

Segmentasi Pelanggan Menggunakan Algoritma *K-Means* Sebagai Dasar Strategi Pemasaran pada LAROIBA Seluler

Oleh: Mike Indra Istiana

Jurusan Sistem Informasi Fakultas Ilmu Komputer Universitas Dian Nuswantoro Semarang

Email: istiana.mike@gmail.com

Abstrak

Penelitian ini bertujuan untuk melakukan segmentasi pelanggan pada LAROIBA Seluler dengan memanfaatkan proses data mining dengan teknik *Clustering*. Metode yang digunakan adalah CRISP-DM dengan melalui proses *business understanding*, *data understanding*, *data preparation*, *modeling* dan *evaluasi*. Algoritma yang digunakan untuk pembentukan kluster adalah algoritma K-Means. K-Means menghasilkan model visual cluster dengan tools Rapidminer 5.2 yang merepresentasikan jumlah pelanggan di setiap kluster dengan menggunakan atribut RFM (Recency, Frequency dan Monetary). Kelompok pelanggan yang terbentuk adalah empat kelas dengan kluster pertama 4 pelanggan, kluster kedua 76 pelanggan, kluster ketiga 20 pelanggan dan kluster keempat dengan jumlah pelanggan 173.

Kata Kunci: Segmentasi, *Clustering*, *Data mining*, *Business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluasi*, *K-Means*, *Rapidminer*, *RFM*

Abstract

This research aims to perform customer segmentation on LAROIBA Celuler by utilizing data mining process with Clustering techniques. The method used is the CRISP-DM with through the process of business understanding, data understanding, data preparation, modeling, deployment and evaluation. The algorithm used to perform the clusters is K-Means algorithm. Rapidminer 5.2 used as a tools to Clustering with K-Means algorithm that represent the number of customers in each cluster by using attributes RFM (Recency, Frequency and Monetary). The groups of LAROIBA's customers perform 4 cluster as result with 4 customers in the first class, 76 customers for second cluster, 20 customers in the thirth cluster and the last is 173 customers in the fourth cluster.

Keyword: *segmentation*, *Clustering*, *Data mining*, *Business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, *K-Means*, *Rapidminer*, *RFM*

Pendahuluan

Perkembangan di bidang telekomunikasi yang dialami Indonesia membawa pengaruh kepada maraknya penggunaan telepon seluler karena alasan fleksibilitas dan mendukung mobilitas yang tinggi. (Erlinda, et. al. 2010)

Berdasarkan penelitian ATSI, pada tahun 2010 pengguna telepon seluler sebesar 180 juta dan diprediksikan akan meningkat menjadi 240 juta pada tahun 2011. Dari jumlah tersebut 95% merupakan pengguna prabayar. Peningkatan jumlah pengguna telepon seluler yang sangat pesat dan cenderung terus meningkat ini mendorong berkembangnya bisnis di masyarakat dalam bidang penjualan pulsa baik berupa *voucher* fisik maupun *voucher* elektrik.

Dalam kurun waktu beberapa tahun terakhir, muncul kios-kios pedagang kecil di pinggir jalan maupun pelosok kampung, bahkan supermarket dan minimarket pun ikut terjun di bisnis ini. (Kompasiana) Namun, mempertahankan bisnis penjualan pulsa tidaklah semudah saat memulainya, saat ini banyak *counter* pulsa yang mengalami kebangkrutan. Berdasarkan pendapat konsultan bisnis usaha server pulsa elektrik Triviji, setelah mengalami peningkatan yang cukup tajam dari beberapa server pulsa yang ada pada tahun 2005, saat ini server pulsa mencapai sekitar 2500 server. Dari 2500 server tersebut diperkirakan hanya 35% saja yang dapat bertahan sampai tahun 2013. Masalah utamanya adalah faktor manajemen dan kurangnya strategi pemasaran. (Reni, 2010)

