



Analisis Pada Produk Susu Berdasarkan Komposisi Yang Terkandung Menggunakan Metode Hierarki Cluster

Puteri Fajar Addini^{1*}, Kevin Raih Saputra Dakhi², Putri Cania Br Tarigan³

^{1,2,3}Teknik Informatika, STMIK Pelita Nusantara, Medan, Indonesia

Email Penulis Korespondensi: ¹pfaddini@gmail.com

Abstrak— Analisis cluster dengan metode hierarki adalah analisis yang penclusteran datanya dilakukan dengan cara mengukur jarak kedekatan pada setiap objek yang kemudian membentuk sebuah dendrogram. Metode *Ward* merupakan salah satu metode hierarki cluster adalah metode pembentukan cluster yang didasari oleh hilangnya informasi akibat penggabungan objek menjadi *cluster*. Hal ini diukur dengan menggunakan jumlah total dari deviasi kuadrat pada *mean cluster* untuk setiap pengamatan. *Error sum of squares* (SSE) digunakan sebagai fungsi objektif. Dua objek akan digabungkan jika mempunyai fungsi objektif terkecil diantara kemungkinan yang ada. Penggunaan analisis *cluster* dapat ditemukan dalam berbagai bidang. Contohnya pada bidang pemasaran dan gizi. Susu memiliki komponen gizi. Susu merupakan pendukung 4 sehat 5 sempurna. Berdasarkan latar belakang tersebut, maka peneliti tertarik melakukan pengelompokan terhadap beberapa jenis susu sesuai dengan kadar zat-zat yang terkandung didalamnya. Hasil penelitian cluster yang terbentuk berdasarkan komposisi kandungan 25 sampel susu adalah 9 cluster dimana cluster pertama memiliki 5 anggota, cluster kedua memiliki 3 anggota, cluster ketiga memiliki 2 anggota, cluster keempat memiliki 3 anggota, cluster kelima 1 anggota, cluster keenam memiliki 3 anggota, cluster ketujuh memiliki 2 anggota, cluster kedelapan memiliki 2 anggota dan cluster kesembilan memiliki 4 anggota.

Kata Kunci: Analisis, Hierarki, Susu, Komposisi, Ward

Abstract— Cluster analysis using the hierarchical method is an analysis in which data is clustered by measuring the proximity distance to each object which then forms a dendrogram. Ward's method is a hierarchical cluster method, a method for forming clusters which is based on the loss of information due to combining objects into clusters. This is measured using the total sum of the squared deviations in the cluster mean for each observation. Error sum of squares (SSE) is used as an objective function. Two objects will be combined if they have the smallest objective function among the existing possibilities. The use of cluster analysis can be found in various fields. For example in the fields of marketing and nutrition. Milk has nutritional components. Milk is a supporter of 4 healthy 5 perfect. Based on this background, researchers are interested in grouping several types of milk according to the levels of substances contained in them. The research results of the clusters formed based on the content composition of 25 milk samples were 9 clusters where the first cluster had 5 members, the second cluster had 3 members, the third cluster had 2 members, the fourth cluster had 3 members, the fifth cluster had 1 member, the sixth cluster had 3 members, the seventh cluster has 2 members, the eighth cluster has 2 members and the ninth cluster has 4 members.

Keywords: Analysis, Hierarchy, Milk, Composition, Ward

I. PENDAHULUAN

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Turban dkk, 2005). Analisis kelompok (*cluster analysis*) adalah pengelompokan data (objek) yang didasarkan hanya pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan antar data (Tan, 2006).

Metode penclusteran dalam analisis cluster ada 2, yaitu metode hierarki dan metode non hierarki. Analisis cluster dengan metode hierarki adalah analisis yang penclusteran datanya dilakukan dengan cara mengukur jarak kedekatan pada setiap objek yang kemudian membentuk sebuah dendrogram.

