

# Data Mining dengan Algoritma Fuzzy C-Means Clustering Dalam Kasus Penjualan di PT Sepatu Bata

Cakra Ramadhana<sup>1</sup>, Yohana Dewi Lulu W<sup>2</sup>, Kartina Diah K. W.<sup>3</sup>

<sup>1,2</sup>Program Studi Sistem Informasi, Politeknik Caltex Riau, Pekanbaru 28265  
E-mail : ramadhana.cakra@yahoo.com, yohana@pcr.ac.id

<sup>3</sup>Program Studi Teknik Informatika, Politeknik Caltex Riau, Pekanbaru 28265  
E-mail : diah@pcr.ac.id

## ABSTRAK

Data transaksi di sebuah perusahaan retail selalu bertambah seiring dengan perubahan waktu. Data-data tersebut menjadi tumpukan data dalam jumlah yang sangat besar. Dengan memanfaatkan data-data tersebut, proses penggalian data sebagai informasi penting bagi perusahaan dapat dilakukan dengan analisis yang lebih dalam. Salah satu informasinya adalah produk yang laris dan tidak laris di pasaran. Untuk mengetahui hal tersebut, maka dirancang sebuah aplikasi yang berbasis Visual Basic yang mengelola informasi dari data penjualan dengan menggunakan metode Fuzzy C-Means. Data-data perusahaan berupa jumlah invoice, jumlah barang, dan jenis barang (artikel) sejumlah 1068 record diolah serta dilakukan analisa untuk periode data selama dua bulan, yakni Januari 2013 hingga Februari 2013. Metode Fuzzy C-Means pada aplikasi ini menghasilkan output berupa tiga cluster yaitu laris sebanyak 84,12% produk, kurang laris sebanyak 12,69% produk, dan sangat laris sebanyak 3,17% produk. Berdasarkan hasil pengujian, algoritma Fuzzy C-Means sesuai pada kasus penjualan di PT Sepatu Bata yang diperoleh dengan adanya 91,2% karyawan toko PT Sepatu Bata yang menyatakan bahwa aplikasi ini bekerja dengan sangat baik.

**Kata kunci :** Perusahaan Retail, Produk, Fuzzy C-Means

## 1. PENDAHULUAN

Perusahaan *retail* setiap harinya mencatat transaksi penjualan yang sangat banyak. Hal ini berdampak pada pertumbuhan jumlah data yang sangat pesat dan menimbulkan tumpukan data yang berjumlah besar dalam basis data. Jika jumlah data sedikit, masih memungkinkan digunakannya kemampuan manusia untuk melakukan analisa dan interpretasi data menggunakan metode tradisional. Dari data-data yang ada pada sebuah toko/perusahaan *retail* dapat digali informasi-informasi baru yang berguna untuk prospek bisnis ke depan. Data-data tersebut digali dengan metode yang disebut dengan data mining. *Data mining* merupakan suatu proses yang interaktif bersifat *automatically* dalam menemukan pola data-data tersebut dan memprediksi kelakuan (*trend*) di masa yang akan datang. PT Sepatu Bata, Tbk. merupakan salah satu perusahaan *retail* terbesar di Indonesia membutuhkan proses data mining untuk menggali informasi yang berguna bagi proses bisnisnya. *Data mining* pada kasus penjualan di PT Sepatu Bata, Tbk diharapkan dapat memudahkan perusahaan dalam mengidentifikasi produk-produk terbaik mereka.

Oleh karena itu, penulis membuat sebuah perhitungan penggalian data untuk membantu proses analisa data yang diperoleh dari *record-record* transaksi pada sistem informasi PT Sepatu Bata, Tbk. Pada penelitian ini digunakan metode *fuzzy clustering*, yaitu dengan algoritma *Fuzzy C-Means Clustering* (FCM). Algoritma *Fuzzy C-Means Clustering* merupakan algoritma *supervised clustering* (jumlah *cluster* ditentukan). FCM memiliki tingkat akurasi yang tinggi dan waktu komputasi yang cepat [9]. Dengan algoritma ini diperoleh informasi mengenai produk-produk laris dan menghasilkan keuntungan bagi perusahaan. Proses penggalian data ini menggunakan *tools* yang dibuat sendiri menggunakan bahasa pemrograman Visual Basic.

