

SEGMENTASI PELANGGAN PADA CUSTOMER RELATIONSHIP MANAGEMENT DI PERUSAHAAN RITEL: STUDI KASUS PT GRAMEDIA ASRI MEDIA

Christina Deni Rumiarti¹, Indra Budi²

^{1,2}Program Studi Magister Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Indonesia, Jl. Salemba Raya No. 4 Jakarta 10430, Indonesia

Email: ¹deni.christina@gmail.com, ²indra@cs.ui.ac.id

Abstract

Advances in information technology produces wide range of choices in accessing information including reading books. The increase in the number of readers who turning to electronic books making sales of printed books has decreased in the recent years. PT Gramedia Asri Media is one of book retail company in Indonesia. Gramedia implement CRM by launching a member card named Kompas Gramedia Value Card (KGVC). Promotion given has not been able to increase book transaction of KGVC members. This research focus on make customer segmentation in CRM at PT Gramedia Asri Media. Data mining process is done by clustering using K-means algorithm for segmenting customers based on RFM, as well as hierarchical clustering algorithms for segmentation of customers based on the number of books type. Evaluation is done on cluster result using elbow method, silhouette method, and Calinski-Harabasz index. Customer segmentation based on the RFM produce two optimal clusters, occasional customers and dormant customers. Customer segmentation based on the number of types of books purchased produce 3 optimal cluster, namely low, medium, and high. With these results, it is expected to help the company classifying KGVC members to determine the appropriate strategies, so company can increase the number of books transactions.

Keywords: *clustering, customer segmentation, hierarchical, K-means, RFM*

Abstrak

Kemajuan teknologi informasi menghasilkan berbagai pilihan dalam mengakses informasi termasuk membaca buku. Peningkatan jumlah pembaca yang beralih ke buku elektronik mengakibatkan angka penjualan buku cetak mengalami penurunan pada beberapa tahun terakhir. PT Gramedia Asri Media merupakan salah satu perusahaan ritel buku di Indonesia. Gramedia menerapkan CRM dengan menerbitkan kartu member Kompas Gramedia *Value Card* (KGVC). Promosi yang diberikan belum mampu meningkatkan transaksi buku *member* KGVC. Penelitian ini bertujuan untuk membuat segmentasi pelanggan pada CRM di PT Gramedia Asri Media. Proses *data mining* dilakukan dengan melakukan *clustering* menggunakan algoritma *K-means* untuk segmentasi pelanggan berdasarkan RFM, serta algoritma *hierarchical clustering* untuk segmentasi pelanggan berdasarkan banyaknya jenis buku. Evaluasi terhadap hasil cluster menggunakan *elbow method*, *silhouette method*, dan *Calinski-Harabasz index*. Segmentasi pelanggan berdasarkan RFM menghasilkan 2 *cluster* yang optimal, yaitu *occasional customers* dan *dormant customers*. Segmentasi pelanggan berdasarkan banyaknya jenis buku yang dibeli menghasilkan 3 *cluster* yang optimal, yaitu rendah, sedang, dan tinggi. Dengan hasil penelitian ini, diharapkan dapat membantu perusahaan mengelompokkan member KGVC dalam menentukan strategi yang sesuai sehingga dapat meningkatkan jumlah transaksi buku.

Kata Kunci: *clustering, segmentasi pelanggan, hierarchical, K-means, RFM*

1. Pendahuluan

Dalam mendukung program CRM, Gramedia menggunakan *data warehouse* untuk mengetahui historis transaksi setiap *member* KGVC. Data yang digunakan untuk menganalisis *member* KGVC hanya sebatas data jumlah distribusi *member* KGVC berdasarkan umur, tingkat pendidikan, jenis kelamin, pekerjaan, lokasi toko tempat registrasi, dan berdasarkan kota di Indonesia [1]. Selain itu, juga

terdapat data mengenai jumlah pembelanjaan oleh *member* KGVC berdasarkan umur, tingkat pendidikan, jenis kelamin, lokasi toko, dan produk. Padahal, banyak pola belanja dari pelanggan yang menarik bagi CRM untuk dianalisis lebih lanjut [2]. Hal ini menunjukkan bahwa data yang dimiliki saat ini belum dimanfaatkan secara maksimal untuk menganalisis karakteristik *member* KGVC.

Berdasarkan data CRM, jumlah *member* KGVC pada bulan Juli 2016 mengalami kenaikan

TABEL 1
DAFTAR INSTRUMEN PENELITIAN

Variabel	Keterangan
Recency	Recency merupakan interval waktu sejak terakhir kali <i>member</i> KGVC melakukan transaksi pembelian buku.
Frequency	Frequency merupakan seberapa sering <i>member</i> KGVC melakukan transaksi pembelian buku.
Monetary	Monetary merupakan seberapa besar jumlah yang dibayarkan oleh <i>member</i> KGVC pada transaksi dalam periode 1 Januari 2016 – 31 Juli 2016.

sebanyak 68.587 sejak bulan Januari 2016 [1]. Permasalahan terjadi ketika jumlah *member* KGVC mengalami kenaikan, tetapi transaksi buku *member* KGVC mengalami penurunan. Total transaksi buku *member* KGVC pada bulan Juli 2016 mengalami penurunan sembilan milyar rupiah sejak bulan Januari 2016. Hal ini menunjukkan bahwa kenaikan jumlah *member* KGVC ternyata tidak berkorelasi dengan kenaikan jumlah transaksi buku. Semakin bertambahnya jumlah *member* KGVC seharusnya nilai transaksi buku yang dilakukan *member* KGVC juga mengalami peningkatan.

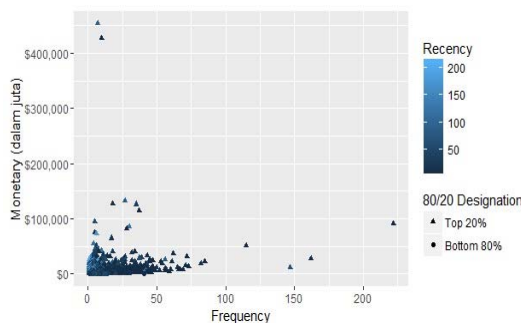
Berdasarkan kondisi dan permasalahan yang telah diuraikan di atas, diidentifikasi mengenai faktor-faktor yang memengaruhi transaksi buku *member* KGVC mengalami penurunan, yaitu (1) produk (adanya buku-buku bajakan yang beredar di pasaran), (2) harga (perbandingan harga buku dan kualitas buku yang relatif sama dengan toko buku lain, serta munculnya toko buku *online* yang menjual buku cetak dengan harga yang kompetitif), dan (3) teknologi (belum memanfaatkan perkembangan teknologi untuk menganalisis pelanggan pada CRM, belum adanya sistem rekomendasi buku, dan belum memiliki sistem untuk mengetahui segmen pelanggan yang dimiliki). Akar masalah yang menjadi dasar penelitian ini adalah belum adanya segmentasi pelanggan pada CRM untuk menentukan strategi promosi yang sesuai dengan segmen pelanggan. Penelitian ini bertujuan untuk menghasilkan segmentasi pelanggan pada *Customer Relationship Management* di PT Gramedia Asri Media.