LAROIBA Seluler sebagai perusahaan dengan kegiatan utamanya sebagai distributor pulsa elektrik di wilayah Kabupaten Boja. Setelah berkecimpung di usaha pengisian pulsa sejak tahun 2005, saat ini LAROIBA mampu menangani lebih dari 1000 transaksi perhari. Dalam menghadapi persaingan dan agar mampu mempertahankan eksistensinya di era pemasaran modern saat ini, yang harus menjadi fokus utama perusahaan adalah pelanggan. CRM adalah salah satu strategi bisnis untuk meningkatkan keuntungan, penghasilan dan kepuasan konsumen dengan cara mengatur segmentasi konsumen, cara memberikan perlakuan terhadap konsumen dan mengimplementasikan proses *customer-centric*. (Buttle, F, 2009)

Segmentasi berdasarkan RFM telah digunakan lebih dari lima puluh tahun yang lalu untuk menargetkan bagian-bagian pelanggannya, meminimalkan biaya perpesanan, dan meningkatkan profitabilitas (Ron, Rajesh,. 2004). Untuk melakukan pengelompokan pelanggan, setiap pelanggan dinilai profitabilitasnya terhadap perusahaan dari transaksi yang telah dilakukan dengan metode RFM (*Recency, Frequency, and Monetary*). *Recency*

merupakan jumlah pembelian sejak transaksi terakhir dilakukan, *Frequency* berkaitan dengan besarnya tingkat transaksi yang dilakukan oleh pelanggan dalam kurun waktu tertentu, dan *Monetary* merupakan besarnya nilai transaksi yang dilakukan oleh pelanggan dalam kurun waktu tertentu.

Landasan Teori

Penelitian di bidang Segmentasi pelanggan telah banyak dilakukan peneliti di dunia, hal ini dapat diketahui dari jumlah jurnal yang dipublikasikan pada IEEE. Pada tanggal 25 Juni 2013 dengan *Keyword* “Customer Segmentation” didapatkan data penelitian berjumlah 271 artikel.

Shin dan Sohn (2004) dalam penelitiannya menggunakan algoritma *KMeans*, *SOM*, dan *Fuzzy C-Means* untuk mencari segmentasi konsumen pasar modal berdasarkan nilai potensialnya, yaitu berdasarkan total nilai transaksi yang dilakukan dalam periode tiga bulan. Dari hasil perbandingan terhadap ketiga metode tersebut, algoritma *K-Means* memiliki hasil segmentasi yang lebih akurat.

Customer Relationship Management

CRM diartikan sebagai satu kesatuan penjualan, pemasaran dan strategi pelayanan yang mencegah terjadinya aktivitas pekerjaan yang tidak terkoordinasi antar bagian dengan baik dan itu tergantung pada aksi – aksi perusahaan yang terkoordinasi (Kalakota, 2001).

Di dalam pengelolaan hubungan pelanggan terdapat empat hal yang menjadi fokus utama. Keempat hal tersebut yaitu *customer identification*, *customer attraction*, *customer retention* dan *customer development*.

Segmentasi Pelanggan

Don Pepper dan Martha Roger dalam bukunya “Managing Customer Relationship : A Strategic Framework”. Melakukan kategorisasi pelanggan sebagai berikut:

- *Most Valuble Customer* (MVC), yaitu pelanggan dengan nilai paling tinggi bagi perusahaan. Merupakan pelanggan yang memberikan keuntungan terbesar bagi perusahaan.
- *Most Growable Customer*, yaitu pelanggan yang tanpa disadari memiliki potensi besar.
- *Below Zeros*, yaitu pelanggan yang memberikan keuntungan lebih sedikit daripada biaya untuk memberikan pelayanan.

- *Migrators*, yaitu pelanggan yang berada pada posisi *diantara below zeros* dan *most growable customer* dan perlu dianalisis agar dapat diketahui kategori asalnya (Pepper, D & Roger, M., 2001).

Model RFM

Model RFM diaplikasikan secara luas pada database pemasaran dan merupakan tool yang umum digunakan untuk membangun strategi pemasaran. (Wei, et al., 2010). RFM berdasarkan segmentasi pelanggan menghasilkan kemampuan segmentasi antara 75% sampai 85%. (Hughes, 2000).

