

**PERBANDINGAN 3 METODE DALAM DATA MINING UNTUK MENENTUKAN  
STRATEGI PENJUALAN PRODUK MAKANAN DAN MINUMAN  
PADA TOSERBA LESTARI BARU GELOMONG**

**NASKAH PUBLIKASI**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA**



Diajukan Oleh :

**Deny Arga Felani**

**Yusuf Sulisty Nugroho, S.T, M.Eng**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA UNIVERSITAS  
MUHAMMADIYAH SURAKARTA**

**MARET 2015**

**HALAMAN PENGESAHAN**

**Publikasi ilmiah dengan judul :**

**PERBANDINGAN 3 METODE DALAM DATA MINING UNTUK MENENTUKAN  
STRATEGI PENJUALAN PRODUK MAKANAN DAN MINUMAN  
PADA TOSERBA LESTARI BARU GELOMONG**

Yang dipersiapkan dan disusun oleh :

**Deny Arga Felani**

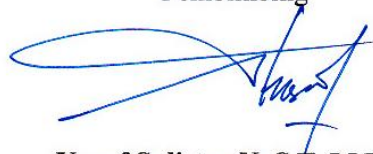
**L200110149**

Telah disetujui pada :

Hari : Jumat .....

Tanggal : 6 Maret 2015 .....

Pembimbing



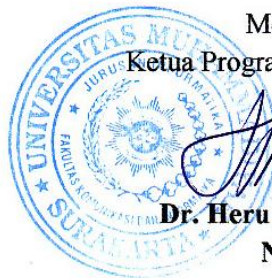
**Yusuf Sulistyo N, S.T, M.Eng**

**NIK : 1197**

Publikasi ilmiah ini telah diterima sebagai salah satu  
persyaratan untuk memperoleh gelar sarjana

Tanggal 16 Maret 2015

Mengetahui,  
Ketua Program Studi Informatika



**Dr. Heru Supriyono, M.sc**

**NIK: 970**



**UNIVERSITAS MUHAMMADIYAH SURAKARTA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
PROGRAM STUDI INFORMATIKA**

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448  
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id>. Email: [informatika@fki.ums.ac.id](mailto:informatika@fki.ums.ac.id)

---

**SURAT KETERANGAN LULUS PLAGIASI**

**/A.3-II.3/INF-FKI/III/2015**

Assalamu'alaikum Wr. Wb

Biro Skripsi Program Studi Informatika menerangkan bahwa :

Nama : DENY ARGA FELANI  
NIM : L200110149  
Judul : PERBANDINGAN 3 METODE DALAM DATA MINING UNTUK  
MENENTUKAN STRATEGI PENJUALAN PRODUK MAKANAN  
DAN MINUMAN PADA TOSERBA LESTARI BARU GEMOLONG  
Program Studi : Informatika  
Status : **Lulus**

Adalah benar-benar sudah lulus pengecekan plagiasi dari Naskah Publikasi Skripsi, dengan menggunakan aplikasi Turnitin.

Demikian surat keterangan ini dibuat agar dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Wr. Wb

*Surakarta, 10 Maret 2015*

Biro Skripsi  
Informatika

  
**Adjie Sapoetra, S.Kom**

**Turnitin Originality Report**

PERBANDINGAN 3 METODE DALAM DATA MINING UNTUK MENENTUKAN STRATEGI PENJUALAN PRODUK MAKANAN DAN MINUMAN PADA TOSERBA LESTARI BARU GELOMONG  
by Deny Arga Felani

Similarity Index <b>22%</b>	Similarity by Source	
	Internet Sources:	16%
	Publications:	1%
	Student Papers:	12%

From publikasi (publikasi)

**sources:**

Processed on 10-Mar-2015 13:08 WIB  
ID: 514646729  
Word Count: 1635

Paper ID: 434538412

- 1 3% match (student papers from 13-Jun-2014)  
Class: publikasi maret 2014  
Assignment:

---

- 2 3% match (Internet from 04-Oct-2012)  
<http://arifds.com/programming/mengenal-algoritma-c4-5-dari-dasar>

---

- 3 2% match (Internet from 04-Apr-2011)  
<http://eprints.undip.ac.id/9335/>

---

- 4 2% match (student papers from 17-Jun-2014)  
Class: publikasi maret 2014  
Assignment:  
Paper ID: 435295687