RFM Analysis

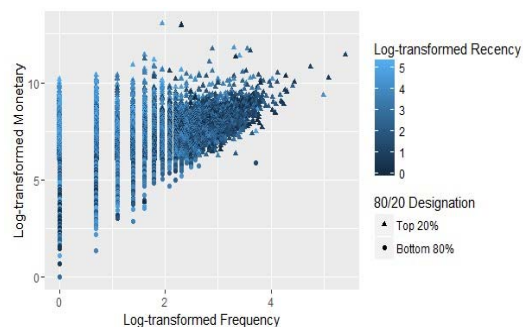
Menurut Tsiptsis dan Chorianopoulos [3], analisis RFM adalah pendekatan yang biasa digunakan di perusahaan ritel untuk memahami perilaku pembelian konsumen. Seperti namanya, analisis RFM melibatkan perhitungan *recency*, *frequency*, dan *monetary*. Perhitungan *recency* menunjukkan waktu sejak terakhir transaksi pelanggan, *frequency* menunjukkan jumlah transaksi pembelian, sedangkan *monetary* menunjukkan nilai pembelian. Pada penelitian ini, label pelanggan yang digunakan adalah (1) *superstars customers*, yaitu pelanggan yang sering melakukan transaksi dan memiliki nilai transaksi yang tinggi, (2) *golden customers*, yaitu pelanggan yang sering melakukan transaksi dan memiliki nilai transaksi kedua yang tinggi, (3) *occasional customers*, yaitu pelanggan yang jarang berkunjung tetapi memiliki nilai transaksi rata-rata yang tinggi, (4) *everyday shoppers*, yaitu pelanggan yang sering melakukan transaksi tetapi memiliki nilai transaksi yang rendah, dan (5) *dormant customers*, yaitu pelanggan dengan tingkat pembelian yang sangat rendah dan sudah lama tidak melakukan transaksi.

Data Mining pada Customer Relationship Management

Data mining memberikan pemahaman yang lebih mendalam tentang pelanggan yang sangat penting untuk membangun strategi CRM yang efektif [3]. Aktivitas CRM yang dapat didukung melalui penggunaan *data mining* meliputi: (1) *customer segmentation* akan menghasilkan segmen pelanggan berdasarkan karakteristik perilaku pelanggan, (2) *direct marketing campaigns* untuk berkomunikasi secara langsung dengan pelanggan dalam meningkatkan jumlah pelanggan baru dan pembelian produk, dan (3) *market basket analysis* untuk mengidentifikasi produk-produk terkait yang biasanya dibeli secara bersama-sama.



Gambar 1. Visualisasi Data berdasarkan Analisis RFM



Gambar 2. Visualisasi Log-transformed RFM

TABEL 2
HASIL CLUSTERING PADA 2 CLUSTER, 3 CLUSTER, DAN 4 CLUSTER

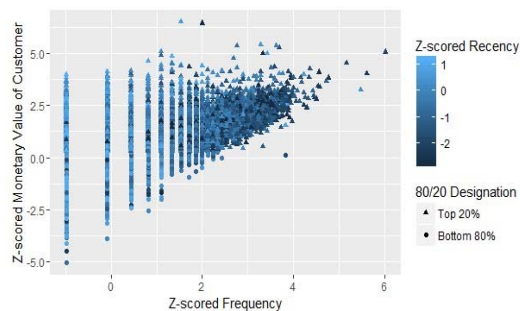
Cluster	Persentase Jumlah Pelanggan (%)	Recency (hari)	Frequency (kali)	Monetary (ribu rupiah)	Label Pelanggan
2 Cluster					
1	40,58	22	4	759	<i>Occasional customers</i>
2	59,42	99	1	177	<i>Dormant customers</i>
3 Cluster					
1	35,90	107	1	132,5	<i>Dormant customers</i>
2	43,62	9	6	982,05	<i>Golden customers</i>
3	20,48	59	3	487	<i>Everyday shoppers</i>
4 Cluster					
1	31,10	122	1	131,9	<i>Dormant customers</i>
2	16,77	12	7	1217,65	<i>Golden customers</i>
3	37,74	10	2	245	<i>Dormant customers</i>
4	14,39	67	3	509,5	<i>Occasional customers</i>

Metodologi Data Mining

Industri *data mining* mengusulkan proses standar *data mining* yang disebut dengan *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Menurut Maheshwari [4], CRISP-DM memiliki urutan fase yang adaptif, dimana tahap berikutnya dalam urutan tergantung hasil dari tahap sebelumnya. CRISP-DM memiliki siklus hidup yang terdiri dari enam fase, yaitu [5]: *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*.

K-means Clustering

Menurut Tsipitsis dan Chorianopoulos [3], *K-means* menggunakan ukuran *Euclidean distance* dan secara iteratif menentukan tiap *record* dari *cluster* asal. Prosedur dimulai dengan memilih *k* dengan *initial records* sebagai pusat *cluster* (*initial seed*)



Gambar 3. Visualisasi Standarisasi data variabel RFM

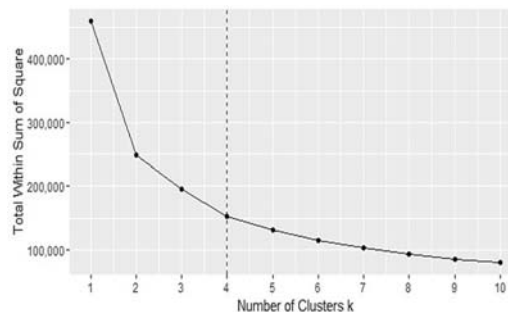
dan menentukan setiap *record* yang paling dekat dengan *cluster*. *Record* baru ditambahkan ke *cluster*, *cluster center* dihitung ulang untuk mencerminkan anggota baru mereka. Prosedur iterasi ini diulang sampai konvergen dan migrasi *record* dengan *cluster* tidak lagi menyaring solusinya.

