

Model Segmentasi Pelanggan Dengan Kernel K-Means Clustering Berbasis Customer Relationship Management

Abdul Haris Lubis

Politeknik Ganesha

Jl. Veteran No. 190 Pasar VI Manunggal

E-mail: haris_lubis@yahoo.com

Abstrak — Jurnal ini mengusulkan sebuah model aturan dalam menentukan pelanggan terbaik dan potensial Kantor Pos Medan. Hingga saat ini dalam menentukan pelanggan terbaik dan potensial menjadi persoalan di Instansi-instansi baik pemerintah maupun swasta khususnya yang bergerak dibidang jasa. Faktor-faktor yang berpengaruh secara dominan dalam menentukan pelanggan terbaik dan potensial masih belum dapat ditentukan secara pasti. Saat ini manajemen Kantor Pos Medan masih menggunakan secara manual dalam menentukan pelanggan terbaiknya, sehingga sangat mungkin terjadi kesalahan pada prosedur yang sudah berjalan. Hal ini akan berpengaruh terhadap hasil keputusan yang akan diambil oleh pihak manajemen Kantor Pos Medan. Untuk itu sangat penting dibuat sebuah model aturan untuk menentukan pelanggan terbaik dan potensial yang dapat digunakan pihak manajemen sebagai sistem pendukung dalam pengambilan keputusan. Data yang digunakan dalam penelitian ini berasal dari database Kantor Pos Medan tahun 2011 – bulan maret 2013. Dalam jurnal ini algoritma Kernel K-Means Clustering telah digunakan untuk mendapatkan suatu model aturan menentukan pelanggan terbaik dan potensial Kantor Pos Medan. Model aturan yang diperoleh menunjukkan bahwa katagori pelanggan terbaik dapat diperoleh jika transaksinya banyak dan besar uangnya sedang dan tinggi.

Kata Kunci — *kernel k-means clustering, model aturan, menentukan*

I. PENDAHULUAN

Ketika suatu organisasi tumbuh semakin besar dan pola tingkatan operasionalnya semakin tidak sederhana dan kompleks, maka secara alamiah tuntutan pihak manajemen akan kebutuhan dan fungsional dari setiap system informasi semakin besar, khususnya pada *fungsionalitas* data, teknologi dan aplikasi.

Suatu organisasi yang berusaha membangun arsitektur-nya terkadang tidak dapat memahami hasil perencanaan sistem informasi dalam bisnisnya, hanya karena bahasa dan model yang digunakan cenderung rumit dan bersifat sangat teknis sehingga sulit dimengerti. Akibatnya tidak ada umpan balik darinya yang justru merupakan faktor utama dalam perencanaan tadi. Keberhasilan pengimplementasian sangat tergantung pada pemahaman dari setiap entitas bisnis itu sendiri termasuk komitmen manajemen organisasi tersebut.

Customer Relationship Management (CRM), dalam Bahasa Indonesia menjadi manajemen hubungan pelanggan. Konsep CRM merupakan spesifikasi dari

konsep *Relationship Marketing* (RM). Konsep *relationship marketing* (pemasaran hubungan) menekankan bahwa perusahaan harus berinteraksi dan menjalin hubungan dengan berbagai pihak yang berkepentingan terhadap perusahaan (*stakeholders*), karena hubungan yang baik merupakan asset yang paling mendasar bagi suatu perusahaan. Adapun konsep CRM lebih menekankan pada menjalin hubungan baik dengan pelanggan sebagai salah satu *stakeholders*, karena pelanggan dianggap sebagai ujung tombak suatu bisnis.

Paradigma baru dunia bisnis menunjukkan bahwa konsumen berperan aktif dalam menciptakan nilai. Hal ini menyebabkan meningkatnya peran konsumen sebagai sumber kompetensi baru bagi perusahaan. Kompetensi yang dimiliki konsumen adalah pengetahuan, keahlian, kemauan untuk belajar dan bereksperimen, serta kemampuan terlibat komunikasi yang aktif dengan perusahaan. (Gaffar, 2007). Hal tersebut dilakukan agar perusahaan dapat menciptakan produk atau jasa yang sangat sesuai dengan yang dibutuhkan oleh konsumen. Untuk itu dibutuhkan adanya saluran komunikasi yang

baik yang memungkinkan adanya interaksi yang intensif antara perusahaan dengan konsumen.