Metode *Ward* merupakan suatu metode pembentukan cluster yang didasari oleh hilangnya informasi akibat penggabungan objek menjadi *cluster*. Hal ini diukur dengan menggunakan jumlah total dari deviasi kuadrat pada *mean cluster* untuk setiap pengamatan. *Error sum of squares* (SSE) digunakan sebagai fungsi objektif. Dua

objek akan digabungkan jika mempunyai fungsi objektif terkecil diantara kemungkinan yang ada. metode *Ward* dimana rata-rata untuk setiap *cluster* dihitung. Lalu, dihitung jarak *Euclidean* antara setiap objek dan nilai rata-rata itu, lalu jarak itu dihitung semua. Pada setiap tahap, dua *cluster* yang memiliki kenaikan *Sum of squares* dalam *cluster* yang terkecil digabungkan (Simamora, 2005).

Penggunaan analisis *cluster* dapat ditemukan dalam berbagai bidang. Contohnya pada bidang pemasaran dan gizi. Susu memiliki komponen gizi. Susu merupakan pendukung 4 sehat 5 sempurna. Berdasarkan latar belakang tersebut, maka peneliti tertarik melakukan pengelompokan terhadap beberapa jenis susu sesuai dengan kadar zat-zat yang terkandung didalamnya.

II. METODE PENELITIAN

1. Metode Ward

Metode varians menghasilkan cluster dengan meminimumkan variansi dalam *cluster*. Metode variansi yang biasa dipergunakan ialah metode ward. Untuk setiap cluster rata-rata dari seluruh variabel dihitung, kemudian



setiap objek, jarak euclid kuadrat ke rata-rata cluster dihitung. Metode ward ditemukan oleh Ward pada tahun 1963 tidak menghitung jarak antar kelompok, namun metode ini membentuk kelompok-kelompok dengan memaksimalkan kehomogenan dalam kelompok. Jumlah kuadrat dalam kelompok yang memuat objek tersebut yang diminimalkan sering disebut *Error Sum of Square* (SSE).

2. Standarisasi Data

Variabel yang memiliki nilai besar mempunyai pengaruh yang lebih besar dalam melakukan prediksi klasifikasi dari pada variabel dengan nilai kecil. Untuk mengatasi masalah tersebut, dapat digunakan teknik normalisasi variabel sehingga semua variabel akan berbeda dalam jangkauan yang sama. Tanpa dilakukan normalisasi, bisa jadi salah satu variabel akan mendominasi di dalam klasifikator. Cara menentukan nilai normalisasi adalah dengan menghitung nilai mean dan variansi dari masing-masing variabel.

3. Deteksi Multikolinieritas

Uji multikolinearitas bertujuan untuk menguji apakah ditemukan adanya korelasi antar variabel bebas (independen) pada model. Asumsi multikolinearitas mengharuskan tidak adanya korelasi yang sempurna atau besar diantara variabel-variabel independen. Analisis koefisien korelasi bertujuan untuk mempelajari apakah ada hubungan antara dua variabel. Koefisien korelasi antar variabel independen haruslah lemah (dibawah 0,8). Jika korelasi kuat, terjadilah problem multikolinearitas.

4. Menghitung Jarak Euclidean

Tujuan analisis cluster adalah mengelompokkan objek yang mirip ke dalam cluster yang sama. Oleh karena itu memerlukan beberapa ukuran untuk mengetahui seberapa mirip atau berbeda objek-objek tersebut. Pendekatan yang biasa digunakan adalah mengukur kemiripan yang dinyatakan dalam jarak (*distance*) antara pasangan objek. Pada analisis cluster terdapat tiga ukuran untuk mengukur kesamaan antar objek jarak euclidean mengukur jumlah kuadrat perbedaan nilai pada masing-masing variabel.

Tahapan-tahapan Kerangka Kerja Penelitian dari mulai analisis masalah, Pengumpulan Data, Pembangunan Sistem, Pengujian Sistem dan Implementasi Sistem untuk penjelasan lebih lengkap dapat dilihat pada diagram alir gambar berikut:

Adapun variabel-variabel yang digunakan dalam penelitian ini yaitu:

- Mendefinisikan masalah dan menentukan solusi yang diinginkan
- Membuat struktur hierarki yang diawali dengan tujuan utama
- Membuat matriks perbandingan berpasangan yang menggambarkan kontribusi relatif atau pengaruh setiap elemen terhadap tujuan atau kriteria yang setingkat diatasnya.
- Mendefinisikan perbandingan berpasangan sehingga diperoleh jumlah penilai seluruhnya.
- Menghitung nilai eigen dan menguji konsistensinya. Jika tidak konsisten maka pengambilan data diulangi.