Hierarchical Clustering

Dalam *hierarchical clustering*, dendrogram dihasilkan melalui partisi rekursif (metode *divisive*) atau menggabungkan *agglomerative cluster* yang ada. Metode *agglomerative clustering* menginisialisasi setiap pengamatan menjadi *cluster* yang kecil. Kemudian, dua *cluster* terdekat dikumpulkan ke dalam *cluster* gabungan baru. Dengan cara ini, jumlah *cluster* dalam kumpulan data berkurang satu di setiap langkah. Akhirnya, semua *record* digabungkan menjadi *cluster* besar yang tunggal. *Divisive methods* dimulai dengan semua *record* dalam satu *cluster* yang besar. *Record* yang berbeda memisahkan diri secara rekursif dalam *cluster* yang terpisah sampai setiap *record* mewakili kelompoknya sendiri. Kebanyakan program komputer menerapkan metode *agglomerative* pada *hierarchical clustering* [5].

Elbow Method

Menurut Kodinariya dan Makwana [6], *elbow method* digunakan untuk menentukan jumlah *cluster* dari dataset. Ide dasar dari *unsupervised model* (misal: *K-means clustering*) adalah untuk menentukan *cluster* sehingga *total intra-cluster variation* (dikenal sebagai *total within-cluster variation* atau *total within-cluster sum of square*) diminimalkan. Metode ini merupakan metode visual. Idennya adalah dimulai dengan $k = 2$, dan terus meningkat dalam setiap langkah dengan ditambah 1 pada nilai k . Pada nilai $k=3$, apabila terjadi perubahan drastis yang berbanding terbalik dengan nilai sebelumnya, maka nilai sebelum terjadinya perubahan tersebut dianggap sebagai jumlah *cluster* yang paling tepat.



Gambar 4. Jumlah Optimal Cluster berdasarkan *Elbow Method*

Silhouette Method

Nilai *silhouette* digunakan untuk mengukur seberapa baik *cluster* pada titik tertentu [5]. Nilai positif menunjukkan hasil yang baik, dengan nilai yang lebih tinggi lebih baik daripada nilai yang lebih rendah. Sebuah nilai yang mendekati nol dianggap tidak bagus. Nilai *silhouette* negatif dianggap sebagai kesalahan klasifikasi.

Calinski-Harabasz Index

Menurut Liu dkk. [7], *Calinski-Harabasz index* (CH) mengevaluasi validitas *cluster* berdasarkan perhitungan *between sum of squares* dan *within sum of squares*. CH mengukur *ratio* dari *separation* berdasarkan jarak maksimum *between cluster centers*, dan mengukur *compactness* berdasarkan jumlah jarak antara objek dan pusat *cluster*. Konfigurasi *compact* dan *well-separated* dari *cluster* diharapkan memiliki nilai *inter-cluster variance* yang tinggi dan *intra-cluster variance* yang relatif rendah [8].

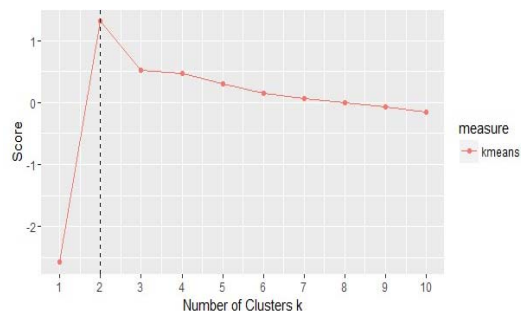
2. Metode Penelitian

Tahapan penelitian dilakukan berdasarkan pada metodologi *data mining*, yaitu CRISP-DM. Tahapan *deployment* dalam CRISP-DM tidak dibahas pada penelitian ini. Berikut ini penjelasan tahapan penelitian yang dilakukan penulis:

Business Understanding (Pemahaman Bisnis): Tahap ini bertujuan untuk memahami kondisi perusahaan dan permasalahan yang dihadapi saat ini.

Tinjauan Pustaka: Penulis melakukan tinjauan pustaka dengan studi literatur terkait teori yang relevan dan melakukan kajian terhadap penelitian sebelumnya yang relevan.

Data Understanding (Pemahaman Data): Pengumpulan data yang dibutuhkan penelitian dilakukan pada tahap pemahaman data.



Gambar 6. Jumlah Optimal Cluster berdasarkan *Calinski-Harabasz index*

Data Preparation (Persiapan Data): Variabel *recency*, *frequency*, dan *monetary* digunakan untuk mengidentifikasi perilaku belanja pelanggan. Kemudian, tahap *preprocessing* dilakukan untuk transformasi data.

Modeling (Pembentukan Model): Algoritma *K-means* digunakan untuk segmentasi pelanggan berdasarkan variabel RFM. Algoritma *hierarchical* digunakan untuk segmentasi berdasarkan jenis buku yang dibeli.

Evaluation (Evaluasi Model): Evaluasi clustering menggunakan tiga metode yaitu, *elbow method*, *silhouette method*, dan *Calinski-Harabasz index*.

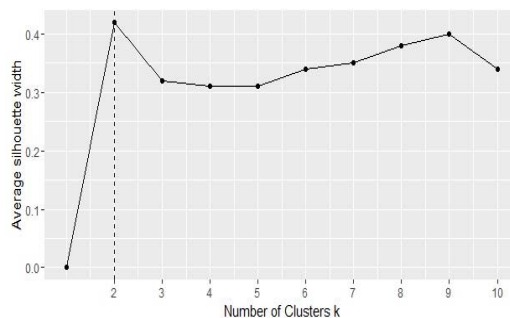
Kesimpulan dan Saran: Penyusunan kesimpulan dan saran dilakukan berdasarkan analisis dan pembahasan yang dilakukan penulis untuk menjawab pertanyaan penelitian terkait segmentasi pelanggan pada CRM di PT Gramedia Asri Media.

Pengumpulan data dilakukan dengan observasi terlebih dahulu terhadap data yang dibutuhkan penelitian ini. Data yang dikumpulkan adalah data *member KGVC* dan data transaksi yang berasal dari basisdata *Oracle*. Data tersebut kemudian diubah ke dalam format *file .csv* agar dapat diolah lebih lanjut untuk pembentukan *clustering*. Tahap *preprocessing*, pembentukan *cluster*, dan evaluasi hasil *cluster* dilakukan menggunakan software R Version 3.3.1.

Instrumen penelitian yang dikaji oleh penulis berdasarkan pada variabel RFM (*recency*, *frequency*, dan *monetary*). Tabel I menunjukkan daftar instrumen penelitian yang menjadi acuan penulis pada saat pembentukan *cluster* menggunakan algoritma *K-means*.