Dengan demikian menurut Gray dan Byun dalam Gaffar (2007:41) keseluruhan proses dan aplikasi dari CRM berdasarkan kepada prinsip-prinsip dasar yaitu perlakuan pelanggan secara individu. Jadi CRM didasarkan pada filosofi *personalisasi*, yaitu tujuan dan pelayanan kepada pelanggan harus dirancang berdasarkan preferensi pelanggan dan perilaku pelanggan.

Konsep CRM banyak diadopsi oleh perusahaan yang bergerak dalam pelayanan jasa. Beberapa alasan disebutkan oleh para ahli marketing berkaitan dengan munculnya konsep CRM, diantaranya pelanggan semakin lebih cerdas, lebih sadar harga, lebih menuntut, kurang memaafkan, dan didekati oleh lebih banyak perusahaan dengan tawaran yang sama atau bahkan lebih baik. Jadi “tidak cukup hanya menghasilkan pelanggan yang puas, tetapi harus sampai pada merasa senang dan menjadi setia (loyal).” (Kotler, 2006). Oleh karena itu terjadi pergeseran dari paradigma pemasaran massal (*mass marketing*) pada pemasaran satu sama-satu (*one to one marketing*).

Dalam hal ini, konsumen yang berbeda mewakili nilai yang berbeda. Pihak manajemen perusahaan jasa harus mampu untuk mengenali konsumen terbaiknya dan mempercayainya dengan meningkatkan pemahaman perusahaan akan kebutuhan mereka sebagai individu sehingga dapat mempertahankan loyalitasnya terhadap perusahaan. Dengan menerapkan konsep CRM (*Customer Relationship Management*), perusahaan dapat melakukan identifikasi konsumen dengan melakukan segmentasi konsumen. Tujuan dari proses segmentasi konsumen adalah untuk mengetahui perilaku konsumen dan menerapkan strategi pemasaran yang tepat sehingga mendatangkan keuntungan bagi pihak perusahaan.

Penelitian ini mengembangkan algoritma K-Means Clustering yang berdasarkan customer relationship management (CRM) pada pelanggan City-Courir. Dari penelitian ini diharapkan didapatkan algoritma K-Means Clustering yang berbasis CRM yang dapat meningkatkan pertumbuhan jangka panjang dan *profitabilitas* perusahaan dengan mengetahui perilaku dan kebutuhan pelanggan,

II. TINJAUAN PUSTAKA

A. Perkembangan Penerapan K-Means

Algoritma K-Means merupakan metode *clustering* berbasis jarak yang mempartisi data ke sejumlah kelompok dan bekerja pada atribut numerik. Algoritma ini dimulai dengan pemilihan jumlah kelompok (K) secara acak serta pengambilan sebagian populasi

sejumlah K untuk dijadikan sebagai titik pusat awal. Salah satu metode perhitungan jarak yang bisa digunakan adalah *Euclidean Distance*. Perhitungan jarak menggunakan metode *Euclidean* dinyatakan sebagai berikut :

$$\text{Euclid}(x,y) = \sqrt{\sum_i^n (x_i - y_i)^2}, i = 1,2 \dots n$$

Dimana

x : obyek ke-1

y : obyek ke-2

n : banyaknya atribut obyek ke-1 dan ke-2

Prosedur dasar *clustering* K-Means adalah sebagai berikut (Johnson & Wicherin 2002):

