

**PEMANFAATAN DATA MINING PADA INDUSTRI  
KNALPOT UNTUK MENERAPKAN *CUSTOMER  
RELATIONSHIP MANAGEMENT***



**TESIS**

Diajukan untuk memenuhi salah satu syarat kelulusan Program Strata 2 (S2)

**INA MARYANI**

14001675

**PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER  
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER**

**NUSA MANDIRI**

**JAKARTA**

2016

ii

**Program Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri**

## SURAT PERNYATAAN ORISINILITAS

Yang bertanda tangan di bawah ini :

Nama : Ina Maryani  
NIM : 14001675  
Program Studi : Magister Ilmu Komputer  
Jenjang : Strata Dua (S2)  
Konsentrasi : *Management Information System*

Dengan ini menyatakan bahwa tesis yang telah saya buat dengan judul: "Pemanfaatan Data Mining Pada Industri Knalpot Untuk Menerapkan *Customer Relationship Management*" adalah hasil karya sendiri, dan semua sumber baik yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar dan tesis belum pernah diterbitkan atau dipublikasikan dimanapun dan dalam bentuk apapun.

Demikianlah surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila dikemudian hari ternyata saya memberikan keterangan palsu dan atau ada pihak lain yang mengklaim bahwa tesis yang telah saya buat adalah hasil karya milik seseorang atau badan tertentu, saya bersedia diproses baik secara pidana maupun perdata dan kelulusan saya dari Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri dicabut/dibatalkan.

Jakarta, 22 Agustus 2016  
Yang menyatakan,



Ina Maryani

## HALAMAN PENGESAHAN

Tesis ini diajukan oleh :

Nama : Ina Maryani  
NIM : 14001675  
Program Studi : Magister Ilmu Komputer  
Jenjang : Strata Dua (S2)  
Konsentrasi : *Management Information System*  
Judul Tesis : "Pemanfaatan Data Mining Pada Industri Knalpot Untuk Menerapkan *Customer Relationship Management*"

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Magister Ilmu Komputer (M.Kom) pada Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri).

Jakarta, 26 Agustus 2016  
Pascasarjana Magister Ilmu Komputer  
STMIK Nusa Mandiri  
Direktur

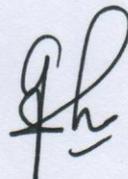
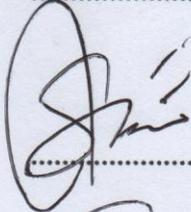
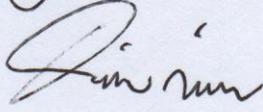
Prof. Dr. Ir. R. Eko Indrajit, M.Sc, MBA

## DEWAN PENGUJI

Penguji I : Dr. Sularso Budilaksono, M.Kom

Penguji II : Dr. Alimuddin, MM.MT

Penguji III /  
Pembimbing : Dr. Dwiza Riana, S.Si, MM, M.Kom

  
.....  
  
25/04/2016  
Jan 15 00  
.....  
  
.....

## KATA PENGANTAR

Puji syukur alhamdulillah, penulis panjatkan kehadirat Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya, sehingga pada akhirnya penulis dapat menyelesaikan tesis ini tepat pada waktunya. Dimana tesis ini penulis sajikan dalam bentuk buku yang sederhana. Adapun judul tesis, yang penulis ambil sebagai berikut “Pemanfaatan Data Mining Pada Industri Knalpot Untuk Menerapkan *Customer Relationship Management*”.

Tujuan penulisan tesis ini dibuat sebagai salah satu syarat untuk mendapatkan gelar Magister Ilmu Komputer (M.Kom) pada Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (PPs MIK STMIK Nusa Mandiri).

Tesis ini diambil berdasarkan data transaksi yang ada pada industri knalpot dan mobil doctor speed Purbalingga. Penulis juga mencari dan menganalisa berbagai macam sumber referensi, baik dalam bentuk jurnal ilmiah, buku-buku literatur, internet, dll yang terkait dengan pembahasan pada tesis ini.

Penulis menyadari bahwa tanpa bimbingan dan dukungan dari semua pihak dalam pembuatan tesis ini, maka penulis tidak dapat menyelesaikan tesis ini tepat pada waktunya. Untuk itu ijinilah penulis pada kesempatan ini untuk mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Ir. R. Eko Indrajit, M.Sc, MBA selaku Direktur Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri.
2. Bapak Dr. Mochamad Wahyudi, M.M, M.Kom, M.Pd selaku Ketua Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri.
3. Ibu Dr. Dwiza Riana, S.Si, MM, M.Kom yang sudah dengan setulus hati membimbing, mengarahkan menyumbangkan ide, waktu, dan tenaganya dalam membimbing penulis untuk menyelesaikan tesis ini.
4. Bapak Ir. Naba Aji Notoseputro selaku Direktur AMIK BSI yang telah mengijinkan penulis untuk menyelesaikan pendidikan S2.
5. Orang tua tercinta serta seluruh keluarga yang telah memberikan dukungan material dan moral kepada penulis.

6. Seluruh staf pengajar (dosen) Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri yang telah memberikan pelajaran yang berarti bagi penulis selama menempuh studi.
7. Seluruh staf dan karyawan Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri yang telah melayani penulis dengan baik selama kuliah.
8. Teman-teman satu kelas, sahabat yang telah berjuang bersama-sama.

Serta semua pihak yang terlalu banyak untuk penulis sebutkan satu persatu sehingga terwujudnya penulisan tesis ini. Penulis menyadari bahwa penulisan tesis ini masih jauh sekali dari sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan karya ilmiah yang penulis hasilkan untuk yang akan datang.

Akhir kata semoga tesis ini dapat bermanfaat bagi penulis khususnya dan bagi para pembaca yang berminat pada umumnya.

Jakarta, 22 Agustus 2016

Ina Maryani

**SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH  
UNTUK KEPENTINGAN AKADEMIS**

Yang bertanda tangan di bawah ini, saya :

Nama : Ina Maryani  
NIM : 14001675  
Program Studi : Magsiter Ilmu Komputer  
Jenjang : Strata Dua (S2)  
Konsentrasi : *Management Information System*  
Jenis Karya : Tesis

Demi pengembangan ilmu pengetahuan, dengan ini menyetujui untuk memberikan ijin kepada pihak Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri) **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalti-Free Right*)** atas karya ilmiah kami yang berjudul : “Pemanfaatan Data Mining Pada Industri Knalpot Untuk Menerapkan *Customer Relationship Management*” beserta perangkat yang diperlukan (apabila ada).

Dengan **Hak Bebas Royalti Non-Eksklusif** ini pihak STMIK Nusa Mandiri berhak menyimpan, mengalih-media atau *bentuk*-kan, mengelolanya dalam pangkalan data (*database*), mendistribusikannya dan menampilkan atau mempublikasikannya di *internet* atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari kami selama tetap mencantumkan nama kami sebagai penulis/pencipta karya ilmiah tersebut.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak STMIK Nusa Mandiri, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 22 Agustus 2016



Ina Maryani



6. Seluruh staf pengajar (dosen) Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri yang telah memberikan pelajaran yang berarti bagi penulis selama menempuh studi.
7. Seluruh staf dan karyawan Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri yang telah melayani penulis dengan baik selama kuliah.
8. Teman-teman satu kelas, sahabat yang telah berjuang bersama-sama.

Serta semua pihak yang terlalu banyak untuk penulis sebutkan satu persatu sehingga terwujudnya penulisan tesis ini. Penulis menyadari bahwa penulisan tesis ini masih jauh sekali dari sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan karya ilmiah yang penulis hasilkan untuk yang akan datang.

Akhir kata semoga tesis ini dapat bermanfaat bagi penulis khususnya dan bagi para pembaca yang berminat pada umumnya.

Jakarta, 22 Agustus 2016



Ina Maryani

Penulis

## DAFTAR ISI

	Halaman
HALAMAN SAMPUL .....	i
HALAMAN JUDUL.....	ii
HALAMAN PERNYATAAN ORISINALITAS.....	iii
HALAMAN PENGESAHAN.....	iv
KATA PENGANTAR.....	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS.....	vii
ABSTRAK.....	viii
ABSTRCT.....	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN.....	xiv
BAB 1. PENDAHULUAN.....	1
1.1. Latar Belakang .....	1
1.2. Masalah Penelitian .....	3
1.2.1. Identifikasi Masalah.....	3
1.2.2. Batasan Masalah .....	4
1.2.3. Rumusan Masalah.....	4
1.3. Tujuan dan Manfaat Penelitian .....	5
1.4. Sistematika Penulisan .....	5
BAB II. LANDASAN TEORI .....	7
2.1. Tinjauan Pustaka .....	7
2.1.1. Data Mining .....	7
2.2. Customer Relationship Management (CRM) .....	11
2.3. Teknologi CRM .....	12
2.4. Data Mining Dalam Kerangka Kerja CRM .....	13
2.5. Segmentasi Pelanggan .....	14
2.6. Clustering.....	17
2.6.1. Keuntungan Algoritma K-Means.....	18
2.7. Klasifikasi (Classification).....	19
2.7.1. Metode Decision Tree.....	19
2.7.2. Implementasi Decission Tree Dalam WEKA.....	19
2.8. Tinjauan Studi.....	20
2.9. Tinjauan Objek Penelitan.....	25
BAB III. METODOLOGI PENELITIAN.....	27
3.1. Metodologi Penelitian.....	27
3.1.1. Jenis Penelitian .....	28
3.1.2. Metode Pengumpulan Data.....	29
3.1.3. Eksperimen .....	29
3.2. Jadwal Penelitan.....	36

BAB IV. HASIL PENELITIAN DAN PEMBAHASAN .....	37
4.1. Hasil Penelitian .....	37
4.1.1. Persiapan Data Mining.....	37
4.2. Pembahasan.....	43
4.2.1. Hasil Penghitungan Cluster .....	43
4.2.2. Hasil Cluster Optimal .....	45
4.2.3. Karakteristik Pelanggan .....	45
4.2.4. Model Klasifikasi.....	46
4.2.5. Deployment.....	53
4.3. Implikasi Penelitian .....	56
4.3.1. Aspek Sistem .....	56
4.3.2. Aspek Manajerial .....	57
4.3.3. Aspek Penelitian Lanjutan .....	57
BAB V. PENUTUP.....	59
5.1. Kesimpulan .....	59
5.2. Saran .....	60
DAFTAR REFERENSI .....	61
SURAT KETERANGAN RISET/PRAKTEK KERJA LAPANGAN .....	64
DAFTAR RIWAYAT HIDUP .....	65
LAMPIRAN-LAMPIRAN .....	66