## 2. TINJAUAN PUSTAKA

### 2.1 Review Penelitian Terdahulu

Penelitian terdahulu telah dilakukan oleh Nugraheni dengan judul “Data Mining Dengan Metode Fuzzy Untuk *Customer Relationship Management* (CRM) pada Perusahaan *Retail*” terhadap perusahaan *retail* UD. Hasil perbandingan dapat dilihat pada Tabel 1.

Tabel 1: Perbandingan Penelitian Terdahulu

No.	Penelitian	Metode	Deskripsi
1	Penelitian Terdahulu	<i>Fuzzy C-Means</i> dan <i>Fuzzy Subtractive Clustering</i> (Nugraheni, 2011)	Melakukan proses <i>clustering</i> dari masing-masing metode dan segmentasi pelanggan pada perusahaan retail UD. Fenny sebanyak 3 juta <i>record</i> data
		K-means, Fuzzy C-Means, SOM, Decision Tree (Shin dan Sohn, 2004)	Mencari segmentasi konsumen pasar modal berdasarkan nilai potensialnya
		K-Means, Fuzzy C-Means, Mountain Clustering, Subtractive Clustering (Hammouda dan Prof. Karaay, 2000)	Melakukan perbandingan keempat algoritma clustering dengan menggunakan data medis untuk mendiagnosa penyakit hati
2	Penelitian Sekarang	<i>Fuzzy C-Means Clustering</i>	PT Sepatu Bata, Tbk sebanyak 1069 <i>record</i> data

## 2.2 Data Mining

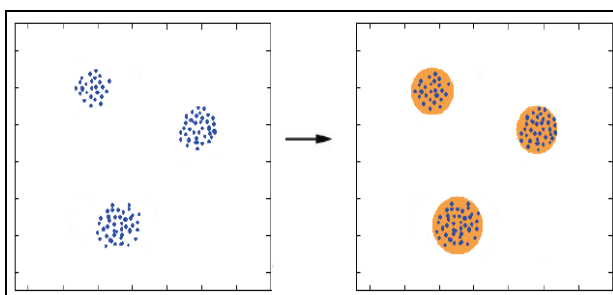
*Data mining* merupakan proses pencarian pola-pola yang menarik dan tersembunyi (*hidden pattern*) dari suatu kumpulan data yang berukuran besar yang tersimpan dalam suatu basis data, *data warehouse*, atau tempat penyimpanan data lainnya [9]. *Data mining* juga didefinisikan sebagai bagian dari proses penggalian pengetahuan dalam *database* yang sering disebut dengan istilah *Knowledge Discovery in Database* (KDD) [8]. KDD merupakan suatu area yang mengintegrasikan berbagai metode. Tahap *Data Mining*.

Karena data mining adalah suatu rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap [1]:

1. Pembersihan data (untuk membuang data yang tidak konsisten dan *noise*).
2. Integrasi data (penggabungan data dari beberapa sumber).
3. Transformasi data (data diubah menjadi bentuk yang sesuai untuk di-*mining*).
4. Aplikasi teknik *data mining*.
5. Evaluasi pola yang ditemukan (untuk menemukan yang menarik/bernilai).
6. Presentasi pengetahuan (dengan teknik visualisasi).

## 2.3 Clustering

Pengelompokan (*clustering*) merupakan teknik yang sudah cukup dikenal dan banyak digunakan untuk mengelompokkan data/objek ke dalam kelompok data (*cluster*) sehingga setiap *cluster* memiliki data yang mirip dan berbeda dengan data yang berada dalam *cluster* lain [7]. Contoh sederhana dari proses *clustering* ditunjukkan pada gambar 1. Tiga *cluster* di dalam data pada gambar 1 dapat teridentifikasi dengan mudah. Dalam kasus ini, kriteria kemiripan yang digunakan adalah jarak (dalam kasus ini jarak geometris). Proses ini disebut *distance-based clustering* [7]. Cara lain untuk melakukan *clustering* adalah *conceptual clustering* di mana objek dikelompokkan menurut konsep deskriptif berdasarkan kecocokannya.



Gambar 1: Contoh Proses *Clustering*

## 2.4 Metode Fuzzy C-Means Clustering

Metode *Fuzzy C-Means Clustering* pertama kali dikenalkan oleh Jim Bezdek pada tahun 1981 [2]. *Fuzzy C-Means* adalah salah satu teknik pengelompokan data yang mana keberadaan tiap titik data dalam suatu kelompok (*cluster*) ditentukan oleh derajat keanggotan. Metode *Fuzzy C-Means* termasuk metode *supervised clustering* dimana jumlah pusat *cluster* ditentukan di dalam proses *clustering* [7].