1. Menentukan *k* sebagai jumlah *cluster* yang ingin dibentuk
2. Membangkitkan *k centroids* (titik pusat *cluster*) awal secara random
3. Menghitung jarak setiap data ke masing-masing *centroids*
4. Setiap data memilih *centroid* yang terdekat
5. Menentukan posisi *centroids* baru dengan cara menghitung nilai rata-rata dari data-data yang berada pada *centroids* yang sama
6. Kembali ke langkah 3 jika posisi *centroids* baru dengan *centroids* lama tidak sama

B. Clustering Hirarki (Hierarchical clustering)

Clustering hierarki membangun sebuah hirarki cluster atau dengan kata lain sebuah pohon cluster yang juga dikenal sebagai *dendogram*. Setiap node cluster mengandung cluster anak, cluster-cluster saudara yang membagi point yang ditutupi oleh induk mereka. Metode-metode *clustering* hirarki dikategorikan kedalam *agglomeratif* (bawah atas) dan *divisive* (atas bawah) (Jain & Murthy, 1999);, Kaufman & Rouseeuw 1990). *Clustering agglomeratif* dimulai dengan cluster satu point (singleton) dan secara berulang-ulang menggabungkan dua atau lebih cluster yang paling tepat. Cluster *divisive* dimulai dengan satu cluster dari semua point data dan secara berulang-ulang melindungi cluster yang paling tepat. Proses erbut berlanjut hingga kriteria penghentian (seringkali, jumlah *k* yang diperlukan dari *cluster*) dicapai. Kelebihan cluster hirarki meliputi :

- (I) Fleksibilitas yang tertanam mengenai level *granularitas*
- (II) Kemudahan menangani bentuk-bentuk kesamaan atau jarak
- (III) Pada akhirnya, daya pakai pada tipe-tipe atribut apapun.

Kelemahan dari clustering hirarki berhubungan dengan :

- (I) Ketidakjelasan kriteria terminasi

(II) Terhadap perbaikan perbaikan hasil clustering, sebagian besar algoritma hirarki tidak mengunjung kembali cluster-clusternya yang telah dikonstruksi.

Untuk *clustering* hirarki, menghubungkan atau memisahkan subset dari point-point dan bukan point-point individual, jarak antara point-point individu harus digeneralikan terhadap jarak antara subset.

Ukuran kedekatan yang diperoleh disebut metrik berhubungan. Tipe metrik hubungan yang digunakan secara signifikan memperoleh algoritma hirarki, karena merefleksikan konsep tertentu dari kedekatan dan konektivitas. Metrik hubungan antar cluster utama (Murtagh 1985, Olson 1995) termasuk hubungan tunggal, hubungan rata-rata, dan hubungan sempurna.

Algoritma *clustering* hirarki populer untuk data kategorikal COBWEB (Fisher, 1987) memiliki dua kualitas yang sangat penting, Pertama menggunakan pembelajaran incrementasi. Daripada mengikuti pendekatan divide atau agglomerative, secara dinamis membangun sebuah dendrogram melalui pengolahan satu point data pada suatu waktu. Kedua COBWEB termasuk pada pembelajaran berdasarkan konseptual atau model, Ini berarti bahwa setiap cluster dianggap sebagai sebuah model yang dapat dijelaskan secara intrinsik, dan bukan sebagai sebuah kumpulan point yang ditentukan terhadapnya.

Dendrogram COBWEB disebut pohon klasifikasi. Setiap node pohon C, sebuah cluster berhubungan dengan probabilitas kondisional untuk pasangan-pasangan nilai-nilai atribut, yakni :

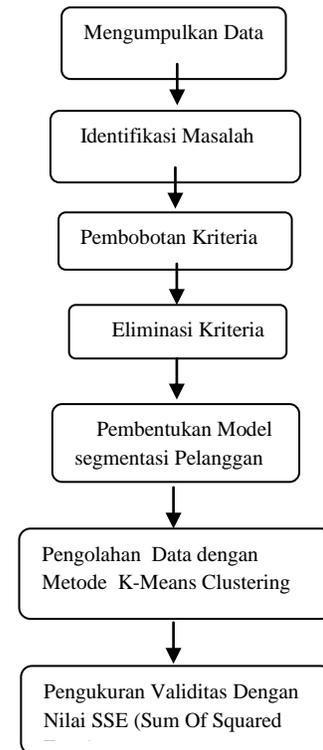
$$Pr(X_i = v_i | C), i = 1; d.p = I \setminus A1$$