---

- 5 2% match (Internet from 25-Jun-2013)  
<http://sariberbagilmu.blogspot.com/2013/06/memo-ty-d-ncans.html>

---

- 6 2% match (Internet from 09-Aug-2013)  
[http://upi-yptk.ac.id/ejournal/File\\_Jurnal/Aldino\\_Herman-Jurnal.pdf](http://upi-yptk.ac.id/ejournal/File_Jurnal/Aldino_Herman-Jurnal.pdf)

---

- 7 2% match (student papers from 04-Feb-2014)  
Class: publikasi maret 2014  
Assignment:  
Paper ID: 393366374

---

- 8 1% match (Internet from 16-Apr-2011)

---

- 9 1% match (student papers from 20-Jul-2013)  
Submitted to Universitas Muhammadiyah Surakarta on 2013-07-20

---

- 10 1% match (student papers from 25-Jul-2013)  
Submitted to Universitas Muhammadiyah Surakarta on 2013-07-25

---

- 11 1% match (student papers from 04-Feb-2014)  
Class: publikasi maret 2014  
Assignment:  
Paper ID: 393332890

---

- 12 1% match (Internet from 31-Dec-2014)  
<http://sonnyintias.blogspot.com/>

---

- 1% match (Internet from 15-Sep-2011)

**PERBANDINGAN 3 METODE DALAM DATA MINING UNTUK MENENTUKAN  
STRATEGI PENJUALAN PRODUK MAKANAN DAN MINUMAN  
PADA TOSERBA LESTARI BARU GELOMONG**

Deny Arga Felani, Yusuf Sulisty Nugroho  
Program Studi Informatika, Fakultas Komunikasi dan Informatika  
Universitas Muhammadiyah Surakarta  
Email : [argaseifer@gmail.com](mailto:argaseifer@gmail.com)

*Abstraksi*

Toserba Lestari Baru adalah Perusahaan besar yang bergerak di bidang penjualan. Data yang tersimpan akan dimanfaatkan menggunakan data Mining sehingga dapat digunakan sebagai pendukung keputusan untuk mencari suatu informasi yang berguna dalam evaluasi data yang digunakan. Berbagai metode yang terdapat dalam data Mining, maka penulis akan melakukan perbandingan 3 metode dari data Mining tersebut.

Penggunaan metode *decision tree*, *K-means (clustering)* dan regresi linear diimplementasikan menggunakan aplikasi RapidMiner, yang nantinya akan dilakukan analisis dari masing-masing metode tersebut untuk menentukan strategi penjualan pada Toserba Lestari Baru Gemolong.

Penelitian ini dilakukan dengan sekelompok data untuk mengetahui persentase nilai *precision*, *recall* dan *accuracy*. Hasil dari penelitian tersebut bahwa metode *K-Means (clustering)* memiliki nilai lebih baik dari metode yang lain pada sisi *precision* dan *accuracy*, sedangkan metode Regresi Linier memiliki nilai *recall* lebih baik dari metode yang lain.

**Kata Kunci : *Decision Tree*, *Clustering*, Regresi linier.**

## **PENDAHULUAN**

Toserba Lestari Baru adalah sebuah perusahaan besar yang berlokasi di Jl. Sukowati Gemolong dan bergerak di bidang penjualan. Perusahaan ini setiap harinya harus memenuhi kebutuhan konsumen dan dituntut untuk dapat mengambil keputusan yang tepat dalam menentukan strategi penjualan. Untuk dapat melakukan hal tersebut, perusahaan membutuhkan sumber informasi yang cukup banyak untuk dapat di analisis lebih lanjut.

Pada Toserba Lestari Baru banyak sekali transaksi yang tercatat. Ketersediaan data penjualan yang sangat besar cenderung hanya untuk dijadikan arsip dan laporan serta belum dimanfaatkan secara optimal hal ini disebabkan belum adanya sistem pendukung keputusan dan metode yang dapat digunakan untuk merancang sebuah strategi bisnis dalam meningkatkan penjualan.

Dengan melakukan data mining, ada beberapa metode yang dapat digunakan untuk melakukan perbandingan dalam kasus diatas, diantaranya menggunakan algoritma C.45, *K-Means* dan Regresi linier.

