

SEGMENTASI PELANGGAN *E-MONEY* DENGAN MENGGUNAKAN ALGORITMA DBSCAN (*DENSITY BASED SPATIAL CLUSTERING APPLICATIONS WITH NOISE*) DI PROVINSI DKI JAKARTA

Windy Rohalidyawati¹, Rita Rahmawati², Mustafid³

^{1,2,3} Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro
e-mail : windyrohali@gmail.com

ABSTRACT

Customer segmentation is one effective way of marketing to determine the most potential target market. Increasing of E-money usage in DKI Jakarta and more banks are providing E-money products. One way to be able to compete in the global market, banks can segment customers. Determining potential customers of E-money users in DKI Jakarta can form segments by applying the DBSCAN (Density Based Spatial Clustering Application with Noise) algorithm. The quality of segments was measured by using the *Silhouette Coefficient*. In this study, E-money customers were grouped by reason of using the bank used, transaction activities, number of transactions, nominal balance, and frequency of top-up. The results of this study were using the density radius of 2 and minimum 3 objects that enter the density radius forming 2 segments and 17 noises. The segment quality value of 0.26. The most potential segment was the segment that has an average greater than the average of all data.

Keywords: Costumer segmentation, *E-money*, clustering, DBSCAN

1. PENDAHULUAN

Menurut Bank Indonesia (BI), secara sederhana uang elektronik didefinisikan sebagai alat pembayaran dalam bentuk elektronik di mana nilai uangnya disimpan dalam media elektronik tertentu. Produk-produk uang elektronik dibagi menjadi dua jenis yaitu *E-money* dan *E-wallet*. *E-money* adalah uang elektronik berbasis chip yang ditanam pada kartu atau media lainnya (*chip based*). Sedangkan *E-wallet* adalah uang elektronik yang berbasis *server* dalam bentuk aplikasi.

Menurut Bank Indonesia (BI), terdapat 36 penyelenggara uang elektronik dan 11 diantaranya adalah penyelenggara *E-money*. Melihat banyaknya penyelenggara *E-money*, pihak penyelenggara harus menentukan target pemasaran yang tepat. Salah satu langkah yang dapat dilakukan dalam menentukan target pemasaran adalah membuat segmen-segmen pasar. Untuk memudahkan segmentasi pasar dapat dilakukan dengan analisis *cluster* atau sering disebut dengan *clustering*.

Salah satu metode *clustering* yang dapat digunakan adalah DBSCAN (*Density Based Spatial Clustering Applications with Noise*). Algoritma DBSCAN termasuk algoritma nonparametrik dalam *unsupervised learning* sehingga tidak memerlukan asumsi dalam pengerjaannya. Secara konsep DBSCAN dapat membentuk *cluster-cluster* yang berbentuk bebas dan acak (tidak bulat) dan dapat mempermudah membentuk *cluster* jika terdapat *noise* atau pencilan pada *cluster-cluster* tersebut. Algoritma ini menumbuhkan daerah yang memiliki kerapatan tinggi menjadi *cluster-cluster* menggunakan dua parameter yang harus ditentukan dengan tepat, parameter tersebut adalah radius ketegangan, yang disimbolkan dengan ϵ , dan minimum objek (minObj) yang menjadi batas kepadatan (*density threshold*) untuk menentukan suatu *region* termasuk padat atau tidak, yang disimbolkan MinObj (Suyanto, 2019).

Penelitian ini bertujuan untuk menerapkan algoritma DBSCAN (*Density Based Spatial Clustering Application with Noise*) dalam membentuk segmen-segmen dari pelanggan *E-money* di provinsi DKI Jakarta untuk menentukan pelanggan potensial serta

menentukan kualitas dari segmen-segmen yang telah terbentuk dengan menggunakan *Silhouette Coefficient*.

2. TINJAUAN PUSTAKA

2.1 Uang Elektronik

Menurut Bank Indonesia (BI), secara sederhana uang elektronik didefinisikan sebagai alat pembayaran dalam bentuk elektronik di mana nilai uangnya disimpan dalam media elektronik tertentu. Penggunaan uang elektronik ini sebagai alat pembayaran yang inovatif dan praktis diharapkan dapat membantu kelancaran pembayaran kegiatan ekonomi yang bersifat massal, cepat dan mikro, sehingga perkembangannya dapat membantu kelancaran transaksi di jalan tol, di bidang transportasi seperti kereta api maupun angkutan umum lainnya atau transaksi di *minimarket*, *food court*, atau parkir.