Algoritma *Fuzzy C-Means* disusun dengan langkah sebagai berikut [6]:

1. Input data yang akan di *cluster* X, berupa matriks berukuran  $n \times m$  ( $n$  = jumlah sampel data,  $m$  = atribut setiap data).  $X_{ij}$  = data sampel ke- $i$  ( $i=1,2,\dots,n$ ), atribut ke- $j$  ( $j=1,2,\dots,m$ ).
2. Tentukan:
  - Jumlah *cluster* yang akan dibentuk =  $c$  ( $c \geq 2$ );
  - Pangkat (pembobot) =  $w$  ( $w > 1$ ), nilai dari  $w$  yang paling optimal dan sering dipakai adalah  $w = 2$  (Klawonn dan Keller, 1997);  $w = w$ ;
  - Maksimum iterasi = MaxIter;
  - Error terkecil yang diharapkan (nilai positif yang sangat kecil) sebagai kriteria penghentian =  $\xi$ ;
  - Fungsi obyektif awal =  $P_0 = 0$ ;
  - Iterasi awal =  $t = 1$ ;
3. Bangkitkan bilangan random  $\mu_{ik} = 1, 2, \dots, n$ ;  $k = 1, 2, \dots, c$ ; sebagai elemen-elemen matriks partisi awal U. Hitung jumlah setiap kolom (atribut):

$$Q_j = \sum_{k=1}^c \mu_{ik} \quad (1)$$

dengan  $j = 1, 2, \dots, m$ .

Hitung:

$$\mu_{ik} = \frac{\mu_{ik}}{Q_j} \quad (2)$$

4. Hitung pusat *cluster* ke- $k$ :  $V_{kj}$ , dengan  $k = 1, 2, \dots, c$ ; dan  $j = 1, 2, \dots, m$ .

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w \times X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \quad (3)$$

5. Hitung fungsi obyektif pada iterasi ke- $t$ ,  $P_t$ :

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left( \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right) \quad (4)$$

6. Hitung perubahan matriks partisi:

$$\mu_{ik} = \frac{\left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{-\frac{1}{w-1}}}{\sum_{k=1}^c \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{-\frac{1}{w-1}}} \quad (5)$$

Dengan:  $i = 1, 2, \dots, n$ ; dan  $k = 1, 2, \dots, c$ .

7. Cek kondisi berhenti:
  - Jika: ( $|P_t - P_{t-1}| < \xi$ ) atau ( $t > \text{MaxIter}$ ) maka berhenti;
  - Jika tidak:  $t = t + 1$ , ulangi langkah ke-4.

## 3. PERANCANGAN

### 3.1 Analisa Perancangan

Kriteria-kriteria untuk menentukan tingkat laris atau tidaknya produk pada PT Sepatu Bata, Tbk antara lain jumlah *invoice* (nomor faktur penjualan), jumlah barang, dan jenis barang/artikel (kode dari setiap jenis barang). Data-data mentah dari toko akan disimpulkan ke dalam tabel yang akan berguna untuk proses perhitungan *Fuzzy C-Means*.

Tabel 2: Contoh Data Penjualan

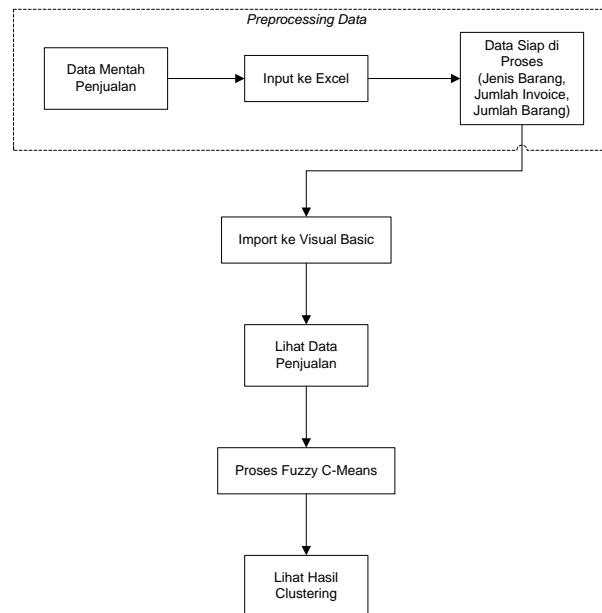
Jenis Barang	Jumlah Invoice	Jumlah Barang
871	5	10
671	4	7
872	8	18
508	10	10
952	6	23