4. Hasil dan Pembahasan

Data Understanding



Gambar 5. Jumlah Optimal Cluster berdasarkan *Average Silhouette*

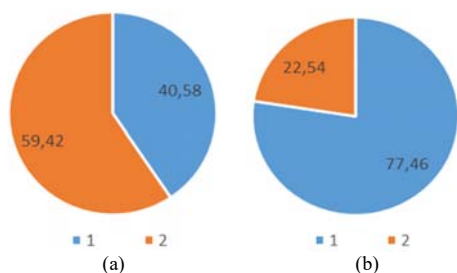
TABEL 3
CONTOH DATASET BUKU

CUSTID	CHIL- DRENS BOOKS	FICTION LITERA- TURE	AGRI- CUL- TURE	ART ARCHI- TECTURE
97244478	1	1	0	0
742689				
85499161	1	0	0	0
76318				
33788457	1	0	0	0
573423				

Data yang dikumpulkan adalah data *member* KGVC dan data transaksi. Pelanggan yang masih aktif per tanggal 1 Agustus 2016 diolah untuk segmentasi. Data *member* KGVC diambil dari data pelanggan seluruh toko Gramedia yang masih aktif pada tanggal 31 Juli 2016. Sedangkan data transaksi merupakan transaksi produk buku dari *member* KGVC pada periode 1 Januari 2016 sampai 31 Juli 2016. Data transaksi yang digunakan dalam penelitian adalah seluruh data transaksi buku dari *member* KGVC yang masih aktif pada tanggal 31 Juli 2016. Data yang diambil merupakan transaksi dengan jumlah nilai lebih dari nol rupiah. Dengan kata lain, transaksi pengembalian buku tidak diambil pada tahap pengumpulan data. Berdasarkan 389.686 pelanggan yang masih aktif pada tanggal 31 Juli 2016, jumlah transaksi pada periode 1 Januari – 31 Juli 2016 adalah 1.535.002 transaksi. Berdasarkan transaksi tersebut, hanya 153.347 pelanggan yang melakukan transaksi pada periode tersebut.

Data Preparation

Variabel input yang digunakan untuk analisis *cluster* adalah *recency* (R), *frequency* (F), dan *monetary* (M). Rentang nilai variabel *recency* adalah 1–213 hari. Semakin kecil nilainya mengindikasikan bahwa aktivitas transaksi terbaru yang dilakukan oleh *member* KGVC. Rentang nilai variabel *frequency* adalah 1–222 transaksi/pelanggan. Semakin besar nilainya mengindikasikan bahwa *member* KGVC semakin sering berbelanja pada periode 1 Januari 2016–31 Juli 2016. Dalam kurun waktu 7 bulan, ada pelanggan yang baru berbelanja satu kali dan ada juga pelanggan yang sudah berbelanja sebanyak 222 kali. Rentang nilai variabel *monetary*



Gambar 7. (a) Persentase Jumlah *Member* KGVC dan (b) Nilai Transaksi Buku *Member* KGVC Tiap Cluster

TABEL 4
CONTOH DATASET BUKU

Cluster	Persentase Jumlah Pelanggan (%)	Mean / Rata-rata Jumlah Pembelian Buku	Label Pelanggan
1	82,80	2	Rendah
2	16,66	5	Medium
3	0,54	7	Tinggi

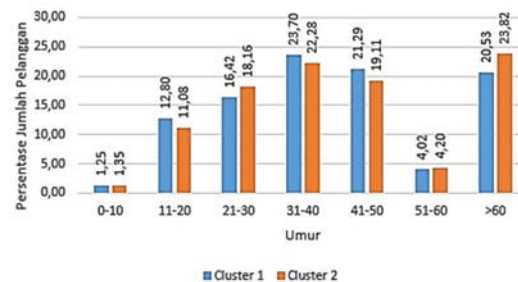
adalah Rp 1.000–Rp 453.657.000. Semakin besar nilainya mengindikasikan bahwa *member* KGVC semakin banyak nilai pembelanjannya pada periode 1 Januari 2016–31 Juli 2016.

Seperti terlihat pada Gambar 1, nilai *monetary* pelanggan berada pada sumbu y dan *frequency* pembelian berada pada sumbu x. Nilai *recency* divisualisasikan dalam 80/20, yaitu 80 persen dan 20 persen berdasarkan total nilai pembelian. Nilai RFM tidak terdistribusi secara normal, yaitu distribusi miring (*skewed distribution*).

Menurut Zumel dan Mount [9], distribusi data positif yang sangat miring, seperti nilai pelanggan, pendapatan, penjualan, dan harga saham dapat dimodelkan sebagai *lognormal distribution*. Untuk keperluan *modeling*, penggunaan *natural logarithm*, yaitu *log base 10*, atau *log base 2* umumnya tidak bermasalah dan tidak memengaruhi nilai hasil. Pada penelitian ini, supaya nilai RFM terdistribusi normal, maka nilai RFM tersebut diubah menjadi *lognormal distribution* terlebih dahulu sebelum dilakukan *modeling*.

Berdasarkan Gambar 2, terlihat bahwa penyebaran data lebih baik dibandingkan Gambar 1, yaitu sebelum data RFM ditransformasikan menjadi *lognormal distribution*. Pelanggan dengan *frequency* tinggi dan *monetary* tinggi dapat terlihat pada sudut kanan atas grafik. Sedangkan pelanggan dengan *frequency* rendah dan *monetary* rendah terlihat pada sudut kiri grafik.

Langkah terakhir *data preparation* adalah standarisasi data. Standarisasi akan membuat variabel RFM sebanding satu sama lain dalam perhitungan algoritma *clustering*. Visualisasi standarisasi data variabel RFM dapat dilihat pada Gambar 3. Pada penelitian ini, variabel RFM memiliki skala



Gambar 8. Persentase Jumlah *Member* KGVC berdasarkan Umur pada Segmentasi RFM

yang berbeda, yaitu *recency* = hari, *frequency* = jumlah, dan *monetary* = rupiah. Skala pada setiap variabel RFM pada dataset diubah menjadi *zero means* dan *unit standard deviation* (*z-scores*). Hal ini dilakukan untuk mencegah salah satu atribut dengan rentang nilai yang besar akan mendominasi variabel yang lain dalam perhitungan jarak.