2.2 Segmentasi Pasar

Segmentasi pasar adalah proses di mana pasar dibagi menjadi para pelanggan yang terdiri atas orang-orang dengan kebutuhan dan karakteristik yang sama yang mengarahkan pelanggan untuk merespon tawaran produk/jasa dan program pemasaran strategis tertentu dalam cara yang sama (Boyd JR *et al*, 2000). Keuntungan dari segmentasi pasar adalah menemukan pasar yang dapat meningkatkan keuntungan perusahaan dan mengidentifikasi pelanggan yang paling menguntungkan (Goyat, 2011). Pelanggan potensial adalah pelanggan yang mendorong keuntungan paling besar untuk perusahaan (Yankelovich, 2006).

2.3 Sampling

Teknik pengambilan sampel (sampling) adalah cara peneliti mengambil sampel atau contoh yang representatif dari populasi yang tersedia. Cara pengambilan sampel dari populasi dapat dilakukan dengan memperhatikan unsur peluang atau tidak. Jika dalam proses mengambil sampel memperhatikan unsur peluang, tipe sampling disebut sampling peluang (probability sampling) atau cara pengambilan sampel secara acak. Jika dalam proses pengambilan sampel tidak memperhatikan unsur peluang, tipe sampling disebut sampling nonpeluang (non-probability sampling) (Sanusi, 2011).

2.4 Quota Sampling

Quota sampling digunakan untuk memastikan bahwa berbagai *subgroup* dalam populasi telah terwakili dengan berbagai karakteristik sampel sampai batas tertentu seperti yang dikehendaki oleh peneliti. Dalam *quota sampling*, peneliti menentukan target kuota yang dikehendaki. Keunggulan sampel kuota adalah memungkinkan pengumpulan data dengan cepat dan biaya yang rendah. Kelemahan dari sampel kuota adalah adanya kemungkinan bias karena responden dipilih berdasarkan prosedur *convenience sampling* dan bukan berdasarkan metode probabilitas (misalnya sampel stratifikasi) (Kuncoro, 2009).

2.5 Data Mining

Data mining sering juga disebut *knowledge discovery in database* (KDD) yang artinya adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam *dataset* berukuran besar. Keluaran dari *data mining* ini dapat digunakan untuk memperbaiki pengambilan keputusan di masa depan (Santosa,

2007). Salah satu langkah dalam *data mining* adalah *preprocessing data*. Pada *preprocessing data*, salah satu yang dapat dilakukan adalah melakukan normalisasi data. Nilai-nilai atribut data yang berbeda-beda rentangnya perlu dinormalisasi atau distandarisasi agar proses *data mining* tidak bias. Salah satu metodenya yaitu normalisasi min-maks.

Misalkan A adalah atribut bertipe numerik, min_A adalah nilai minimum dalam atribut A , $maks_A$ adalah nilai maksimum dalam atribut A . Suatu nilai x_i dapat dinormalisasi menjadi nilai baru x'_i yang berada dalam rentang $[minbaru_A, maksbaru_A]$ dengan formula (Suyanto, 2019):

$$x'_i = \frac{x_i - min_A}{maks_A - min_A} (maksbaru_A - minbaru_A) + minbaru_A$$

2.6 Analisis Cluster

Analisis *Cluster* adalah istilah yang diberikan pada sebuah teknik yang digunakan untuk mengidentifikasi objek atau individu yang serupa dengan memperhatikan beberapa kriteria. Tujuan utama analisis *cluster* adalah untuk menggolongkan individu atau objek yang berhubungan secara *mutually exclusive* ke dalam jumlah yang lebih kecil. Sebuah *cluster* harus memiliki homogenitas internal (dalam satu *cluster*) yang tinggi dan heterogenitas eksternal (antar *cluster*) yang tinggi (Kuncoro, 2009).