Aktivitas Berdasarkan latar belakang di atas, dapat dirumuskan beberapa masalah yang menjadi inti dari penelitian ini, yaitu :

1. Bagaimana menggunakan *metode Kernel* untuk data pelanggan pada perusahaan yang bergerak dibidang jasa.
2. Bagaimana memperoleh model segmentasi pelanggan dengan *Kernel K-Means Clustering* berbasis Customer Relationship Management (CRM).

III, METODOLOGI PENELITIAN

Obyek penelitian ditentukan melalui survei industri dan dipilih perusahaan jasa courier sebagai pilot study. Penelitian dilakukan dengan mengumpulkan data jasa courier. Dari data yang dilakukan identifikasi masalah pembobotan criteria, pembentukan model serta pengukuran validitas yang dapat digambarkan sebagai berikut



Gambar 1 Framework penelitian

A. Analisa dan Pembahasan

Kriteria yang digunakan dalam pemilihan pelanggan yang terbaik dan potensial adalah Recency, Frequency, Monetary. Recency variable linquistik untuk segmentasi jenis layanan yang dimiliki jasa courier. Frequency merupakan jumlah transaksi selama per bulan yang dalam hal ini dibedakan 3 jenis yaitu sedikit dengan jumlah transaksi ≤ 1.000 dalam per bulan, agak banyak untuk jumlah transaksi ≥ 1.000 transaksi dalam per bulan, banyak untuk jumlah transaksi > 7.000 transaksi per bulan.

Monetary merupakan variable linquistik yang digunakan untuk mensegmentasikan besar uang untuk masing-masing pelanggan dalam perbulan, dibagi 3 bagian yaitu rendah untuk besar uang transaksi $> Rp. 5.000.000,-$ dan sedang untuk besar uang transaksi $> Rp. 300.000.000$ per bulan, dan tinggi untuk besar uang transaksi $> 500.000.000$ per bulan.

Untuk menentukan kriteria yang digunakan pada pemilihan segmentasi pelanggan digunakan metode K-Means Clustering.

K-Means adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode k-

means berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain. Dengan kata lain, metode ini berusaha untuk meminimalkan variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di cluster lainnya.

Objective function yang berusaha diminimalkan oleh k-means adalah:

$$J(U, V) = \sum_{k=1}^N \sum_{i=1}^c (a_{ik} * (x_k - v_i)^2)$$

dimana:

U : Matriks keanggotaan data ke masing-masing cluster yang berisikan nilai 0 dan 1

V : Matriks centroid/rata-rata masing-masing cluster

N : Jumlah data

c : Jumlah cluster

a_{ik} : Keanggotaan data ke-k ke cluster ke-i

x_k : data ke-k

v_i : Nilai centroid cluster ke-i

Prosedur yang digunakan dalam melakukan optimasi menggunakan k-means adalah sebagai berikut:

Step 1. Tentukan jumlah cluster

Step 2. Alokasikan data ke dalam cluster secara random

Step 3. Hitung centroid/rata-rata dari data yang ada di masing-masing cluster.

Step 4. Alokasikan masing-masing data ke centroid/rata-rata terdekat

Step 5. Kembali ke Step 3, apabila masih ada data yang berpindah cluster atau apabila perubahan nilai centroid, ada yang di atas nilai threshold yang ditentukan atau apabila perubahan nilai pada objective function yang digunakan, di atas nilai threshold yang ditentukan Centroid / rata-rata dari data yang ada di masing-masing cluster yang dihitung pada Step 3. didapatkan menggunakan rumus sebagai berikut:

$$v_{ij} = \sum_{k=0}^{N_i} (x_{kj}) / N_i$$

dimana:

i,k : indeks dari cluster

j : indeks dari variabel

v_{ij} : centroid/rata-rata cluster ke-i untuk variabel ke-j

x_{kj} : nilai data ke-k yang ada di dalam cluster tersebut untuk variabel ke-j

N_i : Jumlah data yang menjadi anggota cluster ke-i

Sedangkan pengalokasian data ke masing-masing cluster yang dilakukan pada Step 4. dilakukan secara penuh, dimana nilai yang memungkinkan untuk a_{ik} adalah 0 atau 1. Nilai 1 untuk data yang dialokasikan ke cluster

dan nilai 0 untuk data yang dialokasikan ke cluster yang lain. Dalam menentukan apakah suatu data teralokasikan ke suatu cluster atau tidak, dapat dilakukan dengan menghitung jarak data tersebut ke masing-masing centroid/rata-rata masing-masing cluster. Dalam hal ini, a_{ik} akan bernilai 1 untuk cluster yang centroidnya terdekat dengan data tersebut, dan bernilai 0 untuk yang lainnya.

TABEL 1 DATA CLUSTERING THN 2011-2013

CLUSTER	TAHUN		
	2013	2012	2011
TERBAIK	453,019,022	1,997,017,450	1,561,402,620
POTENSIAL	214,224,750	370,979,834	642,276,629
TIAP HARI	91,551,460	180,325,000	109,310,117
KADANG	43,153,028	47,122,000	48,013,725
KURANG	5,217,553	6,500,000	7,965,325

TABEL 2 PERHITUNGAN JARAK ITERASI PERTAMA 3 CENTROID CLUSTER TAHUN 2011

LABEL	3 CENTROID		
	C0	C1	C2
TERBAIK	2,347,944,852	2,096,265,393	2,168,950,891
POTENSIAL	526,445,823	1,065,771,833	945,298,778
TIAP HARI	292,771,126	1,412,211,430	1,219,486,532
KADANG	342,899,815	1,488,487,151	1,290,153,142
KURANG	400,575,878	1,545,524,848	1,345,763,415

TABEL 3 PERHITUNGAN JARAK ITERASI PERTAMA 2 CENTROID CLUSTER TAHUN 2013

LABEL	2 CENTROID	
	C0	C1
TERBAIK	2,396,512,917	2,129,435,005
POTENSIAL	563,371,924	704,759,826
TIAP HARI	222,225,279	983,657,667
KADANG	249,598,920	1,055,159,791
KURANG	306,109,060	1,111,843,900

B. Pengukuran Validitas Dengan Nilai SSE (Sum Of Squared Error)

Dari hasil penelitian yang dilakukan dengan menggunakan data transaksi dari tahun 2011 sampai dengan bulan maret 2013, ada beberapa hal yang dapat dianalisis. Untuk menentukan jumlah cluster yang memiliki nilai fitness terbaik digunakan pengukuran

validitas dengan nilai SSE (sum of squared-error, SSE) . Bentuk grafis nilai SSE untuk algoritma K-Means Clustering ditunjukkan pada Tabel 15.

Fungsi kriteria: jumlah dari kesalahan kuadrat (sum of squared-error, SSE) yang dapat mengukur kualitas Clustering yang dibuat :

$p \in C_i$ = tiap data poin pada cluster i ,

$$SSE = \sum_{i=1}^k \sum_{p \in C_i} d(p, m_i)^2$$

m_i = centroid dari cluster i ,

d = jarak/ distances/ variance terdekat pada masing-masing cluster i .

Nilai SSE tergantung pada jumlah cluster dan bagaimana data dikelompokkan ke dalam cluster-cluster. Semakin kecil nilai SSE semakin bagus hasil clustering yang dibuat.

Pada gambar 4.22 diatas nilai SSE pada 2 Centroid yang paling kecil dibandingkan dengan 3 Centroid, dengan demikian dapat disimpulkan bahwa cluster dengan 2 Centroid yang paling baik dan dijadikan sebagai cluster yang terbaik dalam penelitian ini.