Segmentasi Pelanggan berdasarkan RFM

Pada penelitian ini, algoritma *K-means* digunakan untuk pembentukan *cluster* berdasarkan variabel RFM. Tahap pembentukan model dilakukan menggunakan software R Version 3.3.1. *Cluster* yang diuji coba adalah $k=2$ sampai $k=10$. Hasil *clustering* menggunakan algoritma *K-means* pada $k=2$ sampai $k=4$ dapat dilihat pada Gambar 3.

Metode yang digunakan untuk penentuan jumlah *cluster* optimal berdasarkan RFM ada tiga, yaitu *elbow method*, *silhouette method*, dan *Calinski-Harabasz index*. Berikut ini penjelasan lebih lanjut metode evaluasi *cluster* pada algoritma *K-Means*:

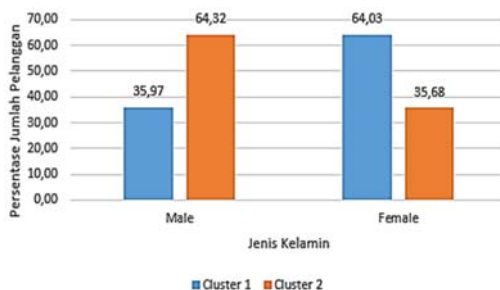
Elbow Method

Tujuan dari *elbow method* adalah untuk memilih nilai k yang kecil dan masih memiliki nilai *withinss* yang rendah. Nilai *withinss* cenderung menurun menuju 0 (nol) ketika nilai k ditingkatkan. Berdasarkan Gambar 4, nilai $k=4$ dengan persentase *variance* = 66,8 persen dan *total within-cluster sum of square* = 152913,93 adalah *cluster* yang optimal.

Silhouette Method

Jumlah optimal *cluster* k dengan nilai *average silhouette* yang tinggi mengindikasikan *cluster* yang baik. *Cluster* dengan $k=2$ memiliki nilai *average silhouette* paling tinggi dibandingkan *cluster* yang lain. Nilai *average silhouette* pada dua *cluster* adalah 0.42. Berdasarkan Gambar 5, dua *cluster* adalah *cluster* yang optimal.

Calinski-Harabasz Index



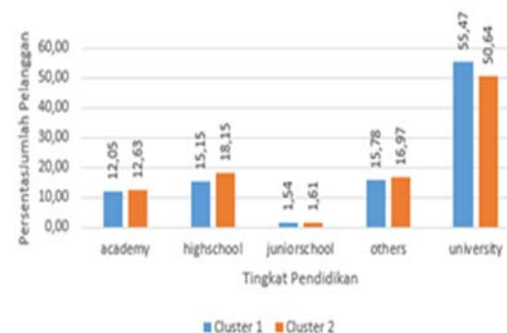
Gambar 10. Persentase Jumlah *Member* KGVC berdasarkan Jenis Kelamin pada Segmentasi RFM

Berdasarkan Gambar 6, dua *cluster* adalah *cluster* yang optimal. *Cluster* dengan $k=2$ memiliki *Calinski-Harabasz index* paling tinggi dibandingkan *cluster* yang lain. *Calinski-Harabasz index* pada dua *cluster* adalah 1.31.

Berdasarkan tahap evaluasi, jumlah *cluster* yang optimal berdasarkan *elbow method* adalah $k=4$, sedangkan jumlah *cluster* yang optimal berdasarkan *silhouette method* dan *Calinski-Harabasz index* adalah $k=2$. Pelanggan pada 2 *cluster* terdiri dari *occasional customers* dan *dormant customers*. Pada penelitian ini, $k=2$ dipilih penulis untuk dianalisis lebih lanjut. Hal ini dikarenakan 2 *cluster* merupakan *cluster* yang optimal berdasarkan dua metode evaluasi.

Persentase jumlah *member* KGVC dibandingkan nilai transaksi pada setiap *cluster* ditunjukkan pada Gambar 7. *Occasional customers* terdiri dari 40,58 persen *member* KGVC, sedangkan *dormant customers* terdiri dari 59,42 persen *member* KGVC. Meskipun *occasional customers* memiliki persentase pelanggan yang lebih sedikit, namun nilai transaksi justru lebih besar yaitu 77,46 persen dari total transaksi buku *member* KGVC. *Member* KGVC pada *occasional customers* merupakan pelanggan yang sangat potensial bagi Gramedia.

Persentase jumlah *member* KGVC ditinjau dari segi umur pada setiap *cluster* ditunjukkan pada Gambar 8. Pelanggan terbagi menjadi kategori umur 0-10 tahun, 11-20 tahun, 21-30 tahun, 31-40 tahun, 41-50 tahun, 51-60 tahun, dan pelanggan yang tidak diketahui umurnya dimasukkan dalam kategori lebih dari 60 tahun. Pelanggan yang tidak diketahui umurnya mendominasi masing-masing *cluster*, yaitu *cluster* 1 sebanyak 20,53 persen dan *cluster* 2 sebanyak 23,82 persen. Segmen pelanggan berdasarkan umur pada masing-masing *cluster* memiliki persentase populasi yang hampir sama pada masing-masing *cluster* dengan persentase jumlah *member* KGVC pada *cluster* 1 lebih banyak dibandingkan *cluster* 2.



Gambar 9. Persentase Jumlah *Member* KGVC berdasarkan Tingkat Pendidikan pada Segmentasi RFM

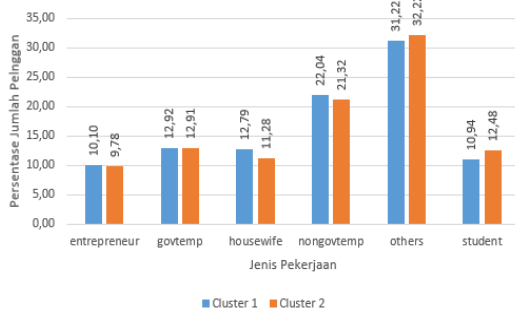
Persentase jumlah *member* KGVC ditinjau dari segi jenis kelamin ditunjukkan pada Gambar 10. *Member* KGVC yang termasuk dalam *occasional customers* lebih banyak pelanggan perempuan dengan persentase 64,03. Sedangkan *member* KGVC yang termasuk dalam *dormant customers* lebih banyak pelanggan laki-laki dengan persen-tase 64,32. Berdasarkan perbandingan ini menunjukkan bahwa *member* KGVC berjenis kelamin perempuan lebih potensial dibandingkan pelanggan laki-laki.