2.7 Jarak Euclidean

Untuk data yang memiliki atribut numerik terdapat formula pengukuran jarak salah satunya adalah Jarak Euclidean. Jarak Euclidean adalah jarak antara dua objek data (i, j) dari n atribut bernilai numerik, yang dinyatakan sebagai $\mathbf{i} = x_{i1}, x_{i2}, \dots, x_{in}$ dan $\mathbf{j} = x_{j1}, x_{j2}, \dots, x_{jn}$ pada ruang dimensi n (R^n) (Suyanto, 2019):

$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{in} - x_{jn})^2}$$

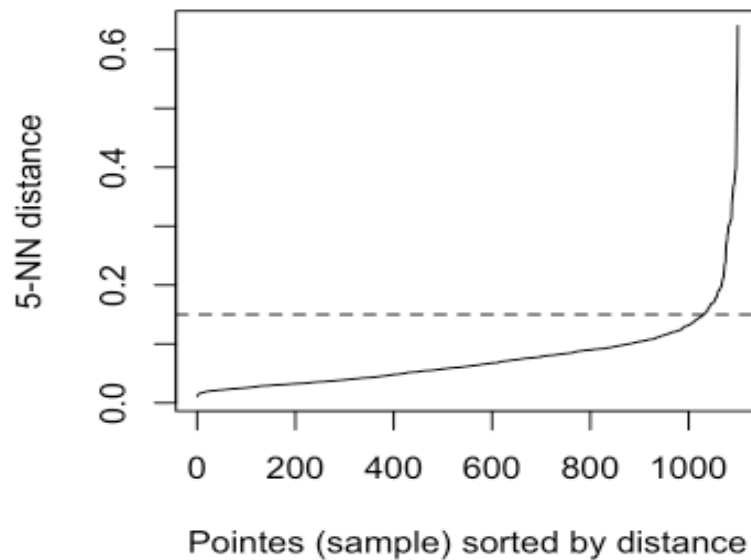
2.8 DBSCAN

Prinsip-prinsip dasar dari DBSCAN adalah sebagai berikut (Silitonga, 2016):

1. *Neighborhood* yang terletak di dalam radius ϵ disebut dengan ϵ -*neighborhood* dari objek data.
2. Jika ϵ -*neighborhood* dari suatu objek adalah jarak terkecil dalam sebuah *cluster*, objek tersebut disebut *core object*.
3. Suatu objek p adalah *density reachable* dari objek q dengan respek ke ϵ dan $minObj$ dalam suatu set objek D jika terdapat suatu rantai objek p_1, p_2, \dots, p_n , dengan $p_1 = q$ dan $p_n = p$, dengan p_{i+1} *density reachable* secara langsung dari p_i dengan respek ke ϵ dan $minObj$, untuk $1 \leq i \leq n, p_i$ anggota D .
4. Suatu objek p adalah *density connected* ke objek q dengan respek ke ϵ dan $minObj$ dalam suatu set objek D jika terdapat suatu objek o anggota D dengan ke dua p dan q adalah *density reachable* dari o dengan respek ke ϵ dan $minObj$.

2.9 Penentuan Eps (ϵ) dan MinObj

Untuk menentukan ϵ dan $MinObj$ yang optimal dapat dilakukan dengan membuat *k-dist graph* dengan melihat pergeseran nilai ϵ dari nilai k yang bervariasi. Titik dimana akan terjadi perubahan tajam pada *k-dist* akan digunakan sebagai ϵ dan nilai k digunakan sebagai $MinObj$ (Purwanto, 2012).



2.10 Algoritma DBSCAN

Pengelompokan dengan menggunakan DBSCAN secara umum dilakukan dengan algoritma sebagai berikut (Suyanto, 2019):

1. Tandai semua D sebagai “*unvisited*”
2. Pilih sebuah objek p kemudian tandai sebagai “*visited*”
3. Uji apakah dalam radius ϵ objek p memiliki minimal MinObj objek tetangga
4. Jika iya sebuah cluster C akan terbentuk, tambahkan objek p ke C
5. Jadikan N sebagai objek-objek dalam ϵ -neighborhood dari p
6. Lakukan secara iteratif kepada objek yang masih bertanda *unvisited*.
7. Jika sebuah objek dalam radius ϵ tidak memiliki minimal MinObj objek tetangga, maka ditandai sebagai noise.