III. HASIL DAN PEMBAHASAN

Standarisasi Data

Tabel 1. Statistik Deskriptif

Var	N	Min	Max	Mean	Std. Dev
Lemak	25	2,00	7,00	4,62	1,24399
Protein	25	3,00	8,00	5,24	1,30000
Karbonhidrat	25	10,00	35,00	20,20	5,88076
Serat_pangan	25	1,00	5,00	1,20	,81650
Glukosa	25	6,00	25,00	16,00	5,45436
Natrium	25	40,00	130,00	83,80	24,07800
Valid N (listwise)	25				

Standarisasi dilakukan agar tidak ada data yang lebih dominan atau tidak ada data yang mempunyai skala data yang lebih besar.

Deteksi Multikolinieritas

Deteksi Multikolinieritas digunakan untuk melihat apakah terdapat multikolinieritas atau tidak. Berikut merupakan nilai koefisien korelasi antar variabel.

Tabel 2. Koefisien Korelasi Antar Variabel

Var	X ₁	X ₂	X ₃	X ₄	X ₅
X ₁	1	0,085	0,164	0,138	-0,326
X ₂	0,085	1	0,308	-0,141	-0,083
X ₃	0,164	0,308	1,000	-0,035	-0,252
X ₄	0,138	-0,141	-0,035	1	-0,766
X ₅	-0,326	-0,083	-0,252	-0,766	1

Berdasarkan hasil Tabel 2 terlihat bahwa nilai mutlak dari koefisien korelasi antar variabel penelitian yang berbeda bernilai di bawah 0,8 yang artinya tidak ada multikolinieritas antar variabel dalam penelitian dan dapat dilanjutkan ke proses pengelompokkan dengan metode Ward.

Menghitung Jarak Euclidean

Tujuan analisis cluster adalah mengelompokkan objek yang mirip ke dalam cluster yang sama. Oleh karena itu memerlukan beberapa ukuran untuk mengetahui seberapa mirip atau berbeda objek-objek tersebut. Berikut nilai jarak antar objek yang telah dihitung.

Tabel 3. Perhitungan Jarak Euclidean

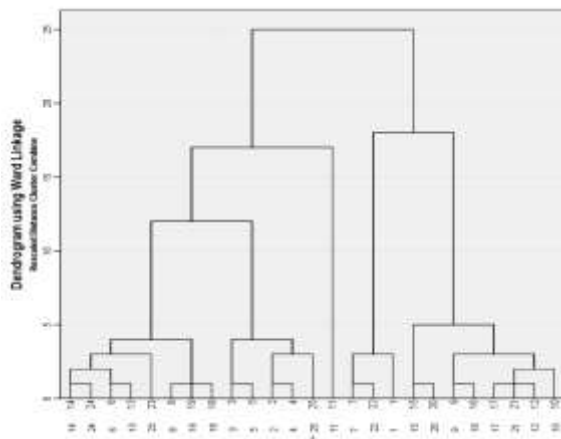


	1	2	3	4	5	...	25
1	0	19,27	12,36	24,20	9,85	...	32,58
2	19,27	0	2,58	0,75	3,09	...	5,06
3	12,36	3,58	0	6,23	0,32	...	6
4	24,20	0,75	6,23	0	5,85	...	6,67
5	9,85	3,09	0,32	5,85	0	...	7,72
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
25	32,58	5,06	6	6,67	7,72	...	0

Berdasarkan hasil perhitungan jarak euclid antara cimory blueberry dan dancow coklat memiliki jarak terdekat sebesar 2,06. Hal ini menunjukkan bahwa cimory blueberry dan dancow coklat memiliki kemiripan secara karakteristik komposisi kandungan mie.

Pembentukan Cluster dengan Metode Ward

Pengelompokkan produk susu berdasarkan komposisi kandungannya dengan metode Ward menggunakan bantuan software SPSS. Hasil pengelompokkan produk susu berdasarkan variable kandungan komposisi yang terkandung disajikan dalam bentuk dendrogram yaitu suatu alat grafis yang digunakan untuk menyajikan hasil pengukuran yang dapat dilihat pada Gambar 1 sebagai berikut :



Gambar 1. Dendrogram Pengelompokan Produk Susu Berdasarkan Komposisi Kandungan

Berdasarkan Gambar 1, terlihat bahwa cluster pertama memiliki 5 anggota, cluster kedua memiliki 3 anggota, cluster ketiga memiliki 2 anggota, cluster keempat memiliki 3 anggota, cluster kelima 1 anggota, cluster keenam memiliki 3 anggota, cluster ketujuh memiliki 2 anggota, cluster kedelapan memiliki 2 anggota dan cluster kesembilan memiliki 4 anggota.