B. Model Aturan (Model Rule)

Dari hasil cluster yang terbentuk dapat dibuat sebuah model aturan. Tabel 4.10 merupakan model aturan untuk menentukan pelanggan terbaik Kantor Pos Medan = Terbaik - A. Jika Jumlah Transaksi Per Tahun = Banyak (lebih dari 7.000,-) dan Besar Uang Transaksi Per Tahun = Tinggi (lebih dari Rp. 400.000.000 – Rp. 500.000.000,-) maka diperoleh Predikat = Terbaik - B.

Jika Jumlah Transaksi Per Tahun = Banyak (lebih dari 7.000,-) dan Besar Uang Transaksi Per Tahun = Sedang (lebih dari Rp. 300.000.000 – Rp.400.000.000) maka diperoleh Predikat = Potensial - A. Jika Jumlah Transaksi Per Tahun = Agak Banyak (lebih dari 1.000,-) dan Besar Uang Transaksi Per Tahun = Sedang (lebih dari Rp. 200.000.000 – Rp. 300.000.000) maka diperoleh Predikat = Potensial - B.

Jika Jumlah Transaksi Per Tahun = Banyak (lebih dari 7.000,-) dan Besar Uang Transaksi Per Tahun = Sedang (lebih dari Rp. 100.000.000 – Rp. 200.000.000) maka diperoleh Predikat = Tiap Hari-A. Jika Jumlah Transaksi Per Tahun = Agak Banyak (lebih dari 1.000,-) dan Besar Uang Transaksi Per Tahun = Sedang (lebih dari Rp. 50.000.000 – Rp. 100.000.000) maka diperoleh Predikat = Tiap Hari - B.

Jika Jumlah Transaksi Per Tahun = Sedikit (kurang dari 1.000,-) dan Besar Uang Transaksi Per Tahun = Rendah (lebih dari Rp. 40.000.000 - Rp. 50.000.000) maka diperoleh Predikat = Kadang-kadang.

Jika Jumlah Transaksi Per Tahun = Sedikit (kurang dari 1.000,-) dan Besar Uang Transaksi Per Tahun = Rendah

(lebih dari Rp. 40.000.000 - Rp. 50.000.000) maka diperoleh Predikat = Kadang-kadang
Jika Jumlah Transaksi Per Tahun = Sedikit (kurang dari 1.000,-) dan Besar Uang Transaksi Per Tahun = Rendah (kurang atau sama dengan dari Rp. 40.000.000) maka diperoleh Predikat = Kurang Aktif - A. Jika Jumlah Transaksi Per Tahun = Sedikit (kurang dari 1.000,-) dan Besar Uang Transaksi Per Tahun = Rendah (lebih dari Rp. 5.000.000,-) maka diperoleh Predikat = Kurang Aktif - B.

TABEL 4. MODEL ATURAN (MODEL RULE)

KETERANGAN RULE		GRADE
FREQUENCY / JLH TRANSAKSI PER THN	MONETARY / BESAR UANG TRANSAKSI PER THN	
Banyak = Transaksi > 7.000	Tinggi = besar uang transaksi > Rp. 500.000.000	Terbaik A
Banyak = Transaksi > 7.000	Tinggi = besar uang transaksi > Rp. 400.000.000	Terbaik B
Banyak = Transaksi > 7.000	Sedang = besar uang transaksi > Rp. 300.000.000	Potensial A
Agak Banyak = transaksi > 1.000	Sedang = besar uang transaksi > Rp. 200.000.000	Potensial B

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Dari penelitian yang telah dilakukan, maka dapat dihasilkan kesimpulan sebagai berikut :

1. Diperoleh suatu model aturan yang dapat digunakan untuk menilai pelanggan terbaik dan potensial sehingga dapat memberi manfaat bagi pihak manajemen dalam pengambilan keputusan.
2. Dalam cluster ini dilakukan perbandingan dengan menggunakan 2 dan 3 nilai centroid, hasilnya 2 nilai centroid lebih bagus jika dibandingkan dengan 3 nilai centroid, yaitu hasil nilai SSE (Sum Of Squared-