2.11 Pengukuran Kualitas Cluster

Kualitas *cluster* diukur menggunakan *Silhouette Coefficient*. Nilai $a(i)$ menyatakan kepadatan *cluster* yang mengandung objek i . Semakin kecil nilai $a(i)$, semakin padat *cluster* tersebut. Sementara itu, nilai $b(i)$ menggambarkan seberapa jauh objek i terpisah dari *cluster-cluster* lain. Jadi, jika $a(i)$ bernilai sangat kecil dan $b(i)$ bernilai sangat besar, maka *Silhouette Coefficient* dari i akan mendekati satu. Sebaliknya, jika $a(i)$ bernilai sangat besar dan $b(i)$ bernilai sangat kecil, maka *Silhouette Coefficient* dari i akan mendekati -1. Jika *Silhouette Coefficient* dari i mendekati 1, maka hal ini berarti *cluster* yang berisi objek i sangat padat dan objek i terpisah jauh dari *cluster-cluster* lain. Sebaliknya, jika *Silhouette Coefficient* dari i mendekati -1, maka hal ini berarti *cluster* yang berisi objek i tidak padat dan objek i sangat dekat (bahkan tumpang tindih) dengan *cluster-cluster* lain (Suyanto, 2019).

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

3. METODE PENELITIAN

3.1. Data Penelitian

Pengambilan sampel dilakukan kepada pelanggan *E-money* yang berdomisili di 5 kota di Provinsi DKI Jakarta, yaitu Jakarta Timur, Jakarta Barat, Jakarta Selatan, Jakarta Utara, dan Jakarta Pusat. Responden pada penelitian ini adalah penduduk DKI Jakarta dari 5 kelompok umur 15-19 tahun, 20-24 tahun, 25-29 tahun, 30-34 tahun, dan 35-39 tahun. Jumlah sampel yang diambil sebanyak 180 orang.

3.2. Tahapan Analisis Data

Analisis *cluster* ini dilakukan dengan bantuan *software* Ri386.3.5.2. Tahap-tahap yang dilakukan dalam segmentasi pelanggan *E-money* adalah sebagai berikut:

1. Menentukan sampel dengan *Quota Sampling*:
 - a. Menentukan total jumlah sampel yang akan diambil.
 - b. Mencari jumlah populasi penduduk Provinsi DKI Jakarta berdasarkan umur dan kota.
 - c. Menghitung kuota sampel berdasarkan umur dari populasi penduduk di kota Jakarta Utara, Jakarta Barat, Jakarta Pusat, Jakarta Timur, dan Jakarta Selatan
2. Membuat kuesioner.
3. Melakukan pengumpulan data melalui kuesioner *online*.
4. Melakukan *preprocessing data* dengan melakukan normalisasi min-max.
5. Melakukan analisis segmentasi pelanggan *E-money* dari data yang telah dinormalisasi. Analisis segmentasi pelanggan *E-money* menggunakan algoritma *Density-Based Spatial Clustering Algorithm with Noise* (DBSCAN):
 - a. Menentukan nilai parameter MinObj dan ϵ berdasarkan *k-dist*
 - b. Memilih titik *p* secara acak, berdasarkan MinObj dan ϵ yang telah dipilih dengan menghitung jarak titik *p* dengan titik lainnya menggunakan Jarak Euclidean.
 - c. Ambil semua titik yang menjadi tetangga titik *p* yang masuk ke dalam radius ϵ dan memiliki MinObj objek tetangga.
 - d. Setelah diketahui semua titik yang menjadi tetangga titik *p*, pilih jarak yang terkecil sebagai *core object*. Apabila titik *p* merupakan *core object* maka *cluster* mulai terbentuk.
 - e. Apabila *p* adalah *border* dan semua titik yang menjadi tetangga titik *p* sudah dikunjungi, maka proses dilanjutkan ke titik yang lain.
 - f. Melakukan kembali langkah a sampai dengan e hingga semua titik telah diproses. Jika terdapat titik yang bukan termasuk dalam *cluster* maka disebut dengan *noise*.
 - g. Melakukan pengukuran kualitas cluster menggunakan *Silhouette Coefficient* terhadap *cluster* yang telah dibentuk dari algoritma DBSCAN.
 - h. Melakukan interpretasi hasil dari terbentuknya *cluster-cluster* yang terbentuk dari algoritma DBSCAN.

