationship Model In Banking Services," *International Research Journal of Applied and Basic Sciences*, pp. 2294-2302, 2012.
- [5] PT. Bank BRISyariah, "Consumer Banking," 2011. [Online]. Available: <http://www.brisyariah.co.id/?q=tabungan-brisyariah>.
- [6] PT. Bank BRISyariah, "Annual Report BRISyariah 2013," PT. Bank BRISyariah, Jakarta, 2013.
- [7] E. Turban, R. Sharda and D. Delen, "Data Mining for Business Intelligence," in *Decision Support and Business Intelligence Systems*, New Jersey, Prentice Hall, 2011, pp. 194-231.

- [8] K. I. Moin and D. Q. B. Ahmed, "Use of Data Mining in Banking," *International Journal of Engineering Research and Application Vol. 2*, pp. 738 - 742, 2012.
- [9] M. Kamber and J. Han, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, 2001.
- [10] D. Larose, *Data mining methods and models*, New Jersey: John Wiley & Sons, Inc., 2006.
- [11] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*, London: Addison-Wesley, 1974.
- [12] L. Rokach and O. Maimon, "CLUSTERING METHODS," *DATA MINING AND KNOWLEDGE DISCOVERY HANDBOOK*, pp. 330-343, 2009.
- [13] S. Ayramo and T. Karkkainen, "Introduction to partitioning-based clustering methods with a robust example," University of Jyväskylä, Agora, 2006.
- [14] J. D. Banfield and A. E. Raftery, "Model-based Gaussian and non-Gaussian," *Biometrics*, 1993, pp. 803-821.
- [15] D. Birant, "Data Mining Using RFM Analysis," in *Knowledge-Oriented Applications in Data Mining*, Turkey, InTech, 2011, p. 93.
- [16] S. A. Rasul, "A Side of Hash for You To Dig Into," *SAS Global Forum 2011 Coders*, pp. 112 - 119, 2011.
- [17] S. K. Bhatia, "A Propound Method for the Improvement of Cluster Quality," *IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 2*, pp. 216-221, 2012.
- [18] M. K. Pakhira, "Finding Number of Clusters before Finding Clusters," pp. 27-37, 2012.
- [19] M. Ashoori and Z. Taheri, "Using Clustering Methods for Identifying Blood Donors Behavior," *5th Iranian Conference on Electrical and Electronics Engineering*, pp. 4055-4058, 2013.
- [20] M. Hardin, D. Hom, R. Perez and L. Williams, "Which chart or graph is right for you?," *Whitepaper Tableau*, pp. 10-15, 2013.
- [21] M. Rueter, "Tableau for the Enterprise : An Overview for IT," *White Paper Tableau for Enterprise*, pp. 1-14, 2012.
- [22] J. J. H. Ward, "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association*, vol. 58, p. 236-244, 1963.
- [23] E. Forgy, "Cluster Analysis of multivariate data: efficiency vs. interpretability of classifications," *Biometrics*, vol. 21, pp. 768-780, 1965.
- [24] B. K. Slinker and S. A. Glantz, *Primer of Applied Regression and Analysis of Variance*, New York: McGraw-Hill, 2000.
- [25] Wikipedia, "F-test," 14 December 2014. [Online]. Available: <http://en.wikipedia.org/wiki/F-test>.
- [26] UCLA: Statistical Consulting Group, "SPSS Annotated Output," 2015. [Online]. Available: http://www.ats.ucla.edu/stat/spss/output/reg_spss.htm.