Persentase jumlah *member* KGVC ditinjau dari segi tingkat pendidikan ditunjukkan pada Gambar 9. *Member* KGVC dengan tingkat pendidikan universitas mendominasi masing-masing pelanggan pada *cluster* 1 dan 2. Sebanyak 55,47 persen pelanggan dengan tingkat pendidikan universitas merupakan *occasional customers*. Demikian juga sebanyak 50,64 persen pelanggan dengan tingkat pendidikan universitas merupakan *domant customers*. Segmen *member* KGVC dengan tingkat pendidikan berbeda memiliki persentase populasi yang hampir sama pada masing-masing *cluster*.

Persentase jumlah *member* KGVC ditinjau dari segi jenis pekerjaan ditunjukkan pada Gambar 12. Pelanggan yang tidak diketahui jenis pekerjaan mendominasi masing-masing *cluster*, yaitu *cluster* 1 sebanyak 31,22 persen sedangkan *cluster* 2 sebanyak 32,22 persen. *Occasional customers* terdiri dari pelanggan dengan sebagian besar variasi jenis pekerjaan dan persentase jumlah *member* KGVC lebih banyak dibandingkan *dormant customers*. Sedangkan persentase *member* KGVC yang masih pelajar (*students*) justru lebih banyak berada pada segmen *dormants customer*.

Segmentasi Pelanggan berdasarkan Pembelian Buku

Dataset yang digunakan untuk segmentasi pelanggan berdasarkan pembelian buku adalah sebanyak 28 kolom, yaitu kode pelanggan, dan 27 jenis buku yang masing-masing berupa *categorical data*. Jenis buku tersebut adalah *agriculture*, *art*, *Archi-*



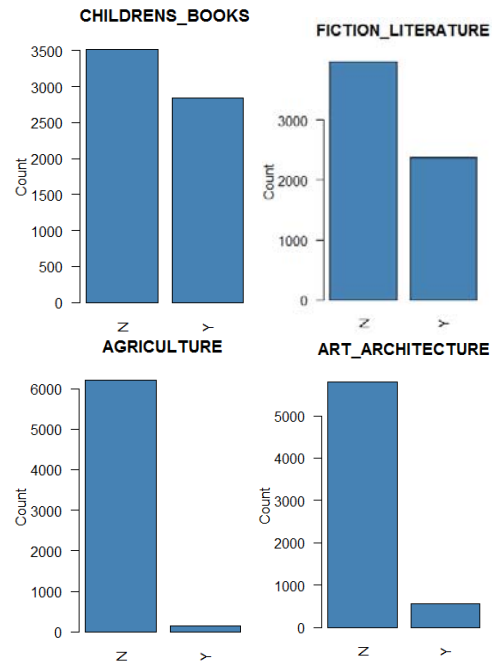
Gambar 12. Persentase Jumlah *Member* KGVC berdasarkan Jenis Pekerjaan pada Segmentasi RFM

ecture & photography, *business & economics*, *children`s books*, *computing & internet*, *cooking*, *diet & health*, *education & teaching*, *engineering*, *entertainment*, *fiction & literature*, *home & garden*, *law*, *magazines.tabloid and journal*, *medical*, *others*, *parenting & family*, *philosophy*, *psychology*, *reference & dictionary*, *religion & spirituality*, *schoolbooks indonesia curriculum*, *schoolbooks singapore curriculum*, *science & nature*, *self-improvement*, *social sciences*, dan *sports & adventure*. Contoh *dataset* yang digunakan pada penelitian ini seperti yang ditunjukkan pada TABEL 3.

Data yang digunakan adalah variabel Y dan N seperti ditunjukkan Gambar 11. Y adalah jumlah pelanggan yang membeli buku jenis tertentu, sedangkan N adalah jumlah pelanggan yang tidak membeli buku jenis tertentu.

Sebelum dilakukan *clustering*, terlebih dahulu dihitung nilai varian pada *dataset* yang terdiri dari *categorical data*. Berdasarkan *elbow method*, seperti terlihat pada Gambar 13, nilai $k=3$ dengan persentase varian = 4,63 persen adalah *cluster* yang optimal.

Visualisasi hasil pembentukan *cluster* menggunakan algoritma *hierarchical clustering* dapat dilihat pada Gambar 14. Hasil *cluster* yang optimal pada segmentasi pembelian buku adalah 3 *cluster*. TABEL 4 menunjukkan persentase jumlah pelanggan, dan nilai *mean* setiap *cluster* yang terbentuk dari hasil segmentasi pembelian buku pada 3 *cluster*. *Cluster* 1 dengan label pelanggan pembelian



Gambar 11. Contoh Distribusi Pelanggan berdasarkan Jenis Buku

rendah terdiri dari 82,80 persen pelanggan dan rata-rata pembelian 2 jenis buku. *Cluster 2* dengan label pelanggan pembelian sedang terdiri dari 16,66 persen pelanggan dan rata-rata pembelian 5 jenis buku. *Cluster 3* dengan label pelanggan pembelian tinggi terdiri dari 0,54 persen pelanggan dan rata-rata pembelian 7 jenis buku.

Persentase jumlah *member KGVC* berdasarkan umur pada segmentasi pembelian buku ditunjukkan pada Lampiran Gambar I. Pelanggan berumur 41-50 tahun merupakan mayoritas pelanggan yang berada pada segmen pelanggan dengan pembelian jenis buku yang “tinggi”, yaitu sebesar 32,35 persen. Pelanggan berumur > 60 tahun merupakan pelanggan yang tidak diketahui umur sebenarnya. Pelanggan tersebut merupakan mayoritas pelanggan yang berada pada segmen pelanggan dengan pembelian jenis buku “rendah” dan “sedang”. Persentase jumlah pelanggan pada cluster “rendah” sebesar 33,59 persen, sedangkan persentase jumlah pelanggan pada cluster “sedang” sebesar 29,87 persen.

Persentase jumlah *member KGVC* berdasarkan umur pada segmentasi pembelian buku ditunjukkan pada Lampiran Gambar II. *Member KGVC* yang termasuk dalam segmen pelanggan dengan pembelian buku “rendah” dan “tinggi” adalah mayoritas pelanggan dengan jenis kelamin perempuan. Persentase jumlah pelanggan pada *cluster* “rendah” sebesar 62,93 persen, sedangkan persentase jumlah pelanggan pada *cluster* “tinggi” sebesar 52,94 persen. *Member KGVC* yang termasuk dalam segmen pelanggan dengan pembelian buku “sedang” adalah mayoritas pelanggan laki-laki dengan persentase 52,08 persen. Berdasarkan perbandingan ini menunjukkan bahwa *member KGVC* berjenis kelamin perempuan lebih potensial dibandingkan pelanggan laki-laki.