4. HASIL DAN PEMBAHASAN

4.1 Penarikan Sampel

Sampel yang diambil dalam penelitian ini adalah pengguna *E-money* yang merupakan penduduk DKI Jakarta yang berusia 15-19 tahun, 20-24 tahun, 25-29 tahun, 30-34 tahun, 35-39 tahun dan berdomisili di Jakarta Pusat, Jakarta Utara, Jakarta Barat, Jakarta Selatan dan Jakarta Timur. Sampel diambil sebanyak 180 orang.

4.2 Analisis Deskriptif

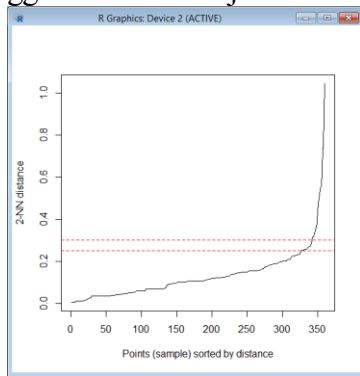
Pertama dilakukan analisis deskriptif untuk mendapatkan gambaran umum mengenai data penelitian yang sudah terkumpul.

Tabel 1. Analisis Deskriptif Pelanggan *E-money*

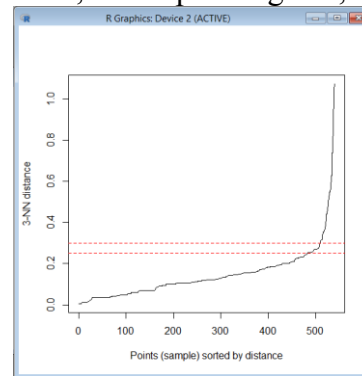
Variabel	N	Minimum	Maksimum	Mean	StDev
Saldo	180	Rp. 2.000	Rp. 850.000	Rp. 113.955	Rp. 125.698
<i>Top-up</i>	180	1	30	3	4
Transaksi	180	0	10	2	1
Alasan	180	1	8	3	2
Kegiatan	180	1	6	3	1

4.3 Penentuan Eps (ϵ) dan MinObj yang Optimal

Pada DBSCAN pemilihan parameter Eps (ϵ) dan MinObj yang optimal dipilih berdasarkan k-dist graph. Komputasi dilakukan untuk mendapatkan nilai k-dist pada nilai k yang bervariasi. K-dist dicari pada titik k=2 dan k=3. Dilihat pada Gambar 1 dan Gambar 2, jika menggunakan MinObj=2 maka ϵ yang optimal adalah 0,25 sampai dengan 0,3 dan jika menggunakan MinObj=3 maka ϵ yang optimal adalah 0,25 sampai dengan 0,3.



Gambar 1. K-dist dengan k=2



Gambar 2. K-dist dengan k=3

4.4 Jarak Euclidean

Segmen pelanggan *E-money* dibentuk menggunakan algoritma DBSCAN memerlukan jarak antar objek. Jarak antar objek dihitung menggunakan jarak Euclidean, hasil perhitungan jarak antar objek dapat dilihat pada Tabel 2.

Tabel 2. Perhitungan Jarak Euclidean

(i,j)	d(i,j)
(1,1)	0
(1,2)	0,335377
(1,3)	0,527741
	...
(2,2)	0
(2,3)	0,835263
(2,4)	0,403512
	...
(179,179)	0
(179,180)	0,825511
(180,180)	0

4.5 Pembentukan Segmen

Hasil pembentukan segmen pada data pelanggan *E-money* menggunakan algoritma DBSCAN di DKI Jakarta dijelaskan pada Tabel 3.