Berdasarkan permasalahan tersebut, maka dalam penelitian ini akan dilakukan perbandingan dengan menggunakan 3 algoritma dalam data mining untuk menentukan metode yang terbaik dalam mengetahui evaluasi dalam penjualan yang bisa dimanfaatkan oleh perusahaan.

## **TINJAUAN PUSTAKA**

### **1. Data Mining**

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Pemilihan metode yang tepat tergantung pada tujuan dan proses KDD (*Knowledge Discovery in Database*) secara keseluruhan. Data mining mampu menganalisis data yang besar menjadi informasi berupa pola yang mempunyai arti bagi pendukung keputusan. (Sunjaya, 2010)

### **2. Algoritma C.4.5**

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Sedang pohon keputusan dapat diartikan suatu cara untuk mengklarifikasi yang sangat kuat.

Pohon keputusan dapat membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan. (Kusrini dan Luthfi, 2009)

### 3. Algoritma *K-Means*

*K-Means* merupakan salah satu metode pengelompokan data *non hierarki* yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain. (Prasetyo, 2012).

### 4. Regresi Linier

Analisis regresi adalah teknik statistik untuk permodelan dan investigasi hubungan dua atau lebih variabel. Yang sering dipakai adalah regresi linear sederhana. Dalam analisis regresi satu atau lebih variabel *independent / predictor* yang biasa diwakili dengan notasi  $x$  dan satu variabel respon yang biasa diwakili dengan notasi  $y$ . (Santosa, 2007)

## METODE PENELITIAN

### 1. Identifikasi Masalah

Penggunaan data mining untuk menentukan strategi penjualan adalah masalah yang harus diselesaikan, untuk mengetahui faktor yang mempengaruhi.

Metode yang digunakan ada beberapa, dari hal tersebut nanti akan dilakukan perbandingan dari beberapa metode tersebut untuk menentukan metode yang paling baik dalam menangani masalah tersebut.

### 2. Analisis Kebutuhan

Analisis kebutuhan adalah menentukan kebutuhan apa yang diperlukan seperti perangkat keras, perangkat lunak dan materi yang mendukung.

Adapun tahap yang diperlukan dalam penelitian antara lain :

- a. Analisis data untuk pemilihan atribut yang diperlukan dalam melakukan penelitian.

Tabel 1. Data Atribut

Atribut	Variabel
Jumlah terjual	Y
Kategori Produk	X1
Harga	X2
Bulan	X3
Tempat	X4



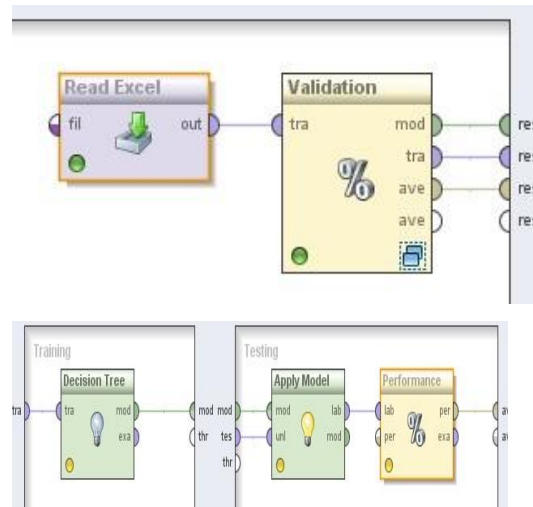
## HASIL DAN PEMBAHASAN

Penelitian yang dilakukan bertujuan melakukan perbandingan metode dalam data mining untuk mengklasifikasikan data penjualan pada Toserba Lestari Baru Gemolong menggunakan variabel-variabel yang sudah ditentukan.