Persentase jumlah *member KGVC* berdasarkan tingkat pendidikan pada segmentasi pembelian buku seperti ditunjukkan pada Lampiran Gambar

III. *Member KGVC* dengan tingkat pendidikan universitas mendominasi masing-masing pelanggan pada *cluster* pembelian buku “rendah”, “sedang”, dan “tinggi”. Persentase pelanggan dengan tingkat pendidikan universitas pada segmen “tinggi” memiliki persentase paling tinggi, yaitu 70,59 persen. Tetapi, pelanggan dengan tingkat pendidikan SMP dan SMA tidak termasuk pada segmen pelanggan “tinggi”.

Persentase jumlah *member KGVC* berdasarkan jenis pekerjaan pada segmentasi pembelian buku ditunjukkan pada Lampiran Gambar IV. Pelanggan yang tidak diketahui jenis pekerjaan (*others*) mendominasi masing-masing *cluster*, yaitu *cluster 1* sebanyak 51,99 persen, *cluster 2* sebanyak 49,81 persen, dan *cluster 3* sebanyak 50,00 persen. Sementara itu, persentase *member KGVC* yang masih pelajar (*students*) tidak termasuk pada *cluster 3*.

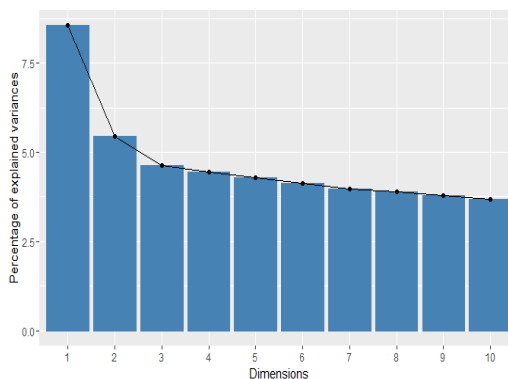
Rekomendasi Strategi Marketing

Berdasarkan segmentasi pelanggan yang dihasilkan, diperlukan strategi marketing untuk meningkatkan profit perusahaan dan loyalitas pelanggan. Rekomendasi strategi marketing yang diusulkan oleh penulis adalah segmen pelanggan berdasarkan nilai RFM. Hal ini dikarenakan variabel yang diambil terdiri dari kapan pelanggan terakhir kali berbelanja, frekuensi kunjungan, dan nilai transaksi buku yang dibeli. Penulis tidak merekomendasikan strategi marketing terhadap segmen pelanggan berdasarkan pembelian jenis buku dikarenakan segmen pelanggan ini hanya berdasarkan jenis buku yang dibeli oleh pelanggan. Berikut penjelasan lebih lanjut terkait beberapa rekomendasi strategi marketing untuk *occasional customers* dan *dormant customers*.

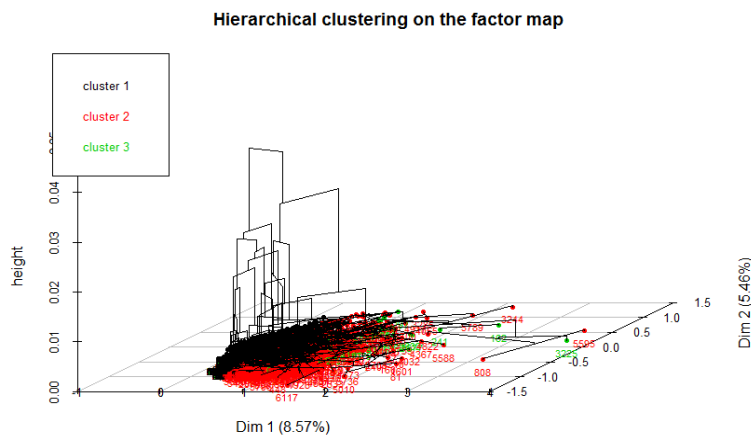
Occasional Customers

Pelanggan *occasional* merupakan pelanggan yang memiliki *KGVC* dengan frekuensi kunjungan yang jarang tetapi memiliki nilai transaksi rata-rata yang tinggi. Gramedia perlu mengembangkan aktivitas promosi untuk meningkatkan frekuensi kunjungan pelanggan pada *cluster* ini. Selain itu, loyalitas pelanggan harus tetap dijaga dan ditingkatkan dengan memberikan benefit terhadap pelanggan setiap kali pelanggan tersebut melakukan kunjungan. Strategi marketing tersebut misalnya dengan memberikan pelanggan *reward*, yaitu diskon atau hadiah tertentu setelah mengumpulkan struk transaksi dengan nilai rupiah tertentu pada periode tertentu.

Peningkatan nilai transaksi pada pelanggan *occasional* dapat dilakukan dengan mengembangkan berbagai aktivitas promosi. Informasi terkait jenis buku baru yang sering dibeli oleh pelanggan *KGVC* perlu diinformasikan oleh Gramedia kepa-



Gambar 13. Persentase Varian pada Setiap Dimensi berdasarkan Dataset Buku



Gambar 14. Segmentasi Pembelian Buku menggunakan *Hierarchical Clustering*

da pelanggan secara periodik. Selain itu, peningkatan nilai transaksi ini dapat dikembangkan dengan metode *data mining* dalam *cross-/up-selling campaign* dan *market basket analysis*. *Cross-/up-selling campaign* bertujuan untuk menjual buku tambahan atau alternatif buku yang lebih memberikan benefit yang lebih banyak kepada pelanggan. *Market basket analysis* bertujuan untuk menawarkan buku yang biasanya dibeli secara bersama-sama oleh pelanggan.

Dormant Customers

Pelanggan *dormant* merupakan pelanggan yang kurang potensial karena memiliki nilai RFM paling rendah. Aktivitas promosi untuk mempertahankan pelanggan yang kurang potensial memang tidaklah mudah. Gramedia perlu mengidentifikasi terlebih dahulu pelanggan KGVC yang lebih memberikan kontribusi dan dapat ditingkatkan tingkat kunjungannya. Hal ini dapat dilakukan dengan menganalisis faktor-faktor yang menyebabkan pelanggan pada segmen ini kurang potensial. Pelanggan *dormant* perlu diberikan perhatian dan Gramedia harus meningkatkan komunikasi untuk mempertahankan pelanggan supaya tetap setia berbelanja menggunakan KGVC di Gramedia. Misalnya, pemberian diskon khusus bagi pelanggan KGVC tertentu untuk meningkatkan frekuensi dan minat berbelanja.