## DAFTAR TABEL

	Halaman
Tabel II.1 Tinjauan Penelitian Sebelumnya Dari <i>International Journal</i> .....	21
Tabel II.2 Tinjauan Penelitian Sebelumnya Dari Jurnal <i>National</i> .....	22
Tabel III.1 <i>Variabel</i> Proses <i>Clustering</i> .....	30
Tabel III. 2 Menunjukkan <i>Score Recency, Frequency dan Monetary</i> .....	31
Tabel III.3 Jadwal Penelitian.....	36
Tabel IV.1 Dataset Transaksi Penjualan .....	38
Tabel IV.2 Kategori <i>Recency</i> .....	39
Tabel IV.3 Kategori <i>Frequency</i> .....	39
Tabel IV.4 Kategori <i>Monetary</i> .....	40
Tabel IV.5 Pembobot RFM.....	40
Tabel IV.6 Data Penjualan Score RFM.....	40
Tabel IV.7 Titik Pusat Awal Setiap Cluster.....	41
Tabel IV.8 Contoh Hasil Perhitungan Setiap Data Ke Setiap <i>Cluster</i> .....	42
Tabel IV.9 Perbandingan <i>Performa Cluster</i> .....	45
Tabel IV.10 Dataset Transaksi Pelanggan .....	46
Tabel IV.11 Format .arff Data Cluster .....	46
Tabel IV.12 Karakteristik Pelanggan Setiap <i>Cluster</i> .....	49
Tabel IV.13 Spesifikasi Hardware dan <i>Software</i> .....	56

## DAFTAR GAMBAR

	Halaman
Gambar II.1 Proses <i>Knowledge Discovery From Data</i> .....	7
Gambar II.2 Proses <i>CRISP-DM</i> .....	9
Gambar II.3 Kerangka Kerja Teknik Data Mining Dalam CRM.....	14
Gambar II.4 WEKA GUI Chooser.....	20
Gambar III.1 Langkah Penelitian.....	27
Gambar III.2 <i>Import Dataset</i> Pada Lembar Kerja Rapid Miner .....	33
Gambar III.3 Dataset Berupa Excel Yang Akan <i>diimport</i> .....	33
Gambar III.4 Tampilan Sheet Data .....	34
Gambar III.5 Menentukan <i>Anotasi</i> dari Setiap <i>Tuple</i> .....	34
Gambar III.6 <i>Attribute Customer</i> .....	35
Gambar III.7 Pembentukan Model Pada <i>Rapid Miner</i> .....	35
Gambar IV.1 Nota Transaksi Penjualan.....	37
Gambar IV.2 <i>Attribute</i> Yang Dihimpun .....	37
Gambar IV.3 Hasil Proses <i>Clustering</i> .....	43
Gambar IV.4 Grafik Hasil <i>Clustering</i> .....	44
Gambar IV.5 Centroid atau Pusat Cluster Yang terbentuk .....	44
Gambar IV.6 Model Tree Untuk Penentuan Karakteristik Pelanggan.....	47
Gambar IV.7 <i>Rule Fitur</i> Untuk Penentuan Karakteristik Pelanggan.....	47
Gambar IV.8 Hasil Prediksi Data Training Fitur Penentuan Karakteristik Pelanggan .....	48
Gambar IV.9 Tampilan Halaman Depan .....	53
Gambar IV.10 Tampilan Form Input Pemetaan Pelanggan .....	54
Gambar IV.11 Tampilan Halaman Setelah Data Diinput .....	54
Gambar IV.12 Tampilan Halaman Output atau Hasil.....	55
Gambar IV.13 Tampilan Halaman Rekomendasi Strategy.....	55

## DAFTAR RUMUS

	Halaman
Rumus 3.1 <i>Euclidean Distance</i> .....	32
Rumus 4.1 <i>Mean</i> .....	38
Rumus 4.2 Standar Deviasi .....	39
Rumus 4.3 Jarak setiap data dengan pusat <i>cluster</i> .....	41

## DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Data Set Histori Transaksi Penjualan .....	66
Lampiran 2. Pembobotan .....	66
Lampiran 3. Data Hasil Pembobotan Siap Untuk Dimining .....	66
Lampiran 4. Penghitungan Cluster Manual .....	66
Lampiran 5. Hasil Output Clustering Menggunakan Rapid Miner .....	66

## ABSTRAK

Nama : Ina Maryani  
NIM : 14001675  
Program Studi : Magister Ilmu Komputer  
Jenjang : Strata Dua (S2)  
Konsentrasi : *Management Information System*  
Judul : “Pemanfaatan Data Mining Pada Industri Knalpot Untuk Menerapkan *Customer Relationship Management*”

Konsumen merupakan aset yang sangat penting bagi keberlangsungan perusahaan. Hal ini menjadi alasan mengapa perusahaan harus merencanakan dan menggunakan strategi yang cukup jelas dalam memperlakukan konsumen. Masalah yang dihadapi perusahaan adalah bagaimana menentukan konsumen potensial. Dengan menerapkan CRM (*Customer Relationship Management*), perusahaan dapat mengidentifikasi konsumen potensial dengan melakukan segmentasi konsumen. Tujuan dari proses segmentasi konsumen adalah untuk mengetahui profil konsumen dan menerapkan strategi pemasaran yang tepat sehingga mendatangkan keuntungan bagi pihak perusahaan. Penelitian ini membahas bagaimana proses data mining dari data konsumen di Industri Knalpot, yaitu salah satu industri yang menjual produk berupa knalpot motor dan mobil di daerah Purbalingga. Data yang diolah dalam penelitian ini berasal dari data histori transaksi yang dilakukan oleh pelanggan dan bertujuan untuk mencari konsumen potensial, Model yang digunakan adalah model RFM, yaitu model segmentasi berdasarkan atribut *Recency*, *Frequency* dan *Monetary*. Model RFM merupakan model yang umum digunakan dalam proses segmentasi konsumen. Proses data mining dimulai dengan melakukan proses *clustering* menggunakan *Algoritma K-means*. selanjutnya diadakan uji *validitas cluster* menggunakan *Davies Bouldin Index* untuk mengetahui cluster yang paling optimal, Selanjutnya untuk Penentuan Karakteristik *Cluster* dilakukan dengan *Algoritma Decision Tree*. Hasil dari penelitian ini adalah empat segmen pelanggan dan karakteristik dari masing-masing pelanggan serta rekomendasi strategi hubungan pelanggan.

Kata kunci: *Clustering, K-Means, RFM, Classification, Decision Tree*

## ABSTRACT

Name : Ina Maryani  
NIM : 14001675  
Study of Program : Magsiter Ilmu Komputer  
Levels : Strata Dua (S2)  
Concentration : MIS  
Titel : “Pemanfaatan Data Mining Pada Industri Knalpot Untuk Menerapkan *Customer Relationship Management*”

*Customers are the most important asset of industry. That is why industry should plan and employ a clear strategy for treating customers. A problem that appears in industry with a great quantity of customers is how to identify potential customers. Industry could identify their best customer through customer segmentation by applying Customer Relationship Management (CRM) concept. The objective of customer segmentation is to understand the customer's behavior and apply the appropriate marketing strategy in order to make a industry profitable. Industry could identify their best customer through customer segmentation by applying Customer Relationship Management (CRM) concept. The objective of customer segmentation is to understand the customer's behavior and apply the appropriate marketing strategy in order to make a profitable for the company. This research presents how the process of data mining of customer data in Industrial muffler, which is one of the industry that sells products such as motorcycle and automobile exhaust in the area Purbalingga. The processed data in this study is derived from historical data transactions made by customers and aims to find potential customers, model used is a model RFM, the segmentation model based on attributes Recency, Frequency and Monetary. RFM Model is a model that is commonly used in the process of customer segmentation. The data mining process begins with the process of clustering using K-means algorithm. subsequently held a validity test cluster using Davies Bouldin Index to determine the most optimal cluster, Cluster Characteristics Determining Furthermore, to do with the Decision Tree Algorithm. The results of this study are four customer segments and the characteristics of each customer and customer relationship strategy recommendations..*

*Keywords: Clustering, K-Means, RFM, Classification, Decision Tree*

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Dengan perubahan serba cepat akibat kemajuan teknologi informasi, organisasi dipaksa untuk siap menjalankan tren terbaru dalam mencapai keunggulan kompetitif. Konsep *data mining* merupakan bagian dari konsep teknologi informasi yang berkaitan dengan data dan informasi. Gupta dan Anggarwal (2012) menyatakan konsep *data mining* sebagai proses atau teknik pemodelan yang mempergunakan analisis dengan variasi data yang banyak untuk mendapatkan pola dan hubungan diantara variasi data tersebut. Kehadiran *data mining* dilatarbelakangi dengan adanya masalah *data explosion* atau ledakan data yang dialami oleh banyak organisasi yang telah mengumpulkan data sekian tahun lamanya (data pembelian, data penjualan, data nasabah, data transaksi, dan data data lainnya) (Fadli,2011).

Data pembelian, data penjualan, data nasabah, data transaksi, dan data Lain dalam perusahaan diperoleh oleh perusahaan dari kegiatan operasional yang dilakukan perusahaan. Perusahaan-perusahaan besar dalam berbagai industri, misalnya manufaktur, perbankan, dan *knalpot* dalam seharinya tentu akan menghasilkan banyak data dan di dalam data tersebut terkandung informasi yang penting bagi perusahaan. Adanya banyak data dalam perusahaan inilah yang disebut dengan *data explosion* atau ledakan data. Dapat dibayangkan besarnya ukuran data yang didapatkan jika nanti proses ini telah berjalan beberapa tahun dan sangat rugi bagi perusahaan jika dari data-data tersebut tidak didapatkan sebuah informasi. Pertanyaannya sekarang, apakah data tersebut akan dibiarkan menggunung, tidak berguna lalu dibuang, ataukah dapat me‘nambang’-nya untuk mencari ‘emas’, ‘berlian’ yaitu Informasi yang berguna untuk organisasi atau perusahaan (Fadli,2011).

Banyak diantara perusahaan-perusahaan yang memiliki banyak data tapi miskin informasi (Fadli, 2011). Rygielski *et al.* (2002) berpendapat bahwa *data*

*mining* bertujuan untuk mengekstrak atau mendapatkan informasi tersembunyi dari *database* yang besar. Informasi yang tersembunyi nantinya dapat dipergunakan oleh manajemen sebagai dasar dalam pengambilan keputusan yang terkait dengan bisnis perusahaan.