Profiling Cluster

Profiling Cluster Pertama

Tabel 4. Cluster Pertama

Susu	Lm	Pr	Kb	Sp	Gl	Na
14	5	5	22	1	15	110
24	5	5	23	1	17	105
6	6	6	20	1	16	90
13	5	6	23	1	11	85
23	7	5	24	2	16	120
Total	28	27	112	6	75	510
Mean	5,6	5,4	22,4	1,2	15	102

Dimana

Lm = Lemak; Pr = Protein; Kb = Karbohidrat;

Sp = Serat Pangan; Gl = Gula; Na = Natrium.

Profiling pada cluster pertama yang beranggotakan 5 jenis susu. Jenis susu pada cluster pertama adalah cimory blueberry, dancow coklat, Frisianflag coconut delight, ultramilk caramel dan frisianflag primago. Pada cluster pertama rata-rata lemak sebanyak 5,6 gram, rata-rata protein sebanyak 5,4 gram, rata-rata karbohidrat 22,4 gram, rata-rata serat pangan sebanyak 1,2 gram, rata-rata glukosa sebanyak 15 gram dan rata-rata natrium sebanyak 102 gram.

Profiling Cluster Kedua

Tabel 5. Cluster Kedua

Susu	Lm	Pr	Kb	Sp	Gl	Na
8	4,5	5	22	1	22	110
19	4,5	6	22	1	22	105
18	4	6	22	1	22	130
Total	13	17	66	3	66	345
Mean	4,3	5,7	22	1	22	115

Dimana

Lm = Lemak; Pr = Protein; Kb = Karbohidrat;

Sp = Serat Pangan; Gl = Gula; Na = Natrium.

Profiling pada cluster kedua yang beranggotakan 3 jenis susu. Jenis susu pada cluster kedua adalah frisianflag strawberry, milku coklat dan milku strawberry. Pada cluster kedua rata-rata lemak sebanyak 4,3 gram, rata-rata protein sebanyak 5,7 gram, rata-rata karbohidrat sebanyak 22 gram, rata-rata serat pangan sebanyak 1 gram, rata-rata glukosa sebanyak 22 gram dan rata-rata natrium sebanyak 115 gram.

Profiling Cluster Ketiga

Tabel 6. Cluster Ketiga

Susu	Lm	Pr	Kb	Sp	Gl	Na
3	4	6	24	1	19	45
5	4,5	6	22	1	19	50
Total	8,5	12	46	2	38	95
Mean	4,25	6	23	1	19	47,5

Dimana

Lm = Lemak; Pr = Protein; Kb = Karbohidrat;

Sp = Serat Pangan; Gl = Gula; Na = Natrium.



Profiling pada cluster ketiga yang beranggotakan 2 jenis susu. Jenis susu pada cluster ketiga adalah ultramilk coklat dan ultramilk strawberry. Pada cluster ketiga rata-rata lemak sebanyak 4,25 gram, rata-rata protein sebanyak 6 gram, rata-rata karbohidrat sebanyak 23 gram, rata-rata serat pangan sebanyak 1 gram, rata-rata glukosa sebanyak 19 gram dan rata-rata natrium sebanyak 47,5 gram.

Profiling Cluster Keempat

Tabel 7. Cluster Keempat

Susu	Lm	Pr	Kb	Sp	Gl	Na
2	5	7	27	1	23	75
4	5	8	28	1	25	75
25	3	6	35	1	22	75
Total	13	21	90	3	70	225
Mean	4,3	7	30	1	23,3	75

Dimana

Lm = Lemak; Pr = Protein; Kb = Karbohidrat;
Sp = Serat Pangan; Gl = Gula; Na = Natrium.