### 1. Hasil Implementasi *Decision Tree* pada Rapid Miner.

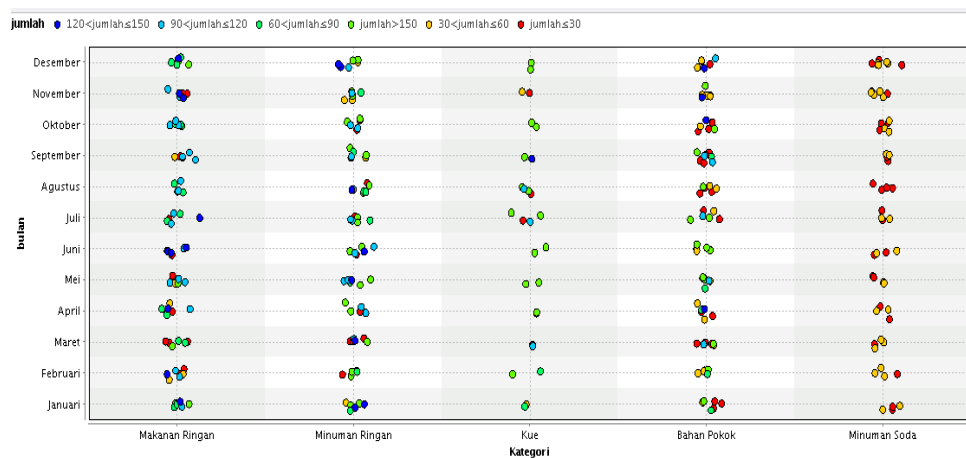
Pada implementasi *data mining* ini untuk menganalisa penentuan kalsifikasi penjualan pada Toserba Lestari Baru menggunakan aplikasi *Rapid Miner 5*. Dalam analisa pohon keputusan (*Decision Tree*) tersebut.

Rancangan proses dengan metode *Decision Tree* menggunakan aplikasi *RapidMiner 5* ditunjukkan pada Gambar 1.



**Gambar 1.** Rancangan *Decision Tree*

Hasil gambar 1 kemudian dieksekusi dan dilihat menggunakan *Scatter Plot* dilihat pada gambar 2.



**Gambar 2.** Hasil RapidMiner pada metode *Decision Tree*



## 2. Hasil Implementasi *K-means* (*Clustering*) pada Rapid Miner

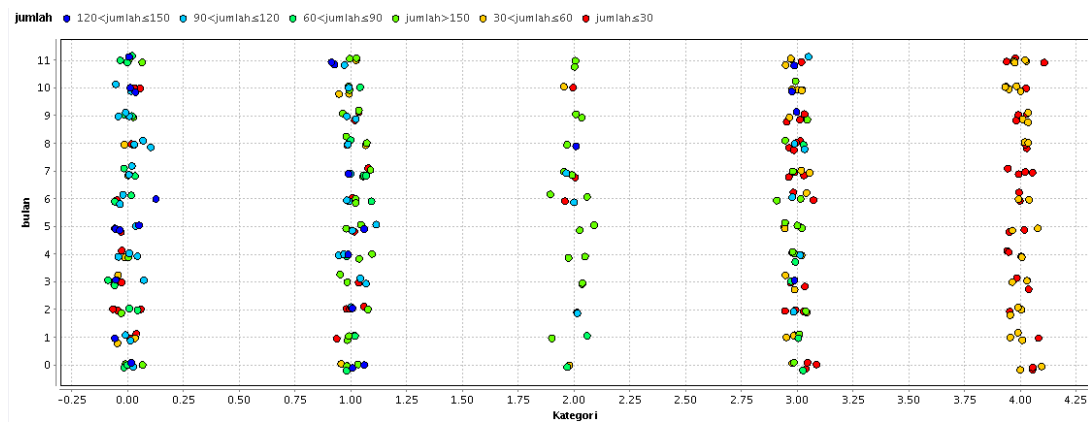
Pada implementasi *data mining* ini untuk menganalisa penentuan klasifikasi penjualan pada Toserba Lestari Baru menggunakan aplikasi *RapidMiner*. Dalam analisa Clustering tersebut,

Rancangan proses dengan metode *Clustering* menggunakan aplikasi *Rapid Miner 5* ditunjukkan pada Gambar 3.



**Gambar 3.** Rancangan *Clustering*

Hasil gambar 3 kemudian dieksekusi dan dilihat menggunakan *Scatter Plot* dilihat pada gambar 4.