Pihak Manajemen Perusahaan harus mampu untuk mengenali Konsumen terbaiknya dan mempercayainya dengan meningkatkan pemahaman Perusahaan akan kebutuhan mereka sebagai individu sehingga dapat mempertahankan loyalitasnya terhadap Perusahaan. Perusahaan dapat melakukan identifikasi konsumen dengan melakukan segmentasi konsumen. Tujuan dari proses segmentasi konsumen adalah untuk mengetahui Perilaku dan Profil Konsumen dan menerapkan strategi pemasaran yang tepat sehingga mendatangkan keuntungan bagi pihak perusahaan.

Kebanyakan Perusahaan akan melakukan pelayanan yang sama kepada semua Pelanggan. Padahal belum tentu semua Pelanggan memerlukan atau menginginkan produk yang sama. Ini sangat terkait dengan *customer relationship management* yang dilakukan oleh pihak Perusahaan kepada Pelanggan. Jika pihak Perusahaan tidak mengetahui kebutuhan dan keinginan dari Pelanggan, tentu saja loyalitas pelanggan kepada perusahaan tersebut akan semakin menurun. Konsep *customer relationship management* merupakan konsep bisnis yang berpusat pada Pelanggan sehingga konsep ini sangat penting agar hubungan antara perusahaan dan pelanggan dapat terjaga dengan baik. Mengingat begitu pentingnya *data mining* terhadap operasional suatu organisasi, khususnya untuk industri-industri, maka sudah seharusnya Manajemen menerapkan konsep *data mining* dan melihat manfaatnya untuk *customer relationship management* Perusahaan kepada Pelanggan. Tulisan ini akan membahas permasalahan yang menyangkut penerapan konsep *data mining* dalam Perusahaan Industri Knalpot dan manfaatnya untuk *Customer Relationship Management* Perusahaan. Riset ini memiliki kontribusi dalam hal memahami pentingnya *data mining* bagi bisnis Industri Knalpot dan memberikan pemahaman mengenai hubungan antara *data mining* dengan *Customer Relationship Management* Perusahaan. Jika *data mining* dan *Customer Relationship Management* Perusahaan dilakukan dengan baik

dalam menjalankan proses bisnisnya, diharapkan perusahaan akan semakin baik kedepannya.

Berbagai penelitian telah dilakukan untuk mengelompokkan Pelanggan yang dikenal dengan istilah *Customer Segmentation* atau *Customer Profiling*. Konsep ini pada dasarnya berusaha mengelompokkan Pelanggan berdasarkan kesamaan karakteristik tertentu yang dimilikinya. Model yang umum digunakan dalam mengelompokkan Pelanggan adalah model RFM (*Recency, Frequency, Monetary*), yaitu melakukan pengelompokan berdasarkan waktu kunjungan terakhir pelanggan, frekuensi kunjungan, dan pendapatan yang didapat oleh perusahaan (shi Zang, 2015; Indika, dkk., 2015). Metode yang sering digunakan dalam pengelompokan pelanggan yaitu dengan teknik *clustering*. Pada penelitian ini dilakukan segmentasi dengan menggunakan teknik *clustering* agar proses dan hasil pengelompokannya lebih baik. algoritma yang digunakan dalam pengelompokan pelanggan Industri Knalpot dengan menggunakan *K-Means Clustering*. Algoritma ini memiliki keunggulan dari kecepatan & kestabilan pada hasil yang diperoleh (Mingoti dkk.,2006; Tsiptsis dkk.,2009). Pada beberapa penelitian terdahulu (Kohavi dkk.,2004; Aggelis dkk.,2005; Niyagas dkk.,2006; Golsefid dkk.,2007; Trappey dkk.,2009; Yanuarita dkk.,2011) telah dilakukan segmentasi dengan menggunakan tiga buah karakteristik, yaitu *Recency, Frequency, Monetary*.

Berdasarkan latar belakang di atas maka peneliti tertarik untuk melakukan penelitian dengan judul **“PEMANFAATAN DATA MINING PADA INDUSTRI KNALPOT UNTUK MENERAPKAN CUSTOMER RELATIONSHIP MANAGEMENT”**

## **1.2 Masalah Penelitian**

### **1.2.1 Identifikasi Masalah**

Dari latar belakang tersebut diatas ditemukan beberapa permasalahan antara lain:

- a. Pada Industri Knalpot Doctor Speed Purbalingga terdapat banyak himpunan data yang belum dimanfaatkan

- b. Pada Industri Knalpot Doctor Speed Purbalingga belum ada Pemetaan Pelanggan
- c. Karakteristik Pelanggan yang berbeda-beda menyebabkan Perusahaan kesulitan menjalankan strategi hubungan Pelanggan dan melakukan pemasaran produk.

### 1.2.2 Batasan Masalah

Dalam rangka proses penelitian ini, penulis membatasi masalah dengan melaksanakan studi kasus di perusahaan Industri Knalpot dan Mobil pada Doctor Speed Purbalingga dengan identifikasi masalah tersebut di atas.

Dengan melaksanakan penelitian ini diharapkan perusahaan dapat memanfaatkan data mining untuk mengetahui segmentasi konsumennya dan pentingnya *data mining* bagi bisnis Industri Knalpot serta memberikan pemahaman mengenai hubungan antara *data mining* dengan *customer relationship management* perusahaan

### 1.2.3 Rumusan Masalah

Berdasarkan latar belakang dan batasan masalah tersebut, masalah-masalah yang dibahas dapat dirumuskan sebagai berikut:

- a. Bagaimana membuat suatu pemetaan atau Profil dari Pelanggan yang dimiliki oleh perusahaan, sehingga dapat diketahui segmen dari para pelanggan tersebut berdasarkan karakteristik yang dimilikinya.
- b. Bagaimana mengklasifikasikan Pelanggan dan mengetahui Karakteristik dari masing-masing Pelanggan
- c. Bagaimana menerapkan *Customer Relationship Management* bagi Perusahaan agar tepat dalam melakukan strategi pemasaran untuk menjaga loyalitas pelanggan

### **1.3 Tujuan dan Manfaat**

#### **1.3.1 Tujuan**

Tujuan dari penelitian ini adalah :

- a. Membuat *Customer Profile* untuk mengelompokkan/memetakan pelanggan berdasarkan analisis RFM (*Recency, Frequency dan Monetary*)
- b. Mengklasifikasikan dan Mengetahui Karakteristik pelanggan serta melakukan Evaluasi Loyalitas Pelanggan.
- c. Mengetahui bagaimana menerapkan *Customer Relationship Management* dalam melakukan strategi pemasaran yang tepat untuk menjaga loyalitas pelanggan.

#### **1.3.2 Manfaat**

Manfaat dari penelitian ini adalah :

1. Membantu manajemen perusahaan untuk dapat mengetahui kategori dari para pelanggan berdasarkan hasil *customer profiling*
2. Mengetahui Karakteristik Pelanggan
3. Memberikan informasi yang dapat digunakan untuk menentukan strategi yang berkaitan dengan peningkatan kualitas pelayanan. Beberapa hal yang dapat dilakukan diantaranya adalah:
  - a. Prioritas pelayanan kepada pelanggan yang termasuk dalam kategori *potential customer*.
  - b. Program diskon atau potongan harga layanan bagi pelanggan yang bertransaksi dengan intensitas cukup tinggi atau rutin.
  - c. Sebagai bahan acuan bagi perusahaan untuk mengembangkan pelayanan atau melihat celah peluang usaha.

### **1.4 Sistematika Penulisan**

Sistematika penulisan tesis ini terdiri dari 5 (lima) bab, dimana tiap bab terdiri dari beberapa sub bab sebagai berikut :

**BAB 1 PENDAHULUAN**

Bab ini membahas tentang Latar Belakang Penulisan, Identifikasi Masalah, Batasan Masalah, Rumusan Masalah, Tujuan Manfaat dan Sistematika Penulisan.

**BAB 2 LANDASAN/KERANGKA PEMIKIRAN**

Pada Bab ini dibahas teori yang melandasi penelitian, dalam bab ini juga diuraikan Tinjauan Pustaka, Tinjauan Studi, Tinjauan Organisasi/Obyek Penelitian, Kerangka Pemikiran.

**BAB 3 METODOLOGI PENELITIAN**

Membahas mengenai Metodologi/langkah penelitian, pengumpulan data, metode analisis data, kerangka pendekatan berikut jadwal penelitian yang digunakan untuk melakukan segmentasi konsumen

**BAB 4 HASIL PENELITIAN DAN PEMBAHASAN**

Menampilkan hasil dari pengelompokkan data dan analisa hasil pengolahan data. Pada bab ini juga ditampilkan hasil *clustering* menggunakan *Algoritma K-Means* berdasarkan analisis *variable Recency, Frequency, Monetary* dan Pembobotan Kriteria, Klasifikasi Pelanggan menggunakan *Algoritma Desicion Tree J48* dan hasil ditampilkan dalam bentuk Tabel Jenis dan Karakteristik Pelanggan.

**BAB 5 PENUTUP**

Membahas kesimpulan dari penelitian yang dilakukan dan saran untuk penelitian selanjutnya.

## BAB II LANDASAN TEORI

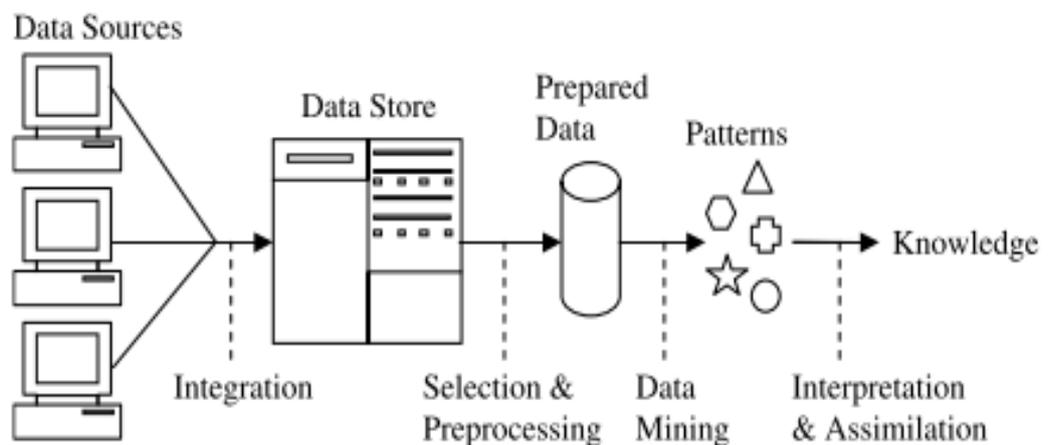
### 2.1 Tinjauan Pustaka

Untuk menghindari kesalahan penafsiran dalam mengartikan istilah-istilah yang digunakan untuk menata konsep di dalam penelitian ini, maka istilah-istilah yang terdapat dalam penelitian ini akan dijelaskan sebagai berikut:

#### 2.1.1. Data Mining

*Data mining* didefinisikan sebagai proses menemukan dan menggambarkan pola struktural dalam data sebagai alat untuk membantu menjelaskan data dan membuat prediksi dari data tersebut (Witten, Frank, & Hall, 2011). Data mining merupakan bagian dari *Knowledge Discovery Data* (KDD) yang merupakan proses ekstraksi informasi yang berguna, tidak diketahui sebelumnya, dan tersembunyi dari data (Bramer, 2007) dan juga mengembangkan model yang digunakan untuk memahami fenomena dari analisis data dan prediksi (Maimon & Rokach, 2010). Data mining merujuk pada ekstraksi pengetahuan dari jumlah data yang besar (Han & Kamber, 2006) yang tersimpan dalam komputer, proses pada data mining lebih banyak berdasarkan pada teknologi informasi (Olson & Delen, 2008).