Berdasarkan Tabel 2, dapat diketahui bahwa untuk jenis barang 871 terjual sebanyak 10 pasang yang tercatat pada 5 buah *invoice*. Lalu untuk jenis barang 671 jumlah barang yang terjual sebanyak 7 pasang yang tercatat pada 4 buah *invoice*, dan seterusnya. Selanjutnya akan ditentukan masuk ke golongan *cluster* mana dari masing-masing jenis barang tersebut dengan menggunakan *Fuzzy C-Means*. Untuk proses perhitungannya, hanya jumlah *invoice* dan jumlah barang yang dapat diolah datanya. *Output* yang diharapkan berupa tiga buah *cluster*, yaitu laris, kurang laris, dan sangat laris.

Untuk melakukan perhitungan dengan algoritma *Fuzzy C-Means*, sebelumnya ditentukan nilai awal sebagai berikut:

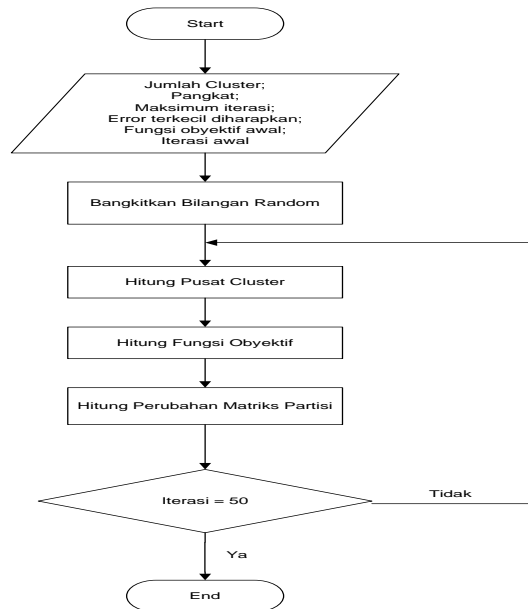
1. Jumlah *cluster* yang akan dibentuk adalah tiga, yaitu laris, kurang laris, dan sangat laris.
2. Pangkat atau pembobot yang digunakan adalah 2.
3. Maksimum iterasi adalah 50.
4. Error terkecil yang diharapkan adalah 10<sup>-5</sup>.
5. Fungsi obyektif awal yang digunakan adalah 0.
6. Nilai iterasi awal adalah 1.

### 3.2 Block Diagram



Gambar 2: Block Diagram Aplikasi Fuzzy C-Means

### 3.3 Flow Chart



Gambar 3: Flow Chart Proses Fuzzy C-Means

## 4. PENGUJIAN

### 4.1 Pengujian Sistem

Pengujian sistem aplikasi yang dibangun dilakukan dengan memeriksa satu persatu bagian bagian aplikasi yang dibangun tanpa memperhatikan struktur logika internal sistem. Metode pengujian ini dilakukan untuk memeriksa apakah sistem telah berjalan sesuai dengan yang diharapkan.

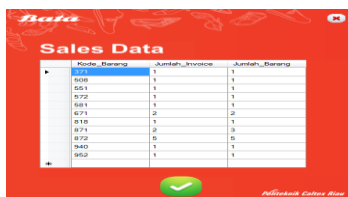
Gambar 4(a) merupakan tampilan “Splash Screen”. Gambar 4(b) adalah tampilan form “Import Data”. Gambar 4(c) merupakan tampilan form “Sales Data”. Gambar 4(d) merupakan tampilan hasil perhitungan. Gambar 4(e) adalah tampilan detail hasil perhitungan.



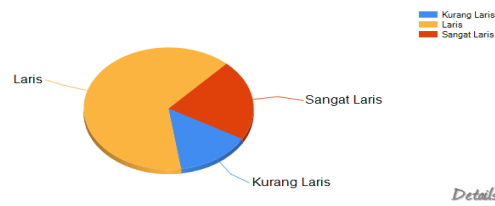
(a)



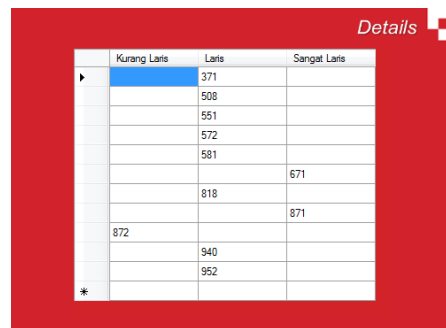
(b)