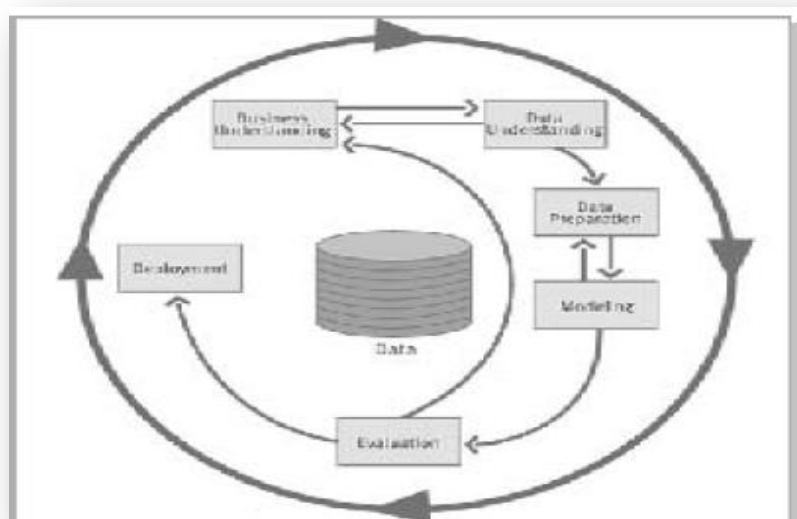
Berikut ini adalah penjelasan mengenai RFM menurut C.H. Cheng, Y.S Chen (2009):

1. Recency, yaitu kapan transaksi terakhir dilakukan
2. Frequency, tingkat keseringan pelanggan melakukan transaksi. Misalkan sekali transaksi tiap bulan, atau 2 kali dalam satu tahun
3. Monetary, besarnya nilai transaksi yang dilakukan

Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database.

Model CRISP-DM



Gambar Metode CRISP-DM

Algoritma K-Means

Metode ini termasuk teknik penyekatan (*partition*) yang membagi atau memisahkan objek ke k daerah bagian yang terpisah. Pada k-means, setiap objek harus masuk dlm kluster tertentu, tetapi dalam satu tahapan berikutnya objek akan berpindah ke kluster lain.

Prosedur dalam algoritma K-Means dijelaskan pada tahap berikut:

Langkah 1 : Kelompokkan item dalam dataset de dalam K cluster

Langkah 2 : Menentukan secara acak nilai *centroid*

Langkah 3 : Memasukkan setiap item dataset yang jaraknya paling dekat dengan nilai *centroid* ke dalam *centroid* cluster tersebut.

Langkah 4 : Menghitung rata-rata nilai item dalam setiap cluster untuk dijadikan sebagai *centroid* yang baru.

Langkah 5 : Melakukan pengulangan langkah 2 da langkah 3 hingga nilai *centroid* sama dengan nilai rata-rata item dalam cluster

Euclidean distance

Perhitungan jarak antar titik dengan menggunakan *Euclidean distance* dapat dilakukan pada titik dalam satu dimensi, dua dimensi, maupun 3 dimensi. Formula Jarak antar dua titik dalam satu, dua dan tiga dimensi secara berurutan ditunjukkan pada formula 1,2,3 berikut ini.

$$\sqrt{(x - y)^2} = |x - y| \quad (2-1)$$

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} \quad (2-2)$$

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2} \quad (2-3)$$

Metode Penelitian

Metode CRISP-DM

Rancangan tahapan penelitian mengacu pada enam tahap CRISP-DM, atau siklus hidup pengembangan data mining sebagai *framework* dari proyek *data mining* (Chapman, 2000).

1. *Bussiness Understanding*

Pada fase ini ada empat tahap yang dilakukan, yaitu :

a. Pemahaman terhadap tujuan bisnis.

Meningkatkan dan mempertahankan jumlah konsumen terutama konsumen yang berpotensi untuk menjadi konsumen tetap (pelanggan).

b. Penilaian situasi (*situation assessment*).

LAROIBA Seluler merupakan suatu perusahaan yang kegiatan utamanya sebagai distributor pulsa elektrik, selain itu LAROIBA juga menyediakan handphone, kartu perdana dan aksesoris handphone sebagai pelengkapannya. LAROIBA belum melakukan pemanfaatan data pelanggan maupun data transaksi secara baik, bahkan data-data transaksi penjualan pada LAROIBA dihapus dari server tanpa adanya back up data dengan alasan akan membebani server. Saat ini LAROIBA telah menangani lebih dari 1000 transaksi penjualan pulsa elektrik.

c. Menerjemahkan tujuan bisnis ke dalam tujuan *data mining*.