Dibawah ini digambarkan proses ekstraksi data menjadi informasi:



Sumber: Bramer (2007)

Gambar II. 1 Proses Knowledge Discovery From Dat

Dari gambar II.1 menunjukkan bahwa data bisa berasal dari mana saja, data-data tersebut kemudian diintegrasikan kedalam sebuah *data store*, dari *data store* data kemudian diseleksi dan diproses sehingga menghasilkan pola-pola dan pengetahuan yang berguna.

Secara garis besar proses KDD dapat jelaskan sebagai berikut (Kusrini & Luthfi, 2009):

1. *Data Selection*

Pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. *Pre-processing/Cleaning*

Proses *cleaning* antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data. Pada proses ini dilakukan juga proses *enrichment*, yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD.

3. *Transformation*

*Coding* adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*.

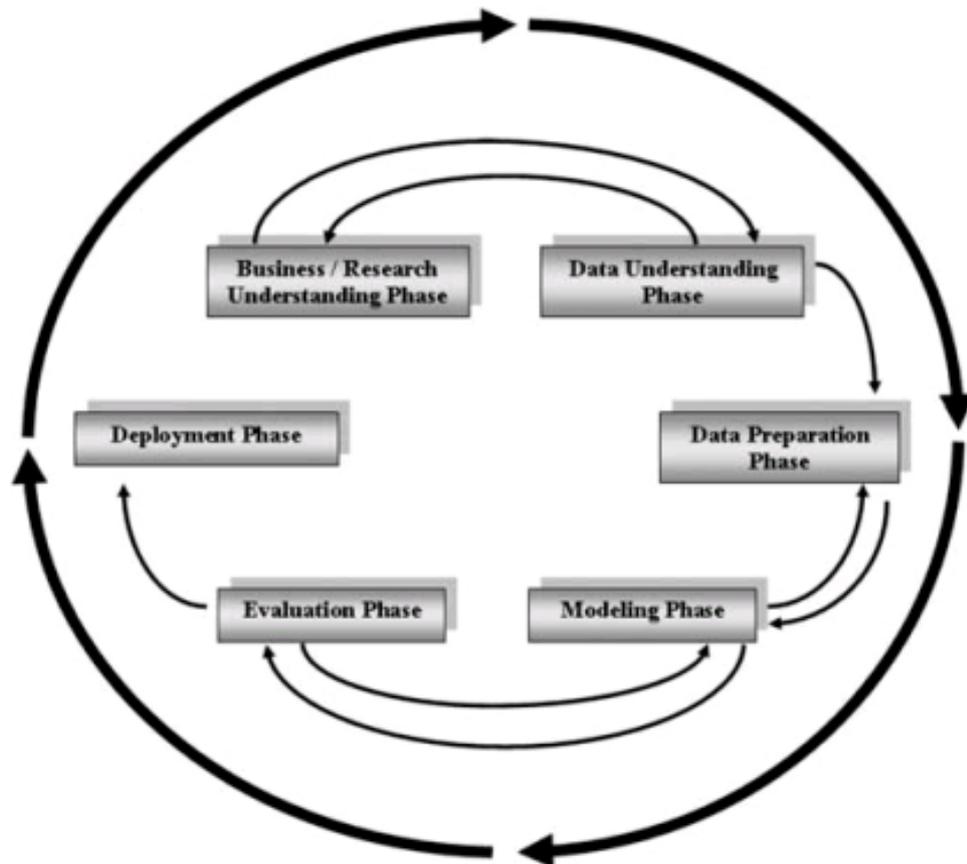
4. *Data Mining*

*Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.

5. *Interpretation/Evaluation*

Pola informasi yang dihasilkan dari proses *data mining* diterjemahkan menjadi bentuk yang lebih mudah dimengerti oleh pihak yang berkepentingan.

Metode data mining dan model melanjutkan cakupan data mining sebagai suatu proses, standar proses yang digunakan adalah kerangka kerja CRISP-DM. Dalam CRISP-DM data mining dilihat sebagai proses keseluruhan dari komunikasi masalah bisnis hingga penerapan model (Larose, 2006).



Gambar II.2 Proses CRISP-DM

Sumber: Larose (2006)

Pada gambar 2.2 Dapat dilihat bahwa proses CRISP-DM terdiri dari enam tahap, yaitu (Larose, 2006):

1. Tahap *Business Understanding*  
Tahap *business understanding* disebut juga sebagai tahap *research understanding*, dalam tahap ini ditentukan tujuan dan requirement secara detail pada keseluruhan penelitian, merumuskan masalah data mining, dan menyiapkan strategi awal untuk mencapai tujuan yang telah ditentukan.
2. Tahap *Data Understanding*  
Pada tahap ini mulai dilakukan proses pengumpulan data, menganalisis data, mengevaluasi kualitas data, dan memilih subset yang mungkin mengandung pola yang ditindaklanjuti.
3. Tahap *Data Preparation*  
Pada tahap ini data akhir yang akan digunakan pada tahap berikutnya mulai disiapkan, memilih kasus dan variabel yang sesuai dengan analisis

yang akan dilakukan, melakukan transformasi pada variabel tertentu jika diperlukan, dan membersihkan data mentah sehingga siap untuk digunakan sebagai alat pemodelan.

4. Tahap *Modelling*

Pada tahap ini memilih dan menerapkan teknik pemodelan yang tepat, mengatur kalibrasi model untuk mendapatkan hasil yang optimal. Pada tahap ini juga dapat diterapkan beberapa teknik yang berbeda untuk permasalahan data mining yang sama, dan jika diperlukan proses dapat kembali ke tahap *data preparation* untuk menjadikan data kedalam bentuk yang sesuai dengan spesifikasi kebutuhan data mining tertentu.

5. Tahap *Evaluation*

Setelah tahap *modelling* selesai dilakukan, model tersebut harus dievaluasi untuk melihat kualitas dan efektivitas sebelum disebarkan untuk digunakan. Pada tahap ini juga ditentukan apakah model dapat mencapai tujuan yang telah ditetapkan pada tahap pertama, apakah terdapat permasalahan penting dari penelitian yang tidak tertangani dengan baik, dan mengambil keputusan berkaitan dengan penggunaan hasil dari data mining.

6. Tahap *Deployment*

Terbentuknya model tidak menandakan telah terselesaikannya proyek, perlu untuk menggunakan model yang dibuat sesuai dengan tujuan bisnis. Contoh sederhana dari tahap ini adalah pembuatan laporan, sedangkan contoh yang lebih kompleksnya yaitu dengan menerapkan proses data mining pada departemen lain secara paralel.

Pada umumnya tugas data mining dapat diklasifikasikan menjadi dua kategori, yaitu deskriptif dan prediktif (Han & Kamber, 2006). Tugas mining deskriptif mengungkapkan pola dalam data dan mudah diinterpretasikan oleh pengguna seperti *clustering*, *association rule*, dll. Tugas mining prediktif menggunakan beberapa variabel yang ada untuk memprediksi nilai dari variabel lain seperti *classification*, *regression*, dll (Gorunescu, 2011).

## 2.2 *Customer Relationship Management (CRM)*

Konsumen merupakan aset yang sangat penting bagi suatu perusahaan. Tidak akan ada prospek bisnis tanpa adanya hubungan antara perusahaan dengan konsumen yang bersifat loyal. Hal ini adalah alasan mengapa perusahaan harus merencanakan dan menggunakan strategi yang cukup jelas dalam memperlakukan konsumen. *Customer Relationship Management (CRM)* telah berkembang dalam beberapa dekade belakangan ini untuk merefleksikan peranan utama dari konsumen untuk pengaturan strategi perusahaan. CRM meliputi seluruh ukuran untuk memahami konsumen dan proses untuk mengeksploitasi pengetahuan ini untuk merancang dan mengimplementasikannya pada kegiatan marketing, produksi, dan rantai *supply* dari pemasok (*supplier*). Berikut ini akan didefinisikan beberapa pengertian CRM yang diambil dari berbagai literatur, antara lain (Tama, 2009) :

1. Dari sisi yang berkaitan dengan teknologi informasi, CRM adalah sebuah strategi untuk mengoptimalkan *customer lifetime value* dengan cara mengetahui lebih banyak mengenai informasi konsumen dan berinteraksi dengan konsumen secara intensif (Todman, 2001).
2. Dari sisi komunikasi dan manajemen, CRM didefinisikan sebagai sebuah pendekatan perusahaan untuk memahami dan mempengaruhi perilaku konsumen melalui komunikasi yang intensif dalam rangka meningkatkan akuisisi konsumen, mempertahankan konsumen, dan loyalitas konsumen (Swift, 2000).
3. Definisi CRM jika dilihat dari segi bisnis dapat diartikan sebagai sebuah strategi bisnis untuk memahami, mengantisipasi dan mengelola kebutuhan konsumen yang potensial dalam suatu organisasi pada saat sekarang dan yang akan datang (Brown, 2000).

Dari ketiga definisi di atas dapat ditarik kesimpulan mengenai definisi CRM yaitu sebuah pendekatan yang komprehensif yang mengintegrasikan setiap bisnis proses yang berhubungan langsung dengan konsumen, yaitu penjualan, pemasaran dan layanan konsumen melalui integrasi filosofi, teknologi, dan juga proses (Tama, 2009). Dengan kata lain, CRM dipandang bukanlah sebagai sebuah produk ataupun sebuah layanan, tetapi sebuah filosofi bisnis yang bertujuan

memaksimalkan nilai konsumen dalam jangka panjang (*customer lifetime value*). Dalam pandangan yang berbeda diungkapkan bahwa CRM adalah suatu strategi untuk mempelajari kebutuhan dan perilaku pelanggan untuk mengembangkan hubungan yang lebih baik dengan mereka (Laudon dkk., 2009).