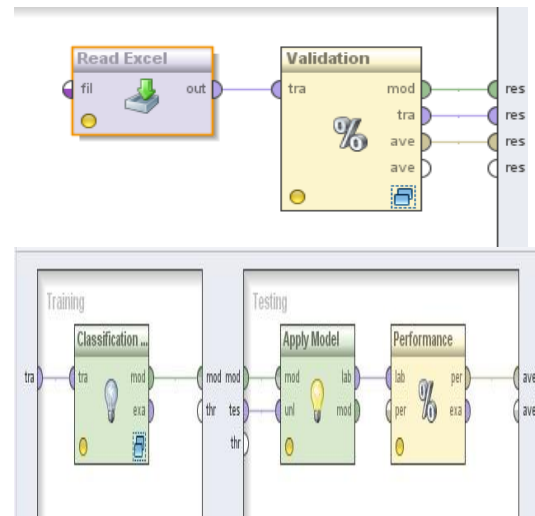


**Gambar 4.** Hasil RapidMiner pada metode *K-Means* (*Clustering*)

### 3. Hasil Implementasi Regresi Linear pada Rapid Miner

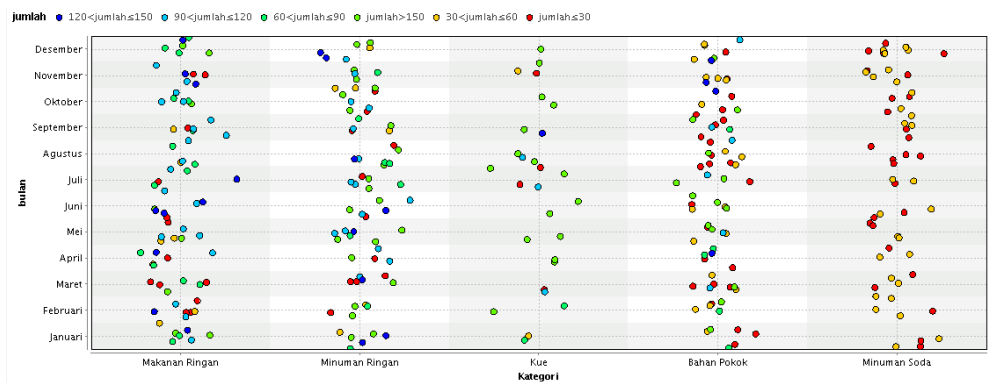
Pada implementasi data mining ini untuk menganalisa penentuan kalsifikasi penjualan pada Toserba Lestari Baru menggunakan aplikasi RapidMiner.

Rancangan proses dengan metode *Regresi Linear* menggunakan aplikasi *RapidMiner* ditunjukkan pada gambar 5.



**Gambar 5.** Rancangan Regresi Linier

Kemudian dieksekusi untuk melihat hasil dari *scatter plot*, hasil dilihat pada gambar 6.



**Gambar 6.** Hasil RapidMiner pada metode Regresi Linier

## ANALISA DAN PEMBAHASAN

### 1. Implementasi *Decision Tree*

Berikut tahap-tahap implementasi pada metode *decision tree* :

#### a. Menentukan Root

Hasil *Information Gain* masing-masing variabel pada **tabel 2**.

**Tabel 2.** Hasil *root node*

Variabel	Keterangan	Gain
X1	Kategori	0,560
X2	Harga	0,659
X3	Bulan	0,292
X4	Tempat	0,012

Berdasarkan perhitungan tabel 2 didapat nilai *Information Gain* tertinggi berupa variabel harga, dimana memiliki nilai sebesar 0,659 bits dan berhak menduduki posisi *root* (akar).

#### b. Menentukan Internal node

Hasil *Information Gain* dilihat dalam tabel 3.

**Tabel 3.** Hasil *Internal node*

Variabel	Keterangan	Gain
X1	Kategori	0,639
X3	Bulan	2,194
X4	Tempat	0,058

Berdasarkan hasil perhitungan tabel 3 didapat nilai *Information Gain* tertinggi berupa variabel Bulan, dimana memiliki

nilai sebesar 2,194 bits dan berhak menduduki *internal node* pada pohon keputusan.

#### c. Menentukan *Leaf node*

Hasil *Information Gain* dari variabel harga yang berupa harga  $\leq 3000$  dan bulan Januari diperoleh seperti dalam tabel 4.

**Tabel 4.** Hasil *leaf node*

Variabel	Keterangan	Gain
X1	Kategori	0,000
X4	Tempat	0,000

Berdasarkan perhitungan tabel 4 didapat nilai *Information Gain* sama-sama memiliki hasil 0,000 bits, sehingga untuk *internal node* selanjutnya diambil dari salah satu variabel. Contohnya adalah variabel Kategori.