Profiling pada cluster keempat yang beranggotakan 3 jenis susu. Jenis susu pada cluster keempat adalah greenfields strawberry, greenfields coklat dan zee vanilla twist. Pada cluster keempat rata-rata lemak sebanyak 4,3 gram, rata-rata protein sebanyak 7 gram, rata-rata karbohidrat sebanyak 30 gram, rata-rata serat pangan sebanyak 1 gram, rata-rata glukosa sebanyak 23,3 gram dan rata-rata natrium sebanyak 75 gram.

Profiling Cluster Kelima

Tabel 8. Cluster Kelima

Susu	Lm	Pr	Kb	Sp	Gl	Na
11	4.5	4	26	5	21	75
Total	4.5	4	26	5	21	75
Mean	4.5	4	26	5	21	75

Dimana

Lm = Lemak; Pr = Protein; Kb = Karbohidrat;
Sp = Serat Pangan; Gl = Gula; Na = Natrium.

Profiling pada cluster kelima yang beranggotakan 1 jenis susu. Jenis susu pada cluster kelima adalah diamond madu. Pada cluster kedua rata-rata lemak sebanyak 4,5 gram, rata-rata protein sebanyak 4 gram, rata-rata karbohidrat sebanyak 26 gram, rata-rata serat pangan sebanyak 5 gram, rata-rata glukosa sebanyak 21 gram dan rata-rata natrium sebanyak 75 gram.

Profiling Cluster Keenam

Tabel 9. Cluster Keenam

Susu	Lm	Pr	Kb	Sp	Gl	Na
7	7	7	10	1	10	90
22	7	6	12	1	7	90
1	6	6	10	1	8	40
Total	20	19	32	3	25	220
Mean	6,7	6,3	10,7	1	8,3	73,3

Dimana

Lm = Lemak; Pr = Protein; Kb = Karbohidrat;
Sp = Serat Pangan; Gl = Gula; Na = Natrium.

Profiling pada cluster keenam yang beranggotakan 3 jenis susu. Jenis susu pada cluster keenam adalah frisianflag fullcream, diamond milk fullcream dan ultramilk full cream. Pada cluster keenam rata-rata lemak sebanyak 6,7 gram, rata-rata protein sebanyak 6,3 gram, rata-rata karbohidrat sebanyak 10,7 gram, rata-rata serat pangan sebanyak 1 gram, rata-rata glukosa sebanyak 8,3 gram dan rata-rata natrium sebanyak 73,3 gram.

Profiling Cluster Ketujuh

Tabel 10. Cluster Ketujuh

Susu	Lm	Pr	Kb	Sp	Gl	Na
15	3.5	3	14	1	9	70
20	4	3	15	1	6	45
Total	7,5	6	29	2	15	115
Mean	3,75	3	14,5	1	7,5	57,5

Dimana

Lm = Lemak; Pr = Protein; Kb = Karbohidrat;
Sp = Serat Pangan; Gl = Gula; Na = Natrium.

Profiling pada cluster ketujuh yang beranggotakan 2 jenis susu. Jenis susu pada cluster ketujuh adalah cimory strawberry dan pediasure vanila. Pada cluster ketujuh rata-rata lemak sebanyak 3,75 gram, rata-rata protein sebanyak 3 gram, rata-rata karbohidrat sebanyak 1,5 gram, rata-rata serat pangan sebanyak 1 gram, rata-rata glukosa sebanyak 7,5 gram dan rata-rata natrium sebanyak 57,5 gram.

Profiling Cluster Kedelapan

Tabel 11. Cluster Kedelapan

Susu	Lm	Pr	Kb	Sp	Gl	Na
9	4	4	20	1	19	105
16	3.5	3	17	1	15	105
Total	7,5	7	37	2	34	210
Mean	3,75	3,5	18,5	1	17	105

Dimana

Lm = Lemak; Pr = Protein; Kb = Karbohidrat;
Sp = Serat Pangan; Gl = Gula; Na = Natrium.

Profiling pada cluster kedelapan yang beranggotakan 2 jenis susu. Jenis susu pada cluster delapan adalah



frisianflag susu sereal rasa kopi dan milo. Pada cluster kedelapan rata-rata lemak sebanyak 3,75 gram, rata-rata protein sebanyak 3,5 gram, rata-rata karbohidrat sebanyak 18,5 gram, rata-rata serat pangan sebanyak 1 gram, rata-rata glukosa sebanyak 17 gram dan rata-rata natrium sebanyak 105 gram.