### 2.3 Teknologi CRM

Sebuah definisi standar mengenai komponen teknologi CRM diberikan oleh META Group di dalam “*The Customer Relationship Management Ecosystem*”. Kategorisasi ini bertujuan untuk lebih memahami bagaimana suatu strategi CRM dan teknologinya mampu menyediakan seluruh arsitektur secara menyeluruh yang terfokus pada layanan konsumen. Ada tiga tipe utama dari teknologi CRM yaitu (Tama, 2009) :

1. *Operational CRM*, yang dikenal sebagai CRM “*front office*”. *Operational CRM* merupakan area dimana terjadi hubungan dengan konsumen secara langsung. Otomasi dari proses bisnis yang terintegrasi yang melibatkan *front office customer touch points*, seperti penjualan, pemasaran, dan layanan konsumen, termasuk integrasi antara *front office* dan *back office*, seperti manajemen pesanan, dan otomasi pemasaran. Dengan perkembangan teknologi informasi dan komunikasi yang pesat, *operational CRM* kemudian juga dapat dibedakan menjadi dua bagian yaitu e-CRM dan m CRM.
2. *Collaborative CRM*, merupakan bentuk perluasan dari CRM tradisional. *Collaborative CRM* berkaitan dengan manajemen hubungan antara *stakeholder* eksternal, seperti *supplier*, distributor dan *reseller*. Aplikasi utama dari *collaborative CRM* adalah *enterprise portal* yang berbasis pada infrastruktur ekstranet dan perangkat lunak manajemen hubungan mitra (*partner relationship management software*) yang memberikan akses kepada konsumen, *reseller* dan mitra bisnis melalui internet.
3. *Analytical CRM* yang dikenal sebagai CRM “*back Office*” atau “*strategic*” CRM. *Analytical CRM* biasanya berhubungan dengan penggunaan data secara efektif, efisien, strategis sehingga memungkinkan pengambilan keputusan yang tepat bagi pihak manajemen. Pengambilan keputusan

dilaksanakan setelah melalui proses analisis, pemodelan dan evaluasi terhadap data yang tersimpan di basis data untuk menghasilkan sebuah hubungan yang saling menguntungkan antara perusahaan dan konsumennya. Contoh dari Analytical CRM adalah business intelligence seperti data warehouse, OLAP dan data mining.

Hal yang penting dalam memahami perilaku konsumen tidak hanya memahami siapa mereka (dalam hal ini *Customer profiling* atau *segmentation*) tetapi juga perilaku mereka dan pola-pola (*pattern*) yang mereka ikuti. Pemahaman perilaku pelanggan harus ditetapkan sebagai suatu proses yang dinamis dan berkelanjutan (Xu dan Walton, 2005)

#### **2.4 Data Mining dalam Kerangka Kerja CRM.**

Menurut Swift (2001), Parvatiyar dan Sheth (2001), serta Kracklauer, Mills, dan Seifert (2004), CRM memiliki empat dimensi, yaitu :

1. Identifikasi konsumen (*customer identification*).
2. Membangun daya tarik terhadap konsumen (*customer attraction*).
3. Mempertahankan konsumen yang ada (*customer retention*).
4. Pengembangan konsumen (*customer development*).

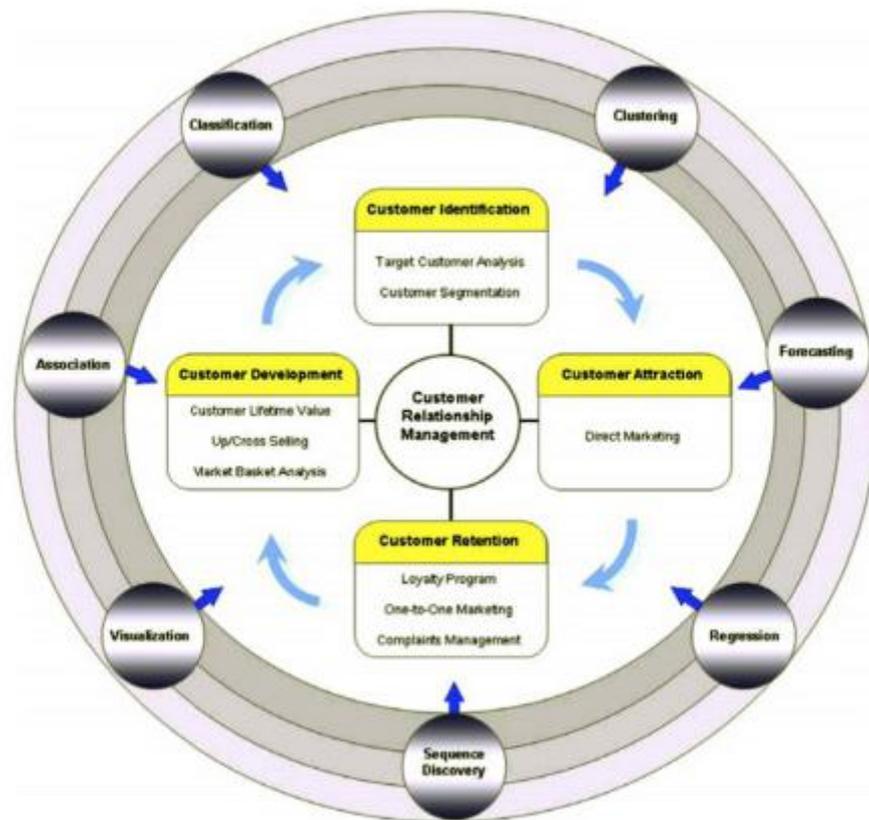
Keempat dimensi dari CRM tersebut dapat dilihat sebagai siklus tertutup dalam *Customer Management System*. Setiap dimensi memiliki tujuan umum yang sama yaitu untuk membangun pemahaman yang lebih mendalam dari konsumen dan meningkatkan nilai konsumen (*customer value*) dari perusahaan. Dalam kaitannya dengan pencapaian tujuan tersebut, teknik *data mining* dapat digunakan untuk menemukan karakteristik dan perilaku konsumen yang tersembunyi dalam *database* yang berukuran besar. Untuk menemukan pola karakteristik dan perilaku konsumen, dapat digunakan metode pemodelan data dalam teknik *data mining*.

Beberapa metode pemodelan data yang dapat digunakan, yaitu :

1. *Association*
2. *Classification*
3. *Clustering*
4. *Forecasting*
5. *Regression*

6. *Sequence discovery*
7. *Visualization*

Kerangka kerja teknik *data mining* dalam keempat dimensi CRM dijelaskan melalui Gambar 2.3.



Gambar II.3 Kerangka Kerja Teknik Data Mining Dalam CRM (Ngai dkk, 2008)

## 2.6 Segmentasi Pelanggan

Segmentasi pelanggan adalah proses membagi pelanggan yang berbeda ke kelompok *homogen* untuk membangun strategi pemasaran berdasarkan karakteristik pelanggan. Banyak tipe segmentasi yang berbeda berdasarkan kriteria fisik atau atribut yang digunakan untuk melakukan segmentasi. Sifat dari segmentasi pelanggan dikelompokkan berdasarkan sifat dan karakteristik yang digunakan. Sifat dari segmentasi dapat dibuat dengan aturan bisnis. Pendekatan yang melekat ini tidak menguntungkan akan lebih efisien jika ditangani oleh beberapa segmentasi dasar dan objektivitas berdasarkan persepsi personal dari ahli

bisnis. Algoritma *clustering* dapat menganalisa sifat data, mengidentifikasi kelompok pelanggan secara alami dan menawarkan solusi yang diterima dari hasil pengamatan pola data. Pelayanan Model Data Mining dibangun dengan baik dapat menemukan kelompok dengan profil dan karakteristik yang jelas dan memperkaya skema segmentasi dengan pengertian dan nilai bisnis. (Tsiptsis Antonios Chorianopoulos, 2009,h.30)

Untuk dapat menjadi pelanggan yang loyal, seorang pelanggan harus memulai beberapa tahapan. Proses ini berlangsung lama dengan penekanan dan perhatian yang berbeda untuk masing-masing tahap karena setiap tahap mempunyai kebutuhan yang berbeda. Dengan memperhatikan masing-masing tahapan dan memenuhi kebutuhan dalam setiap tahap tersebut, perusahaan memiliki peluang yang lebih besar untuk membentuk calon pembeli menjadi pelanggan loyal dan klien perusahaan. Hill (1996 ;332) menjelaskan bahwa tingkatan loyal terbagi atas 6 tingkat, yaitu :

1. *Suspect*

Bagian ini termasuk semua pembeli produk atau jasa dalam pemasaran, jadi *suspects* adalah menyadari akan produk atau jasa perusahaan atau tidak mempunyai kecenderungan terhadap pembelian.

2. *Prospects*

*Prospects* adalah pelanggan *potensial* yang mempunyai daya tarik terhadap perusahaan tetapi belum mengambil langkah untuk melakukan bisnis dengan perusahaan.

3. *Customers*

Suatu tipe pembelian produk (walaupun dalam kategori ini termasuk beberapa pembelian ulang) yang tidak memiliki loyalitas pada perusahaan.

4. *Clients*

Pembelian ulang yang menunjukkan *loyalitas* pada perusahaan tetapi lebih memiliki dorongan pasif daripada aktif terhadap perusahaan.

5. *Advocates*

Client yang memberikan dorongan yang positif pada perusahaan dengan merekomendasikannya kepada orang lain.

## 6. *Partners*

*Partners* adalah hubungan yang sangat erat antara konsumen dengan supplier yang keduanya saling memperlihatkan keuntungan.

Hill (1996) membagi tahapan loyalitas pelanggan menjadi 6 tahap dari tingkat *suspects* hingga tahap *partner*, untuk lebih jelasnya lagi di bawah ini digambarkan mengenai piramida tentang loyalitas pelanggan.

### *The Profit Generator System*

#### 1. *Suspects*

Adalah semua orang yang mungkin akan membeli produk atau jasa perusahaan. Kita menyebutnya sebagai *suspects* karena yakin bahwa mereka akan membeli tetapi belum tahu apapun mengenai perusahaan dan barang/jasa yang ditawarkan.

#### 2. *Prospects*

Adalah orang-orang yang memiliki kebutuhan akan produk atau jasa tertentu dan mempunyai kemampuan untuk membelinya. Para *prospects* ini meskipun mereka belum melakukan pembelian, mereka telah menyatakan keberadaan perusahaan, barang dan jasa yang ditawarkan karena seseorang telah merekomendasikan barang/jasa tersebut padanya.

#### 3. *Disqualified Prospect*

Yaitu *prospect* yang telah mengetahui keberadaan barang/ jasa tertentu, tetapi tidak mempunyai kebutuhan akan barang/ jasa tersebut atau tidak mempunyai kemampuan untuk membeli barang/ jasa tersebut.