### 2. Implementasi *K-means (clustering)*

Pada metode ini dilakukan perhitungan dengan berbagai tahap, diantaranya :

#### a. Menentukan *cluster*

Pada tahap ini menggunakan 5 cluster

#### b. Menentukan *Centroid*

**Tabel 5.** Data *Centroid*

<i>Cluster</i>	<i>Centroid</i>			
	X1	X2	X3	X4
1-60	5,4	6,2	52,4	0
61-120	6,2	7	59,2	9,4
121-180	21,2	22,8	76,8	5
181-240	33,6	36,8	75,6	9,6
241-302	47,6	37,8	77,6	6,2

c. Menentukan *Ecludian distance*

Pada tahap ini menggunakan rumus *ecludian distance* untuk mencari jarak antar data dalam melakukan pengelompokan.

3. Implementasi Regresi Linier

Pada metode ini menggunakan hasil rumus yang dihasilkan oleh RapidMiner.

- a.  $- 0.065 * \text{Kategori} - 0.725$
- b.  $- 0.239 * \text{Kategori} + 0.130 * \text{Harga} - 0.496$
- c.  $- 0.169 * \text{Kategori} + 0.091 * \text{Harga} - 0.642$
- d.  $0.136 * \text{Kategori} - 0.236 * \text{Harga} - 0.421$
- e.  $0.194 * \text{Kategori} - 0.949$
- f.  $0.131 * \text{Kategori} + 0.028 * \text{Harga}$

**PERBANDINGAN METODE**

Setelah semua metode dianalisa maka selanjutnya melakukan perbandingan 3 metode tersebut dalam hitungan *Precision*, *Recall*, dan *Accuracy*.

**Tabel 6.** Data Perbandingan.

<b>Metode</b>	<b><i>Precision</i></b>	<b><i>Recall</i></b>	<b><i>Accuracy</i></b>
<i>Decision Tree</i>	32,94%	45.90%	25,50%
<i>K-Means</i>	42,60%	40,31%	41,40%
Regresi Linier	32,35%	49,18%	29,47%

Pada perbandingan Tabel 6 menunjukkan bahwa *K-Means (Clustering)* merupakan metode yang memiliki tingkat *precision* dan *accuracy* yang lebih tinggi dari metode yang lain, kemudian pada *recall* regresi linear memiliki nilai lebih tinggi daripada metode yang lain.

**KESIMPULAN**

Berdasarkan hasil penelitian, maka dapat ditarik kesimpulan bahwa :

1. Berdasarkan hasil perbandingan dari 3 metode yaitu *Decision tree*, *K-means* dan Regresi Linear, variabel yang paling berpengaruh terhadap hasil penjualan adalah variabel harga.
2. Berdasarkan dari nilai *precision* dan *accuracy*, metode *K-Means* memiliki nilai lebih tinggi daripada algoritma yang lain dengan nilai *precision* sebesar 42,60% dan nilai *accuracy* sebesar 41,40%.
3. Berdasarkan nilai *recall*, regresi linier memiliki nilai lebih tinggi dibandingkan dengan metode yang lain dengan nilai *recall* sebesar 49,18% .

## DAFTAR PUSTAKA

- Kusrini dan Luthfi, E.T, 2009, *Algoritma Data Mining*, Andi Offset, Yogyakarta
- Prasetyo, Eko.2012. *Data Mining Konsep dan Aplikasi Menggunakan Matlab*.Yogyakarta: Andi
- Santoso, Budi. 2007. *Data Mining:Teknik Pemanfaatan data untuk keperluan bisnis*, Yogyakarta : Graha Ilmu.
- Santoso, S. 2010. *Statistik Multivariat*. Jakarta : Elex Media Komputindo.
- Sunjaya. 2010. *Aplikasi Mining Data Mahasiswa dengan Metode Klasifikasi DecisionTree*. Seminar Nasional Aplikasi Teknologi Informasi 2010. Yogyakarta.

## **BIODATA PENULIS**

Nama : Deny Arga Felani  
NIM : L200110149  
Tempat, Tanggal Lahir : Surakarta, 21 Desember 1993  
Jenis Kelamin : Laki-laki  
Agama : Islam  
Pendidikan : S1  
Fakultas : Informatika  
Universitas : Universitas Muhammadiyah Surakarta  
Alamat : Jl. A. Yani Tromol Pos 1 Pabelan, Kartasura  
Telp./Fax : (0271)717417, 719483 / (0271)714448  
Alamat Rumah : Karangjati, Kalijambe, Sragen  
Nomor HP : 085642188421  
Email : argaseifer@gmail.com