- Error) untuk 2 nilai centroid sebesar 3,425,922,878 lebih kecil dari pada 3 nilai centroid dengan nilai SSE sebesar 5,035,230,050.
3. Pada cluster ini diperoleh pelanggan yang memiliki katagori Terbaik-A apabila jumlah transaksi selama setahun banyak ($> 7.000,-$) dan besar uang transaksi selama setahun tinggi ($> \text{Rp. } 500.000.000,-$).
 4. Proses clustering juga menghasilkan jenis layanan yang terbanyak digunakan dikalangan para pelanggan, seperti layanan city kurir dengan prosentase tertinggi sebesar 41%, Pos Express sebesar 23%, Surat biasa sebesar 21%, Surat kilat khusus 10%, Paketpos 4%.

B. Saran

Untuk penelitian lebih lanjut dari tesis ini dapat dikembangkan dengan metode lain seperti metode SSVM (Smooth Support Vector Machine) sehingga dapat dibandingkan kinerja dari masing-masing metode untuk menghasilkan segmentasi pelanggan yang terbaik.

Pengukuran menggunakan Nilai SSE (Sum Of Squared-Error) sangat tergantung pada jumlah cluster dan bagaimana data dikelompokkan ke dalam cluster-cluster. Semakin kecil nilai SSE semakin bagus hasil clustering yang diperoleh..

- [8] Gaffar. 2007. *Customer Relationship Management and Marketing Public Relations*. Bandung : Alfabeta
- [9] Haskett, 2000. *An Introduction to Data Mining, Part 1: Understanding The Critical Data Relationship In The Corporate Data Warehouse*, Enterprise System Journal, V.15:32-34
- [10] Hammouda, K., Karaay, F. 2000. *A Comparative Study of Data Clustering Techniques*. University of Waterloo, Ontario, Canada.
- [11] Jain, A.K., Murthy, M.N. 1999. *Data Clustering : A Review*. ACM Computing Surveys, Vol. 31, No.3.
- [12] Jansen, S.M.H. 2007. "Customer Segmentation and Customer Profiling for a Mobile Telecommunications Company Based on Usage Behavior : A Vodafone Case Study" (thesis). Maastricht : University of Maastricht.
- [13] Johnson, R.A. and D.W. Wicherin. 2002. "Applied Multivariate Statistical Analysis, New Jersey : Prentice Hall.

REFERENSI

- [1] Berkin. 2002. "A Survey Of Clustering Data Mining Techniques". Technical Report, Accrue Software
- [2] Brown, S.A. 2000. *Customer Relationship Management : A Strategic Imperative in the World of e-Business*, Jhon Wiley and Sons Ltd, Canada.
- [3] Cao, Zhu, dan Hou. 2009. "Fuzzy Fisher Criterion Based Semi-Fuzzy clustering algorithm", *Journal Of Electronics and Information Technology*, Vol 30, No. 9, 2008, PP.2162-2165 (In Chinese).
- [4] Cen. 2007. "Customer Segmentation Based On Survival Character", *Journal Of Intelligent Manufacturing*, Vol 18, No. 8, 2007, PP.513-517.
- [5] Collazo-Cuevas, dkk. 2010. *Comparison Between Fuzzy C-Means Clustering And Fuzzy Clustering Subtractive In Urban Air Pollution*. International Journal of Electronics, Communications And Computer (CONIELE COMP).
- [6] Chapman, dkk. 2000. *CRISP-DM 1.0 : Step-by-Step Data Mining Guide*. Tersedia di http://www.community.udayton.edu/provost/it/training/documents/SPSS_CRI SPWPlr.pdf. [diunduh : 08 April 2013].
- [7] Fisher Douglas H. 1987, *Knowledge Acquisition Via Incrementasl Conceptual Clustering*, jornal Machine Learning Volume 2 Issue 2, September 1987