#### 4. *First Time Customer*

Yaitu pelanggan yang membeli untuk yang pertama kalinya, mereka masih menjadi pelanggan yang baru.

#### 5. *Repeat Customers*

Yaitu pelanggan yang telah melakukan pembelian suatu produk sebanyak 2 kali atau lebih. Mereka adalah yang melakukan pembelian atas produk yang sama sebanyak 2 kali atau membeli dua macam produk yang berbeda dalam 2 kesempatan yang berbeda pula.

#### 6. *Clients*

Membeli semua barang/jasa yang ditawarkan yang mereka butuhkan, mereka

membeli secara teratur, hubungan dengan jenis pelanggan ini sudah kuat dan berlangsung lama yang membuat mereka tidak terpengaruh oleh tarikan pesaing produk lain.

#### 7. *Advocates*

Seperti layaknya klien, *advocates* membeli seluruh barang/jasa yang ditawarkan yang ia butuhkan, serta melakukan pembelian secara teratur sebagai tambahan mereka mendorong teman-teman mereka yang lain agar membeli barang/ jasa tersebut. Ia membicarakan tentang barang/jasa tersebut, melakukan pemasaran untuk perusahaan tersebut dan membawa pelanggan untuk perusahaan tersebut.

## 2.7 Clustering

*Clustering* merupakan teknik pengolahan data mining yang termasuk pada kategori *unsupervised learning* yaitu pembelajaran tanpa guru atau tanpa menggunakan label kelas seperti pada klasifikasi. *Clustering* pada algoritma K-Means mencari nilai kemiripan antara data dengan teknik penghitungan jarak salah satunya menggunakan *euclidean distance*,

Analisa cluster yaitu menemukan kumpulan objek hingga objek-objek dalam satu kelompok sama (atau punya hubungan) dengan yang lain dan berbeda (atau tidak berhubungan) dengan objek-objek dalam kelompok lain. Tujuan dari analisa cluster adalah meminimalkan jarak di dalam cluster dan memaksimalkan jarak antar cluster.

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. Jika diberikan himpunan data yang berjumlah terhingga, yaitu  $X$ , maka permasalahan *clustering* dalam  $X$  adalah mencari beberapa pusat *cluster* yang dapat memberikan ciri kepada masing-masing *cluster* dalam  $X$

Algoritma K-means merupakan salah satu algoritma dengan partitional, karena K-Means didasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai *centroid* awalnya (Madhulatha, 2012). Algoritma K-means menggunakan

proses secara berulang-ulang untuk mendapatkan basis data *cluster*. Dibutuhkan jumlah *cluster* awal yang diinginkan sebagai masukan dan menghasilkan titik *centroid* akhir sebagai output. Metode K-means akan memilih pola k sebagai titik awal *centroid* secara acak atau random. Jumlah iterasi untuk mencapai *cluster centroid* akan dipengaruhi oleh calon *cluster centroid* awal secara random. Sehingga didapat cara dalam pengembangan algoritma dengan menentukan *centroid cluster* yang dilihat dari kepadatan data awal yang tinggi agar mendapatkan kinerja yang lebih tinggi (HUNG et al., 2005, Saranya & Punithavalli, 2011, Eltibi & Ashour, 2011).

Dalam penyelesaiannya, algoritma K-Means akan menghasilkan titik *centroid* yang dijadikan tujuan dari algoritma K-Means. Setelah iterasi K-Means berhenti, setiap objek dalam dataset menjadi anggota dari suatu *cluster*. Nilai *cluster* ditentukan dengan mencari seluruh objek untuk menemukan *cluster* dengan jarak terdekat ke objek. Algoritma K-means akan mengelompokkan item data dalam suatu dataset ke suatu *cluster* berdasarkan jarak terdekat (Bangoria et al., 2013). Nilai *centroid* awal yang dipilih secara acak yang menjadi titik pusat awal, akan dihitung jarak dengan semua data menggunakan rumus *Euclidean Distance*. Data yang memiliki jarak pendek terhadap *centroid* akan membuat sebuah *cluster*. Proses ini berkelanjutan sampai tidak terjadi perubahan pada setiap kelompok (Agrawal & Gupta, 2013, Chaturved & Rajavat, 2013, Bhatia & Khurana, 2013).

### **2.7.1. Keuntungan Algoritma K-Means**

Algoritma K-Means juga memiliki keuntungan yaitu :

1. Dalam implementasi menyelesaikan masalah, algoritma K-Means sangat *simple* serta *fleksibel*. Artinya perhitungan komputasinya tidak terlalu rumit dan algoritma ini dapat diimplementasikan pada segala bidang.
2. Algoritma K-Means sangat mudah untuk dipahami, terutama dalam implementasi data yang sangat besar serta dapat mengurangi kompleksitas data yang dimiliki (Bangoria et al., 2013).

## **2.8. Klasifikasi (*Classification*)**

Klasifikasi adalah proses untuk menentukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya diketahui. Salah satu metode klasifikasi adalah *decision tree*, dengan cara ini akan mudah mengelompokkan obyek ke dalam beberapa kelompok. Dalam *decision tree* setiap atribut dinyatakan di simpul. Jawaban dari atribut ini dinyatakan dalam cabang sampai akhirnya ditemukan kategori atau jenis dari suatu objek di simpul terakhir. Proses klasifikasi biasanya dibagi dua fase yaitu *learning* atau *test*. Pada fase *learning*, data yang sudah diketahui kelas datanya dijadikan sebagai model perkiraan. Pada fase *test*, model yang sudah terbentuk diuji dengan data lainnya untuk mengetahui akurasi dari model tersebut. bila akurasinya mencukupi model ini dapat dipakai untuk prediksi kelas data yang belum diketahui.

### **2.8.1. Metode *Decision Tree***

Konsep klasifikasi dengan pengawasan (*supervised clasification*) bertujuan untuk membangun sebuah model dari data yang telah diketahui. Model atau fungsi yang dibangun tersebut kemudian dapat digunakan untuk memetakan data dalam suatu basis data ke suatu atribut target. Jika ada data baru maka model dapat digunakan untuk memperkirakan kelas dari data baru tersebut. Banyak algoritma yang dapat digunakan untuk melakukan klasifikasi.

### **2.8.2. Implementasi *Decision Tree* dalam Weka**

WEKA menurut Witten (2000) mudah digunakan dan diterapkan pada beberapa tingkatan yang berbeda. Tersedia implementasi algoritma-algoritma pembelajaran state-of-the-art yang dapat diterapkan pada dataset dari command line. WEKA mengandung tools untuk pre-processing data, klasifikasi, regresi, clustering, aturan asosiasi, dan visualisasi. User dapat melakukan preprocess pada data, memasukannya dalam sebuah skema pembelajaran, dan menganalisa classifier yang dihasilkan dan performansinya. Semua itu tanpa menulis kode program sama sekali. Contoh penggunaan WEKA adalah dengan menerapkan sebuah metode pembelajaran ke dataset dan menganalisa hasilnya untuk memperoleh informasi tentang data, atau menerapkan beberapa metode dan

membandingkan performansinya untuk dipilih. Weka menyediakan implementasi dari algoritma C45 yang dapat dengan mudah diterapkan dalam dataset yaitu J48.

Salah satu manfaat dari WEKA adalah dapat digunakan untuk menerapkan metode machine learning dimana dari output dapat digunakan untuk mempelajari lebih lanjut tentang data. Penggunaan WEKA yang lain adalah digunakan sebagai model pembelajaran untuk menghasilkan prediksi instances.

Fungsi berikutnya adalah menerapkan beberapa metode klasifikasi yang berbeda untuk tujuan membandingkan kinerja antar metode klasifikasi. Dalam modul Weka semua disediakan sebagai fungsi classifiers (Whitten, 2000).



Gambar II.4 WEKA GUI Chooser

WEKA GUI Chooser adalah tampilan utama yang akan dilihat user pada saat pertama kali membuka perangkat lunak WEKA. Tampilan utama tersebut memberikan 4 pilihan GUI WEKA, yaitu *Explorer*, *Experimenter*, *Knowledge Flow* dan *Simple CLI*.

## 2.9. Tinjauan Studi

Penelitian tentang *data mining* untuk CRM, khususnya untuk proses segmentasi konsumen, serta metode RFM telah beberapa kali dilakukan sebelumnya.

**Tabel II.1 Tinjauan Penelitian Sebelumnya dari *international journal***

No	Penelitian	Judul	Metode	Deskripsi
1.	Vahid Golmah (2014)	A case study of applying SOM in market	SOM, CRM	Mencari segmentasi pelanggan asuransi mobil
2.	Zhang shi, zheng wen, Jin Xia (2015)	An intelligent recommendation system based on customer segmentation	Fuzzy C-Means, RFM	
3.	Mrs Pradnya muley & Dr. Annirudha Joshi (2015)	Application of data mining techniques for customer segmentation in real time busniness intelligence		Predictive data mining, segmentasi pelanggan, industry ritel online
4.	Prof. Archana Raje, Dr. R.K Srivastava (2014)	The Connection between data mining and segmentation in marketing area	Clustering, targetting, positioning	Diperlukan untuk memahami koneksi antara data mining dan segmentasi dalam pemasaran
5.	Dr. Sankar Rajagobal (2011)	Customer data clustering using data mining techniques	Clustering, I-Miner	
6.	Tannane Parsa Kord Aslabi, Rezatoval (2015)	A review of different data mining techniques in customer segmentation	CRM	Menjelaskan CRM dan data mining beserta metode-metode yang biasa digunakan untuk segmentasi pelanggan
7.	V.L.Migoesis, AS. Camanho, Joao Falcao e Cunha (2012)	Customer data mining for life style segmentation	Algoritma clustering	Mengusulkan sebuah metode untuk segmentasi pasar di perusahaan ritel di eropa berdasarkan gaya hidup pelanggan yang informasinya diambil dari set keranjang belanja yang ada di database

No	Penelitian	Judul	Metode	Deskripsi
8.	Hasita Indika Aruwawadu, R.M. Kapila Tharanga Rathnayaka, S.K. illangarithne (2015)	Mining profitability of telecommunication customer using K-Means clustering	K-Means clustering, RFM Model dan CRM	Mencari segmentasi pelanggan yang mendatangkan keuntungan menggunakan K-Means Clustering, RFM Model dan CRM
9.	Hasan Ziafat, Majid shakeri (2014)	Using data mining techniques in customer segmentation	Clustering, CRM, Segmentasi konsumen	