(c)



(d)



Kurang Laris	Laris	Sangat Laris
	371	
	508	
	551	
	572	
	581	
		671
	818	
		871
872		
	940	
	952	
*		

(e)

Gambar 4: Tampilan Aplikasi *Fuzzy C-Means*

## 4.2 Pengujian K-Means

Setelah dilakukan pengujian sistem, maka selanjutnya dilakukan pengujian terhadap metode *clustering* biasa, yaitu K-Means. Pengujian ini dilakukan dengan menggunakan fungsi K-Means yang disediakan oleh perangkat lunak Matlab. Metode K-Means membagi data menjadi beberapa kelompok berdasarkan data yang ada, dimana data dalam satu kelompok memiliki karakteristik yang sama dengan lainnya dan memiliki karakteristik yang berbeda dengan data yang ada pada kelompok yang lain. Secara umum, perbedaan mendasar antara K-Means dengan *Fuzzy C-Means* terletak pada data yang di *cluster*. Pada K-Means, data ter-*cluster* secara tegas, sedangkan pada *Fuzzy C-means* data ter-*cluster* pada semua *cluster* dengan derajat keanggotaan yang berbeda.

Dibandingkan dengan K-Means, algoritma *Fuzzy C-Means* lebih unggul dalam mengatasi data outlier. Pada *Fuzzy C-Means*, data outlier bergabung menjadi satu *cluster* dengan pusat *cluster*, sedangkan data outlier pada K-Means cenderung membentuk *cluster* tersendiri. Hal ini dikarenakan *Fuzzy C-Means* memiliki derajat keanggotaan yang berguna untuk mengelompokkan data ke dalam *cluster* yang semestinya.

## 4.3 Analisa

### 1. Analisa Tahap *Data Mining*

Berikut adalah tahap-tahap *data mining* berdasarkan yang sebelumnya telah dibahas pada Bab 2:

#### a. Pembersihan data

Pada penelitian ini pembersihan data dilakukan secara manual. Data yang didapatkan dari pihak PT Sepatu Bata yaitu berupa *print out*. Data-data tersebut harus diinputkan kembali ke dalam komputer dalam format Microsoft Excel (.xls) agar dapat di *import* ke dalam aplikasi. Tidak seluruh data akan dipakai dalam penelitian ini, dan hanya dipilih beberapa jenis data seperti nomor *invoice* dan jumlah barang.

#### b. Integrasi data

Penggabungan data pada penelitian ini juga dilakukan secara manual. Dari pihak PT Sepatu Bata data yang didapatkan adalah *print out* per hari. Data-data tersebut digabungkan ke dalam satu *file* Microsoft Excel (.xls) yang nantinya akan digunakan untuk proses *import* data.

#### c. Transformasi data

Pada tahap ini, transformasi data telah dilakukan sebelumnya yaitu pada tahap pembersihan data dan integrasi data.

#### d. *Data mining*

Pada penelitian ini adalah proses dimana *clustering* dicari, yaitu metode perhitungan berdasarkan *Fuzzy C-Means* dan akan menghasilkan *output* berupa kelompok-kelompok barang berdasarkan tingkat kelarisan penjualannya.

#### e. Evaluasi pola

Pada tahap ini didapatkan hasil dari proses *data mining* berupa tiga *cluster* tingkat kelarisan produk-produk di PT Sepatu Bata, yaitu kurang laris, laris, dan sangat laris. Berikut ini adalah contoh hasil analisa yang bisa disimpulkan dari hasil proses perhitungan teknik *data mining* menggunakan *Fuzzy C-Means* pada aplikasi ini:

1. Produk-produk PT Sepatu Bata yang sangat laris antara lain: 572 dan 872.
2. Produk-produk PT Sepatu Bata yang laris antara lain: 371, 381, 551, 552, 571, 581, 871, dan 873.

3. Produk-produk PT Sepatu Bata yang kurang laris antara lain: 001, 058, 150, 151, 181, 189, 309, 351, 361, 389, 461, 471, 481, 489, 508, 511, 518, 531, 561, 589, 591, 631, 640, 651, 661, 671, 761, 771, 801, 808, 809, 811, 814, 818, 821, 831, 832, 839, 848, 851, 861, 864, 874, 881, 889, 900, 902, 910. 930, 940, 952, 980, dan 982.

f. Presentasi pengetahuan

Pada tahap terakhir dari *data mining* ini, presentasi pengetahuan mengenai metode *Fuzzy C-Means* yang digunakan untuk memperoleh informasi mengenai kelarisan produk-produk PT Sepatu Bata adalah *output* berupa grafik berbentuk *pie* dengan keterangan-keterangan seperti kurang laris, laris, dan sangat laris dan tabel untuk mengetahui lebih *detail* mengenai produk-produk yang masuk ke dalam masing-masing kategori tersebut untuk mempermudah pengguna dalam memperoleh informasi yang dihasilkan.