5. Kesimpulan

Kesimpulan yang diperoleh berdasarkan analisis dan pembahasan adalah bahwa segmentasi pelanggan yang dihasilkan pada CRM di PT Gramedia Asri Media sebanyak 2 cluster, yaitu *occasional customers*, dan *dormant customers*. *Occasional customers* merupakan pelanggan yang jarang ber-

kunjung tetapi memiliki nilai transaksi rata-rata yang tinggi. *Dormant customers* merupakan pelanggan yang tidak potensial. Sedangkan jumlah cluster optimal berdasarkan pembelian buku sebanyak tiga cluster.

Jumlah *cluster* optimal pada segmentasi nilai pelanggan yang dihasilkan dalam penelitian ini berdasarkan pada nilai tertinggi *silhouette method* dan *calinski-harabasz index*. Nilai *average silhouette* pada dua cluster adalah 0.42. Nilai *Calinski-Harabasz index* pada dua cluster adalah 1.31.

Faktor penting yang harus diperhitungkan ketika mengembangkan strategi pemasaran adalah segmentasi pelanggan. Segmentasi pelanggan ini merupakan langkah awal dalam menganalisis pelanggan supaya dapat meningkatkan transaksi buku yang dilakukan oleh member KGVC. Segmentasi pelanggan yang dihasilkan dalam penelitian ini adalah segmentasi nilai pelanggan dan segmentasi pembelian buku. Dengan adanya segmentasi ini diharapkan dapat membantu perusahaan dalam mengelompokkan member KGVC sehingga perusahaan dapat menentukan strategi yang sesuai untuk setiap kelompok pelanggan.

Penelitian ini masih memiliki kekurangan yang dapat disempurnakan pada penelitian selanjutnya. Berdasarkan hasil penelitian dan kesimpulan yang telah dibuat, penulis memberikan saran untuk penelitian selanjutnya. Segmentasi pelanggan pada penelitian ini menggunakan satu algoritma *clustering*, yaitu algoritma *K-means*. Penelitian selanjutnya dapat bereksperimen menggunakan berbagai algoritma *clustering*.

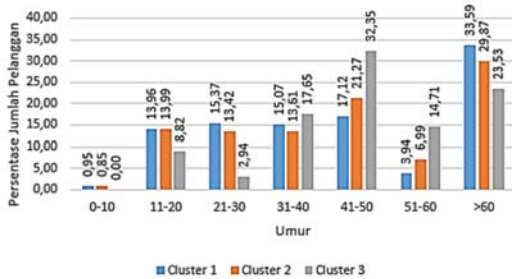
Penelitian selanjutnya dapat dikembangkan dengan metode *data mining* yang lain untuk melakukan *cross-/up-selling campaign* dan *market basket analysis*. *Cross-/up-selling campaign* digunakan untuk menjual produk tambahan atau alternatif produk yang lebih menguntungkan pelang-

gan. *Market basket analysis* digunakan untuk mengidentifikasi produk-produk terkait yang biasanya dibeli secara bersama-sama oleh pelanggan.

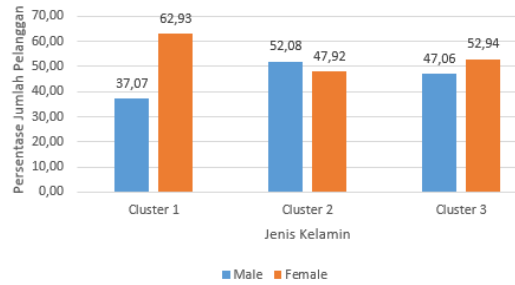
Referensi

- [1] Kompas Gramedia. (2016). *Data Warehouse*. Jakarta: Kompas Gramedia.
- [2] Berry, M. J., & Linoff, G. S. (2004). *Data Mining Techniques for Marketing, Sales, and Customer Relationship Management Second Edition*. Indianapolis, Indiana: Wiley Publishing.
- [3] Tsipstis, K., & Chorianopoulos, A. (2009). *Data Mining Techniques in CRM: Inside Customer Segmentation*. West Sussex: John Wiley & Sons Ltd.
- [4] Maheshwari, A. K. (2015). *Business Intelligence and Data Mining*. New York: Business Expert Press.
- [5] Larose, D. T., & Larose, C. D. (2015). *Data Mining and Predictive Analytics*. New Jersey: John Wiley & Sons.
- [6] Kodinariya, T. M., & Makwana, P. R. (2013). Review on Determining Number of Cluster in K-Means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*, 90-95.
- [7] Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. (2010). Understanding of Internal Clustering Validation Measures. *International Conference on Data Mining* (pp. 911-916). Beijing: IEEE.
- [8] Celebi, M. E., & Aydin, K. (2016). Clustering Evaluation in High-Dimensional Data. In N. Tomašev, & M. Radovanović, *Unsupervised Learning Algorithms* (pp. 71-107). Switzerland: Springer International Publishing.
- [9] Zume, N., & Mount, J. (2014). *Practical Data Science with R*. Shelter Island: Manning Publications Co.

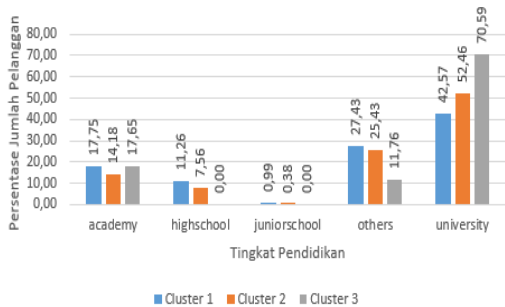
Lampiran



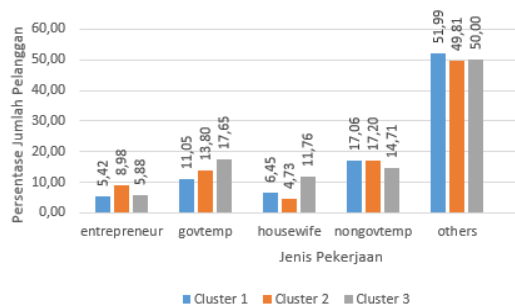
Lampiran Gambar I. Persentase Jumlah Member KGVC berdasarkan Umur pada Segmentasi Pembelian Buku



Lampiran Gambar II. Persentase Jumlah Member KGVC berdasarkan Jenis Kelamin pada Segmentasi Pembelian Buku



Lampiran Gambar III. Persentase Jumlah Member KGVC berdasarkan Tingkat Pendidikan pada Segmentasi Pembelian



Lampiran Gambar IV. Persentase Jumlah Member KGVC berdasarkan Jenis Pekerjaan pada Segmentasi Pembelian Buku