**Tabel II.1 Tinjauan Penelitian Sebelumnya dari jurnal nasional**

No	Penelitian	Judul	Metode	Deskripsi
1.	Sandy Kurniawan, Taufiq Hidayat (2007)	Penerapan data mining dengan metode interpolasi untuk memprediksi minat konsumen asuransi	Interpolasi	Segmentasi pelanggan asuransi
2.	Dini Hmidin (2009)	Model customer relationship management (CRM) Di institusi pendidikan		
3.	Cakra Ramadhan, Yohana Dwi lulu W., Kartina Diah K.W., (2013)	Data mining dengan algoritma fuzzy C-Means clustering dalam kasus penjualan di PT Sepatu BATA	Fuzzy C-Means	Metode perhitungan fuzzy C-Means cocok digunakan pada aplikasi ini, karena bisa menghasilkan ouput berupa tingkat kelarisan produk pada penjualan sepatu bata

No	Penelitian	Judul	Metode	Deskripsi
4.	Putu Sukma Kurniawan (2015)	Perancangan data mining untuk analisis kriteria nasabah kredit yang potensial dan manfaatnya untuk customer relationship management perbankan	CRM	Manfaat data mining untuk konsep Customer Relationship Management perbankan adalah untuk membentuk kelompok nasabah (customer profiling) dan untuk bidang marketing dan customer care
5.	Armandyah amborowati, M. Suyanto (2015)	Studi dukungan marketing intelligence pada strategi pemasaran		Hasilnya adalah bahwa dukungan marketing intelligence pada strategi pemasaran dari sisi teknologi adalah melakukan segmentasi pasar / konsumen yang salah satunya adalah data mining
6.	A.A. Gde Bagus Ariana (2011)	Customer segmentation dengan metode self organizing map (Study Kasus UD: Fenny)	SOM, clustering	Berdasarkan penelitian yang telah dilakukan maka dapat disimpulkan bahwa metode SOM dapat digunakan untuk mengetahui pola-pola cluster pada sebuah basis data
7.	Trikaridha Hardiani, Selo Sulisty, Rudy Hartanto (2014)	Kajian Data mining Customer Relationship Management pada lembaga keuangan mikro	Fuzzy Clustering, Fuzzy C-Means, Fuzzy RFM, CRM	

No	Penelitian	Judul	Metode	Deskripsi
8.	Ferry Haryono Setiawan (2011)	Penerapan Fuzzy C-Means dan Apriori untuk rekomendasi promosi produk berdasarkan segmentasi konsumen	Fuzzy C-Means, Apriori	
9.	Bertha primaria bainana, Ari Basuki, Rahmat Hidayat	Identifikasi preferensi konsumen dalam minat beli rumah dengan pendekatan metode decision tree	Decision tree, AHP	Menggunakan 7 atribut
10.	Mike indra Istiara	Segmentasi pelanggan menggunakan algoritma K-Means sebagai dasar strategi pemasaran pada laroba seluler	Algoritma K-Means, RFM	
11.	Gita Febriana Wulandari	Segmentasi pelanggan menggunakan algoritma K-Means untuk Customer Relationship Management (CRM) studi kasus hijab miulan	K-Means, CRM	
12.	Alvian Kusumja Wijaya	Implementasi data mining dengan algoritma Fuzzy C-Means, Studi kasus penjualan UD Subur Baru	Fuzzy C-Means	
13.	Muhammad thoriq Agung, Bowo Nuhadiyono	Penerapan data mining pada data transaksi penjualan untuk mengatur penempatan barang	Algoritma apriori	

		menggunakan algoritma apriori		
14.	Yohana nugraheni (2011)	Data mining dengan metode fuzzy untuk customer relationship management (CRM) Pada perusahaan retail	Fuzzy , CRM, Clustering	

Berdasarkan *Studi Literature Review* pada tabel II.1 dan II.2, Pada penelitian ini Penulis mengangkat masalah yang diberi judul “Pemanfaatan Data Mining Pada Industri Knalpot Untuk Menerapkan *Customer Relationship Management*” dimana data yang diolah adalah data histori transaksi pelanggan Pada Industri Knalpot *Doctor Speed* Purbalingga, Metode yang digunakan adalah *clustering* menggunakan *Algoritma K-Means* berdasar analisis model RFM dan untuk mengetahui *cluster* yang paling optimal peneliti menguji *validitas cluster* menggunakan pendekatan *Davies Bouldein Indeks*, kemudian setelah *cluster* terbentuk akan diklasifikasikan menggunakan *Decission Tree Algoritma J48* sehingga diketahui masing-masing jenis dan karakteristik pelanggannya.

## 2.10. Tinjauan Objek Penelitian

Industri knalpot di Kabupaten Purbalingga, Jawa Tengah, hingga saat ini telah mampu menyerap 873 tenaga kerja, kata Kepala Dinas Perindustrian Perdagangan dan Koperasi (Dinperindagkop) Purbalingga. "Sementara untuk kondisi sentra knalpot Purbalingga saat ini didukung oleh 173 IKM (Industri Kecil dan Menengah) yang aktif dengan kapasitas produksi secara keseluruhan mencapai 313.380 unit dan nilai produksi mencapai Rp43,8 miliar," katanya di Purbalingga,.

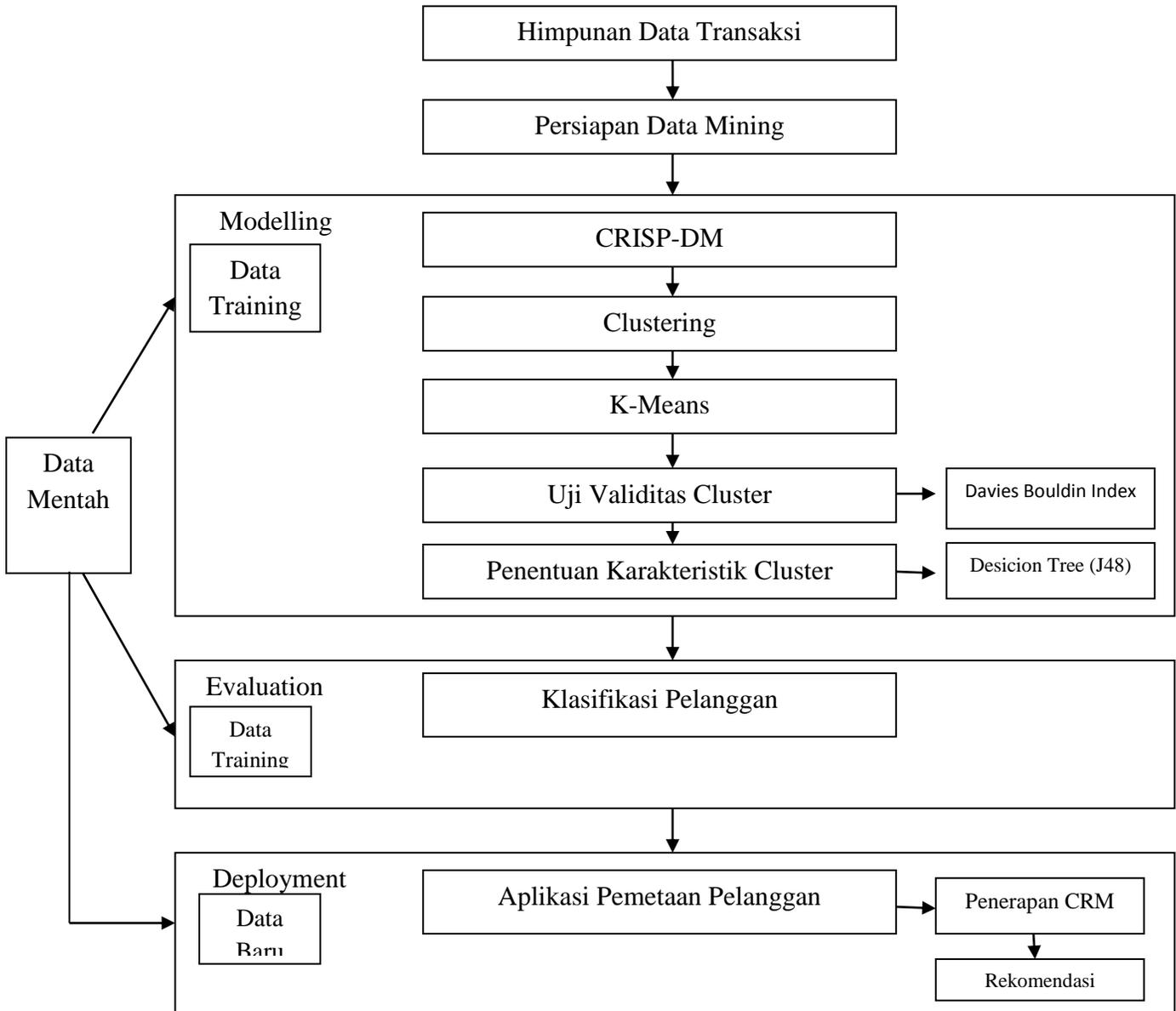
Industri Knalpot Motor dan Mobil *Doctor Speed* Purbalingga merupakan salah satu dari sekian banyak yang menjual knalpot di wilayah Pesayangan, Purbalingga. Di tengah persaingan bisnis yang sangat ketat industri *doctor speed* knalpot purbalingga terus berupaya mempertahankan bisnisnya, namun dalam prosesnya banyak terjadi kendala agar terus dapat bersaing dengan industri yang sejenis. Untuk itu dibutuhkan analisis terhadap segmentasi pelanggannya agar

mengetahui profil dari pelanggannya yang dapat digunakan untuk menerapkan *Customer Relationship Management*.

## BAB III METODOLOGI PENELITIAN

### 3.1. Metodologi Penelitian

Penelitian ini dilakukan dalam beberapa tahapan mulai dari pra-proses data, pengelompokan data (*clustering*), *validasi cluster*, dan *klasifikasi*. Berikut langkah penelitian pada Gambar III.1.



**Gambar III.1 Langkah Penelitian**

### 3.1.1. Jenis Penelitian

Dua pendekatan utama dalam penelitian yaitu pendekatan kualitatif dan pendekatan kuantitatif. Metode penelitian kualitatif berhubungan dengan penilaian subjektif dari sikap, pendapat, dan perilaku. Secara umum teknik yang digunakan adalah *interview* pada kelompok tertentu dan wawancara yang mendalam (Kothari, 2004).

Metode penelitian kuantitatif digunakan untuk meneliti pada *populasi* atau *sample* tertentu, pengumpulan data menggunakan instrumen penelitian, analisis data bersifat kuantitatif/statistik dengan tujuan untuk menguji hipotesis yang telah ditentukan. Metode penelitian kuantitatif disebut juga dengan metode *discovery* karena dengan metode ini dapat ditemukan dan dikembangkan berbagai iptek baru (Sugiyanto, 2008).