Profiling Cluster Kesembilan

Tabel 12. Cluster Kesembilan

Susu	Lm	Pr	Kb	Sp	Gl	Na
17	4	4	18	1	15	80
21	3,5	4	17	1	16	80
12	4	5	18	1	14	60
10	2	5	14	1	11	80
Total	13,5	18	67	4	56	300
Mean	3,75	4,50	16,75	1,00	14,00	75,00

Dimana

Lm = Lemak; Pr = Protein; Kb = Karbohidrat;

Sp = Serat Pangan; Gl = Gula; Na = Natrium.

Profiling pada cluster kesembilan yang beranggotakan 4 jenis susu. Jenis susu pada cluster kesembilan adalah indomilk black latte, indomilk banana, indomilk taro dan diamond vanilla. Pada cluster kesembilan rata-rata lemak sebanyak 3,75 gram, rata-rata protein sebanyak 4,50 gram, rata-rata karbohidrat sebanyak 16,75 gram, rata-rata serat pangan sebanyak 1 gram, rata-rata glukosa sebanyak 14 gram dan rata-rata natrium sebanyak 75 gram.

IV. KESIMPULAN

Berdasarkan hasil penelitian cluster yang terbentuk berdasarkan komposisi kandungan 25 sampel susu adalah 9 cluster dimana cluster pertama memiliki 5 anggota yaitu cimory blueberry, dancow coklat, Frisianflag coconut delight, ultramilk caramel dan frisianflag primago. Cluster kedua memiliki 3 anggota yaitu frisianflag strawberry, milku cokelat dan milku strawberry. Cluster ketiga memiliki 2 anggota yaitu ultramilk cokelat dan ultramilk strawberry. Cluster keempat memiliki 3 anggota yaitu greenfields strawberry, greenfields cokelat dan zee vanilla twist. Cluster kelima 1 anggota yaitu diamond madu. Cluster keenam memiliki 3 anggota yaitu frisianflag fullcream, diamond milk fullcream dan ultramilk full cream. Cluster ketujuh memiliki 2 anggota yaitu cimory strawberry dan pediasure vanilla. Cluster kedelapan memiliki 2 anggota yaitu frisianflag susu sereal rasa kopi dan milo. Cluster kesembilan memiliki 4 anggota yaitu indomilk black latte, indomilk banana, indomilk taro dan diamond vanilla.

UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini.

V. REFERENSI

- Agresti, A. (1990). *Categorical Data Analysis*. New York: John Wiley & Sons.
- Agresti, A. (2002). *Categorical Data Analysis (second edition)*. New York: John Wiley & Sons.
- De Muth, J. E. (2014). *Basic Statistics and Pharmaceutical Statistical Applications (third edition)*. New York: Taylor & Francis Group, LLC.
- Draper, N. R. & H. Smith. (1998). *Analisis Regresi Terapan*. New York : John Wiley and Sons, Inc
- Hendayana, R. (2012). *Application Method of Logistic Regression Analyze the Agricultural Technology Adaption*.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression (Second Edition)*. New York: John. Wiley & Sond, INC. Juanda, B (2009). *EKONOMETRIKA: Pemodelan dan Pendugaan*. IPB Press
- Kleinbaum, D. G. & Klein, M. (2010). *Logistic Regression (third edition)*. New York: Springer
- Ningtias, I. P., & Rahayu, S. P. (2017). *Pemodelan Faktor-faktor yang Mempengaruhi Tingkat Pengangguran Terbuka di Provinsi Jawa Timur Tahun 2015 Menggunakan Regresi Spasial*. Jurnal SAINS DAN SENI ITS, 212-218
- Oke & Oyeka, I. C. A. (2013). *Estimating The Fisher's Scoring Matrix Formula from Logistic Model*. American Journal Of Theoretical and Applied Statistics. 2013
- Simamora, B. (2005). *Analisis Multivariate Pemasaran*. Jakarta : Gramedia Pustaka Umum.
- Tan, P. (2006). *Introduction To Data Mining*. Boston: Pearson Education.
- Turban, E. (2005). *Decision support system and intelligent system*. Yogyakarta: ANDI.