Customer segmentation dapat digunakan oleh pihak manajemen untuk menemukan segmen-segmen konsumen yang bertujuan untuk :

Membangun profil dari konsumen (*customer profiling*) yang terkait dengan *purchase history* dan menentukan pemasaran yang tepat pada setiap segmen yang terbentuk.

2. *Data Understanding*

Pada proses ini akan dilakukan pemahaman terhadap kebutuhan data terkait dengan pencapaian tujuan dalam CRM dan *data mining*.

Penelitian ini dilakukan pada salah satu Distributor pulsa di daerah Kabupaten Boja. Obyek penelitian dalam kasus ini digunakan untuk melakukan segmentasi pelanggan pembeli pulsa elektronik yang tersimpan dalam database server penjualan pulsa di LAROIBA Seluler. Data mining yang digunakan untuk segmentasi adalah dengan algoritma K-Means. Data yang digunakan adalah data transaksi selama 30 hari di bulan Juni 2013 dengan format Rtf.

3. *Data Preparation*

Pada tahap ini struktur basis data akan dipersiapkan sehingga mempermudah proses mining. Proses preparation ini mencakup tiga hal utama yaitu:

- a. *Data Selection*: Memilih data yang akan digunakan dalam proses data mining. Dalam proses ini dilakukan juga pemilihan atribut-atribut yang disesuaikan dengan proses data mining.
- b. *Data Preprocessing*: Memastikan kualitas data yang telah dipilih pada tahap data selection, pada tahap ini masalah yang harus dihadapi adalah noisy data dan missing values. Proses pembersihan data (*cleansing*) dilakukan dengan melakukan metode-metode query sederhana untuk menemukan anomali-anomali data yang bisa saja masih terdapat pada sistem.
- c. *Data Transformation*: Mengelompokkan atribut-atribut atau field-yang telah terpilih menjadi 1 tabel dengan cara melakukan denormalisasi dari basis data OLTP.

4. Pemodelan (*modeling*)

Metode yang digunakan yaitu algoritma *K-Means*. Untuk menghitung data dalam penelitian ini akan menggunakan *framework* RapidMiner versi 5.2 sehingga akan ditemukan nilai akurat.

5. Evaluasi

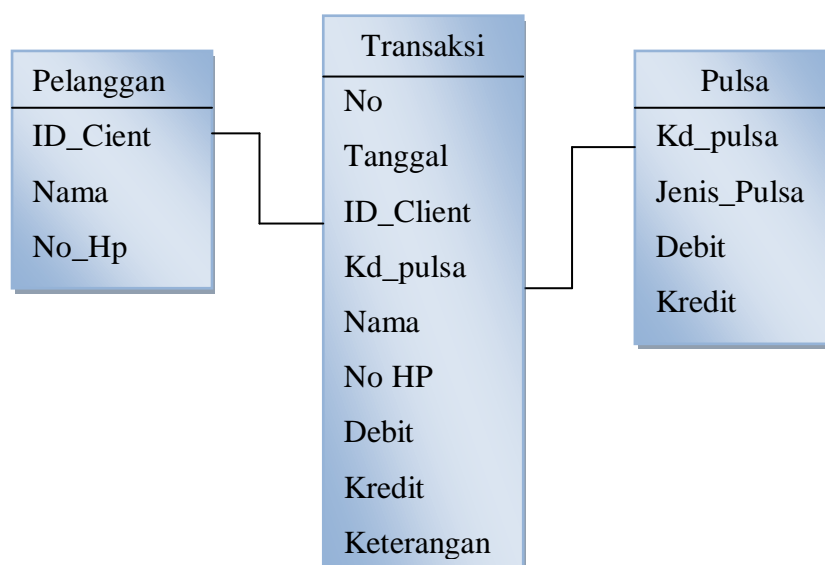
Berdasarkan hasil clustering dengan metode Rapidminer versi 5.2 dilakukan evaluasi kluster yang terbentuk dan diselaraskan dengan tujuan penelitian yang dilakukan.