Tabel 3. Hasil Segmentasi Pelanggan E-money dengan Algoritma DBSCAN

MinObj	2		3	
Eps	0,25	0,3	0,25	0,3
Banyaknya segmen yang terbentuk	4	1	3	1
Noise	13	8	17	8

4.6 Interpretasi Segmen

Berdasarkan Tabel 4, parameter terbaik untuk membentuk segmen pelanggan *E-money* di DKI Jakarta adalah $\epsilon=0,3$ pada $\text{minObj}=2$ dan $\text{minObj}=3$ karena memiliki lebih sedikit *noise* dan memiliki *Silhouette Coefficient* yang paling besar dibandingkan parameter lainnya. Namun, parameter tersebut hanya menghasilkan 1 segmen sehingga tidak ada segmen pembanding untuk menentukan segmen yang paling potensial. Oleh karena itu, parameter yang digunakan untuk membentuk segmentasi pelanggan *E-money* di Provinsi DKI Jakarta adalah $\epsilon=0,25$ dan $\text{minObj}=3$. Parameter tersebut dipilih karena membentuk 2 segmen sehingga dapat dilihat segmen yang paling potensial.

Tabel 4. Interpretasi Segmen yang Terbentuk

MinObj	2		3	
Eps (ϵ)	0,25	0,3	0,25	0,3
Banyaknya segmen yang terbentuk	4	1	2	1
Noise	13	8	17	8
Rata-rata <i>Silhouette Coefficient</i>	0,15	0,47	0,26	0,47

Setelah diketahui jumlah segmen yang terbentuk, maka dilakukan interpretasi masing-masing segmen yang dijelaskan pada Tabel 5

Tabel 5. Perbandingan Rata-rata Segmen 1 dan Segmen 2

Mean	Alasan	Kegiatan	Saldo	Top-up	Transaksi
Seluruh data	3	3	Rp. 113.955	3	2
Segmen 1	2	3	Rp. 87.433	2	2
Segmen 2	1	2	Rp. 500.000	2	4

Dapat dilihat pada Tabel 5 bahwa pengguna *E-money* dalam segmen 2 adalah pelanggan potensial karena memiliki rata-rata nominal saldo dan banyaknya transaksi lebih besar dari rata-rata seluruh data.

5. KESIMPULAN

Berdasarkan k-dist *graph* parameter Eps (ϵ) dan MinObj yang optimal yang $\text{MinObj}=2$ dan 3, $\epsilon=0,25$ dan $\epsilon=0,3$. Dengan menggunakan Eps (ϵ) sebesar 0,25 dan MinObj sebesar 3 menghasilkan nilai *Silhouette Coefficient* sebesar 0,26 dan 17 *noise*. Parameter tersebut membentuk 2 segmen yaitu segmen 1 dan segmen 2. Pengguna *E-money* dalam segmen 2 adalah pelanggan potensial karena memiliki rata-rata nominal saldo dan banyaknya transaksi lebih besar dari rata-rata seluruh data.

DAFTAR PUSTAKA

- Boyd JR, H. W., Walker, J. O., & Larreche, J. C. (2000). *Manajemen Pemasaran Suatu Pendekatan Strategis dengan Orientasi Global*. Jakarta: Penerbit Erlangga.
- Goyat, S. (2011). The basis of market segmentation: a critical review of literature. *European Journal od Business and Management*, 2222-2839.
- Indonesia, B. (n.d.). *Edukasi Perlindungan Konsumen*. Retrieved December 19, 2018, from Bank Indonesia Web Site: <https://www.bi.go.id/id/edukasi-perlindungan-konsumen/edukasi/produk-dan-jasa-sp/uang-elektronik/Pages/default.aspx>
- Kuncoro, M. (2009). *Metode Riset untuk Bisnis & Ekonomi*. Jakarta: Penerbit Erlangga.
- Purwanto, Barus, U. Y., Adrianto, B., & Agung, H. (2012). *Spatial Hotspots Clustering of Forest and Land Fires using DBSCAN and ST-DBSCAN*. Bogor.
- Santosa, B. (2007). *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Sanusi, A. (2011). *Metodologi Penelitian Bisnis*. Jakarta: Penerbit Salemba Empat.
- Silitonga, P. (2016). ANALISIS POLA PENYEBARAN PENYAKIT PASIEN PENGGUNA BADAN NEGARA JAMINAN SOSIAL (BPJS) KESEHATAN DENGAN MENGGUNAKAN METODE DBSCAN CLUSTERING. *Jurnal TIMES*, 36-39.
- Suyanto, D. (2019). *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Penerbit Informatika.
- Yankelovich, D. (2006). New Criteria for Market Segmentation. *Harvard Business Review*.