2. Analisa Kuesioner

Untuk melihat tingkat keberhasilan dari penelitian yang telah dibuat ini, dilakukan analisa melalui media kuesioner. Kuesioner diberikan kepada karyawan toko PT Sepatu Bata. Berdasarkan pernyataan-pernyataan yang diberikan kepada responden, persentase jawaban sebesar 91,67% yaitu sangat setuju dari segi pernyataan yang berhubungan dengan *Fuzzy C-Means*. Aplikasi ini mempermudah pengguna dalam menemukan pola atau pengetahuan baru dalam penjualan produk, mengidentifikasi tingkat penjualan produk-produk, serta penyajian informasi kelarisan produk yang sudah tepat sasaran. Dari segi penggunaan, persentase jawaban sebesar 90,5% sangat setuju. Aplikasi ini mudah digunakan dan pengguna merasa nyaman akan antarmuka aplikasi ini (*user friendly*).

## 5. PENUTUP

### 5.1 Kesimpulan

Setelah dilakukan pengujian beserta analisa pada penelitian ini, maka dapat diambil kesimpulan sebagai berikut:

1. Metode perhitungan *Fuzzy C-Means* cocok digunakan pada aplikasi ini, karena bisa menghasilkan *output* berupa tingkat kelarisan produk-produk pada penjualan di PT Sepatu Bata.
2. Dari hasil kuesioner dengan pegawai PT Sepatu Bata terhadap aplikasi ini sebanyak 91,2% responden menyatakan bahwa aplikasi ini sudah bekerja dengan sangat baik dalam penyajian informasi yang dibutuhkan.

### 5.2 Saran

Untuk pengembangan aplikasi ini, maka beberapa hal yang dapat penulis sarankan adalah:

1. Diharapkan pemilihan event dalam melakukan data mining tidak hanya dua bulan saja seperti kasus yang diangkat tetapi juga berdasarkan event tertentu seperti lebaran, natal, dan tahun baru serta berdasarkan rentang waktu tertentu seperti per tahun atau beberapa bulan sekali. Aplikasi ini dapat dikembangkan dengan meng-*cluster* produk-produk berdasarkan sub-modelnya.
2. Aplikasi ini dapat dikembangkan dengan meng-*cluster* produk-produk berdasarkan varian dari tiap-tiap modelnya (misalnya sepatu-sepatu yang laris berdasarkan warna merah dari model 'Bubblegummers').

## DAFTAR PUSTAKA

- [1] Han, Jiawei, Micheline Kamber. 2001. *Data Mining Concepts and Techniques*. Morgan Kaufmann: California.
- [2] Jain, A. K., Murthy, M. N. & Flynn, P. J. 1999. *Data Clustering: A Review*. ACM Computing Surveys, Vol. 31, No. 3.
- [3] 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.
- [4] Klawonn, F. Keller, A. 1997. *Fuzzy Clustering and Fuzzy Rules*. Science Journal.
- [5] Klir, G. J., Yuan, B. 1995. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice-Hall International, Inc. New Jersey, USA.
- [6] Kusumadewi, S., Purnomo, H. (2004). *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*. Yogyakarta: Graha Ilmu.
- [7] Nugraheni, Yohana. (2011). *Data Mining Dengan Metode Fuzzy Untuk Customer Relationship Management (CRM) pada Perusahaan Retail*. Diambil 29 Oktober 2012 dari [http://www.pps.unud.ac.id/thesis/pdf\\_thesis/unud-212-866836442-thesis\\_yohana.pdf](http://www.pps.unud.ac.id/thesis/pdf_thesis/unud-212-866836442-thesis_yohana.pdf).
- [8] Sumathi, S., Sivanandam, S.N. 2006. *Introduction to Data Mining and its Applications*. Spinger, Verlag Berlin Heidelberg.
- [9] Tan, P.N., Steinbach, M., Kumar, V. 2006. *Introduction to Data Mining*. Pearson Education, Inc.