6. Strategi Pemasaran

Strategi pemasaran LAROIBA Celuler didefinisikan dan diterapkan ke anggota dari setiap kelas yang terbentuk sebagai hasil klustering.

Pembahasan

Pembuatan Database dengan model Star



Berdasarkan relasi tabel pada LAROIBA Seluler, dalam penelitian dengan menggunakan model RFM akan digunakan data-data transaksi yang tersimpan dalam database perusahaan. Dalam Kasus ini, untuk mengakomodasi keperluan segmentasi dengan teknik clustering yang akan digunakan adalah tabel transaksi Laroiba Seluler. Berikut ini merupakan atribut yang akan digunakan pada proses data mining.

Tabel Pemilihan Atribut

Field	Keterangan
ID Client	Kode Pelanggan
Tanggal	Tanggal transaksi pembelian terakhir yang dilakukan pelanggan Digunakan untuk memodelkan <i>Recency</i> dan <i>Frequency</i> . <i>Recency</i> , dalam waktu 30 hari dalam bulan Juni 2013 berapa kali pelanggan melakukan transaksi dengan Laroiba Seluler. <i>Frequency</i> merupakan jumlah transaksi yang dilakukan oleh pelanggan dalam jangka waktu 30 hari dalam bulan Juni 2013.
Kredit	Untuk memodelkan atribut Monetary, yaitu dengan cara menjumlahkan semua transaksi yang dilakukan pelanggan dalam waktu 30 hari dalam bulan Juni 2013.

Analisis RFM

Langkah Selanjutnya yaitu memetakan R,F,M setiap pelanggan dalam waktu satu bulan. Setiap pelanggan dihitung R, F dan M dalam akumulasi waktu satu bulan seperti terlihat pada tabel 4.9 Dibawah ini

Tabel Tabel RFM Pelanggan

▲ Kd_plg	R	F	M
a001	28	380	4273.0
a002	28	310	3658.3
a003	28	336	366
a004	28	307	2656.0
a005	26	73	636.5
a006	24	97	990.3

Setelah diketahui nilai RFM dari setiap pelanggan, kemudian dilakukan pembobotan RFM untuk memudahkan analisa yaitu dengan cara membagi nilai R, F dan M masing-masing menjadi 5 bagian. Pembobotan atribut R, F, dan M dapat dilihat pada tabel berikut:

Tabel 4.11 Pembobotan Nilai Atribut RFM

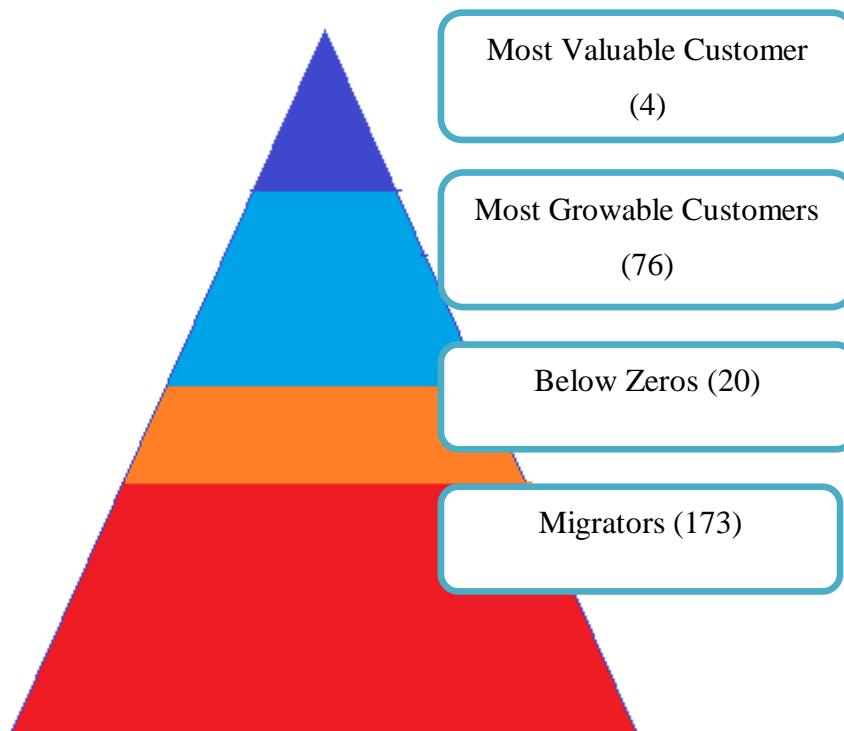
Bobot	R	F	M
1	1-5	1-200	0-4000
2	6-10	201-400	4001-8000
3	11-16	401-600	8001-12000
4	17-22	601-800	12001-16000
5	23-28	801-1000	16001-20000

Analisis dengan Rapidminer

Tabel Centroid Cluster

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
att2	27	27.700	17.457	26.815
att3	455.250	316.350	50.671	171.113
att4	11624.722	3829.460	519.194	1791.587

Hasil Segmentasi Akhir



PENUTUP

5.1 Kesimpulan

1. Setelah dilakukan segmentasi data pelanggan LAROIBA Celuler dengan menggunakan algoritma *K-Means* terbentuk empat cluster yang memenuhi kelas pada Piramida Pelanggan yaitu *Most Valuable Customer* ada 4 pelanggan, *Most Growable Customer* ada 76 pelanggan, *Below Zero* 20 pelanggan dan *Migrators* ada 173 pelanggan.
2. Strategi Potongan harga 10% yang digunakan untuk meningkatkan penerimaan perusahaan dari pelanggan dapat diterapkan pada kelas pertama dan kedua yaitu *Most Valuable Customer* dan *Most Growable Customer* dengan cara mengirimkan SMS (*Short Message Service*) atau pemberitahuan informasi ke nomor pelanggan dalam database pelanggan LAROIBA Celuler dengan id pelanggan tersebut.

DAFTAR PUSTAKA

1. Berry, Michael J.A dan Linoff, Gordon S. (2000), *Mastering Datamining*, John Wiley & Sons, Inc., New York. *Systems with Applications*, Vo.36, hal. 4176-4184
2. Kotler, Philip dan Keller, Kevin L. (2009), *Manajemen Pemasaran edisi ke-13 jilid 1*, Penerbit Erlangga, Jakarta
3. Ngai, E.W.T., Xiu, Li dan Chau, D.C.K. (2009), “Application Of Data Mining Technique In Customer Relationship Management: A Literature Review And Classification”, *Expert System with Applications*, Vol.36, hal. 2592-2602.
4. Tan, Steinbach, Kumar (2004), *Introduction to Data Mining: Chapter-1*, Lecture handouts: Department of Statistics, Standford University.
5. Arthur M. Hughes, *Strategic Database Marketing*, 2nd edition, 2000.
6. Kohavi, Ron dan Parekh, Rajesh. (2004). *Visualizing RFM Segmentation*.
7. Wei Jo-Ting , Lin Shih-Yen and Wu Hsin-Hung , 2010. *A review of the application of RFM model* . *African Journal of Business Management* Vol. 4(19), Pp. 4199-4206.
8. Larose , Daniel T, 2005, *Discovering Knowledge in Data: An Introduction to Data Mining*, John Willey & Sons. Inc
9. Agusta Y, 2007. K-Means-Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika Vol 3*. STIMIK. Bali.
10. Erlinda, dkk.2010. *Analisis Industri Telekomunikasi di Indonesia*. Volume 9 Number 1 Departemen Teknik Industri, Fakultas Teknik UI)
11. <http://Wikipedia.co.id> diakses 20 Agustus 2013 15:15
12. <Http://kompas.com/>, diakses 20 Agustus 2013 22:56
13. Choiriah, Reni R. 2010. *Penerapan Pencatatan Keuangan Pada Industri Kecil Rumahan*. Fakultas Ekonomi Universitas Pembangunan Nasional “Veteran”, Jawa Timur.
14. Cheng, Ching-Hsue dan Chen, You-Shyang (2009), “*Classifying the segmentation of customer value via RFM model and RS Theory*”, *Expert*
15. <http://ieeexplore.ieee.org/Xplore/home.jsp>,