Metode penelitian kuantitatif dapat dibagi menjadi tiga sub kategori yaitu *inferentia*, *experimental*, dan *simulasi*. Desain riset yang digunakan dalam penelitian ini adalah eksperimen. Metode penelitian eksperimen dapat diartikan sebagai metode penelitian yang digunakan untuk mencari pengaruh perlakuan tertentu terhadap yang lain dalam kondisi yang terkendalikan (Sugiyanto, 2008). Misalnya pengaruh ruang kerja terhadap produktivitas kerja karyawan.

Metode eksperimen dikarakterisasi dengan lebih banyak mengontrol lingkungan penelitian, dalam hal ini beberapa *variabel* dimanipulasi untuk mengobservasi efek terhadap *variabel* lain. Dalam metode eksperimen peneliti mengukur efek dari eksperimen yang dilakukan dengan sengaja. Tujuan akhir dari eksperimen adalah untuk men-generalisasi hubungan *variabel* sehingga dapat diterapkan pada populasi yang lebih luas diluar laboratorium (Kothari, 2004).

Metode eksperimen dibagi menjadi dua, yaitu eksperimen absolut dan eksperimen komparatif. Eksperimen absolut mengarah pada dampak yang dihasilkan dari eksperimen, sedangkan eksperimen komparatif membandingkan dua objek yang berbeda (Kothari, 2004). Dalam penelitian ini digunakan metode eksperimen absolut.

### 3.1.2. Metode Pengumpulan Data

#### 1. Sumber Data

Metode pengumpulan data untuk mendapatkan sumber data dibagi menjadi dua, yaitu data primer dan data sekunder. Data primer yaitu data yang dikumpulkan pertama kali, dan untuk melihat apa yang sesungguhnya terjadi melalui observasi, *interview*, *kuesioner*, dll. Data sekunder adalah data yang telah dikumpulkan dan dianalisis oleh orang lain baik yang telah dipublikasikan maupun yang belum dipublikasikan, misalnya dari dokumentasi, literatur, buku, jurnal, dan informasi lainnya yang ada hubungannya dengan masalah yang diteliti.

Dalam penelitian ini metode pengumpulan data untuk mendapatkan sumber data yang digunakan adalah metode pengumpulan data primer. Data utama diperoleh dari data transaksi *industri knalpot motor dan mobil doctor speed Purbalingga*, sedangkan data pendukung didapatkan dari buku, jurnal dan publikasi lainnya.

#### 2. Sampel Penelitian

Sampel dari Penelitian ini adalah data transaksi dari *industri knalpot motor dan mobil doctor speed Purbalingga*

### 3.1.3. Eksperimen

Dalam melakukan eksperimen ini, penulis menggunakan model *Cross- Industry Standard Process for Data Mining* (CRISP-DM) yang terdiri dari enam tahap, yaitu (Larose, 2006):

#### 1. Tahap *Business understanding*

Tahapan ini dilakukan sebelum segmentasi pelanggan. Tahapan ini terdiri dari tiga bagian, yaitu :

- a. Pemahaman terhadap tujuan bisnis
- b. Penilaian sesuatu
- c. Penerjemahan tujuan bisnis dalam data Mining

#### 2. Tahap *data understanding*

Data transaksi diperoleh dari *industri knalpot motor dan mobil doctor speed Purbalingga*, yang terdiri dari 326 data konsumen.

### 3. Tahap *data preparation*

Persiapan data merupakan tahapan dimana peneliti melakukan *query* data untuk mendapatkan data yang siap untuk di *mining*. Tahapan ini menggunakan *Microsoft Excel*.

### 4. Tahap *data modelling*

Segmentasi pelanggan menggunakan *proses clustering-data mining* metode *Hierarchical K-Means*. Proses *clustering* ini dilakukan dengan tiga *variabel*, yaitu *Recency-Frekuensi-Monetary*. Dalam konsep RFM ini, setiap dimensi memberikan informasi yang unik mengenai perilaku transaksi masing-masing pelanggan yang dimiliki perusahaan:

- a. *Recency* – Berdasarkan penelitian yang telah dilakukan bertahun-tahun, statistik memperlihatkan bahwa pelanggan yang belakangan ini melakukan transaksi pembelian akan cenderung melakukannya lagi dalam waktu dekat.
- b. *Frequency* – Sejarah memperlihatkan bahwa pelanggan yang secara berkala melakukan pembelian, akan cenderung melakukannya lagi di kemudian hari.
- c. *Monetary Value* – Pelanggan yang telah mengalokasikan cukup banyak dana untuk melakukan pembelian akan cenderung melanjutkan kebiasaan membelinya. Tabel III.1 menunjukkan pengertian ketiga *variabel* tersebut.

Tabel III.1 *Variabel Proses Clustering*

Variabel	Keterangan
<i>Recency</i>	Jarak pemakaian (interval) antara pembelian terakhir pada periode tertentu
<i>Frequency</i>	Banyaknya jumlah pembelian dalam periode tertentu
<i>Monetary</i>	Jumlah uang selama periode tertentu

Berikut adalah tahapan penentuan variabel *recency-frekuensi-monetary*:

- a. Menghitung frekuensi pembelian *customer* setiap bulan dan menjumlahkannya untuk mendapatkan frekuensi pembelian *customer* dalam 1 tahun.
- b. Mengitung *recency* setiap *customer*
- c. Menghitung total beli selama setahun

Setelah proses tersebut berikutnya adalah pembentukan *data processing* pada studi kasus ini bertujuan untuk mencari bentuk aturan data yang sesuai dengan proses data mining. Proses ini dilakukan melalui pengubahan isi data pada *attribute* yang ada pada data transaksi penjualan industri knalpot doctor speed Purbalingga dengan melakukan proses *konversi* data dari data sumber menjadi data *numerik* sesuai dengan tabel *konversi* sehingga didapatkan angka yang *valid*. Hasil dari proses tersebut kemudian diintegrasikan menjadi sebuah tabel baru sehingga nantinya didapatkan sebuah tabel nilai transaksi yang akan digunakan untuk melakukan proses menjadi beberapa *cluster*. Proses *clustering* ini bertujuan untuk mengidentifikasi dan menentukan jumlah pelanggan pada kelas *profitable* dan *less/unprofitable customer* Setelah mendapatkan tiga variabel tersebut dilakukan proses *scoring*.

Tabel III. 2 menunjukkan score *Recency-Frekuensi-Monetary*

Score	R		F		M	
5	Sangat Pendek	< 3 Bulan	Sangat Tinggi	> 20 Kali	Sangat Banyak	>80 Juta
4	Pendek	3 - 10 Bulan	Tinggi	10 - 20 Kali	Banyak	40 - 80 Juta
3	Biasa	10 - 2 Tahun	Biasa	5 - 10 Kali	Normal	20 - 40 Juta
2	Jauh	2 - 3 Tahun	Rendah	2 - 5 Kali	Sedikit	500 - 20 Juta
1	Sangat Jauh	> 3 Tahun	Sangat Rendah	< 2 Kali	Sangat Sedikit	< 500 Ribu

*Score 5* merupakan *score* tertinggi yang artinya terbaik dari data yang ada. Untuk *recency*, *score 5* menunjukkan *interval* terpendek. Kemudian untuk frekuensi dan *Monetary score 5* menunjukkan nilai frekuensi dan nilai *monetary* terbesar. Namun, untuk *recency* pada *score 1* menunjukkan *interval* terpanjang. Kemudian untuk frekuensi dan *Monetary score 1* menunjukkan nilai frekuensi dan *Monetary* terkecil. Setelah tahapan *scoring*, dilakukan proses *clustering* dengan *Hierarchichal K-Means* secara perhitungan manual dan diuji pada *software Rapid Miner*. Berikut adalah tahapan metode *Hierarchichal K-Means*:

1. Melakukan algoritma *Hierarchichal*:
  - a. Input hasil *scoring* data pelanggan berupa *Recency-Frekuensi-Monetary*
  - b. Menghitung jarak antar data dengan *Euclidean Distance*

$$D(i,j) = \sqrt{(X1i - X1j)^2 + (X2i - X2j)^2 + \dots + (XKi - XKj)^2}$$

Dimana:

$D(i,j)$  = Jarak data ke  $i$  ke pusat cluster  $j$

$Xki$ =Data ke  $i$  pada atribut data ke  $k$

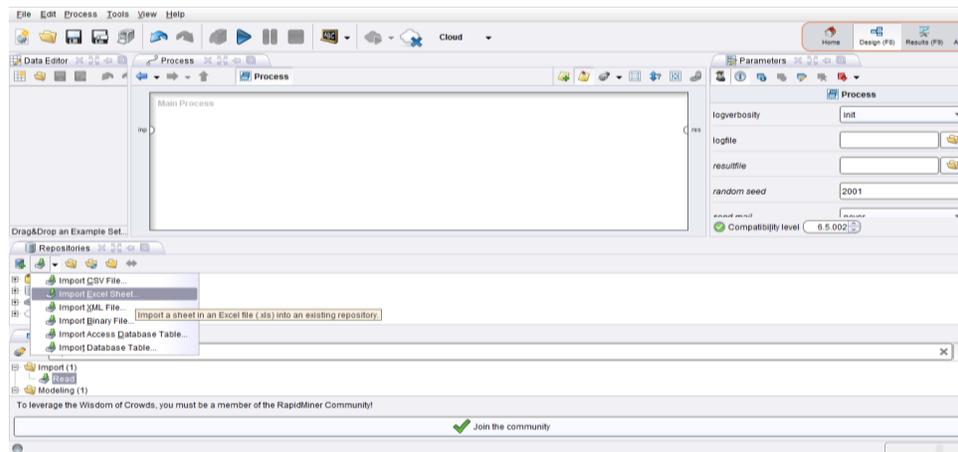
$Xkj$ =Titik pusat ke  $j$  pada atribut ke  $k$

- c. Memilih jarak terdekat untuk digabungkan dalam *euclidean distance*
- d. Mengetahui jumlah *cluster*
2. Melakukan *algoritma K-Means*
  - a. Input jumlah *cluster*
  - b. Menghitung nilai rata-rata pusat *cluster*
  - c. Menghitung jarak masing-masing pusat *cluster* dengan *euclidean distance*
  - d. Memilih *centroid* terdekat untuk menghitung pusat *cluster* terbaru
  - e. Menyimpan hasil *clusterisasi* jika *centroid* sudah tidak berubah dan *konvergen*.

Tahap *modelling* dilakukan dengan dua cara yaitu dengan cara manual dan dengan menggunakan *software rapid miner*.

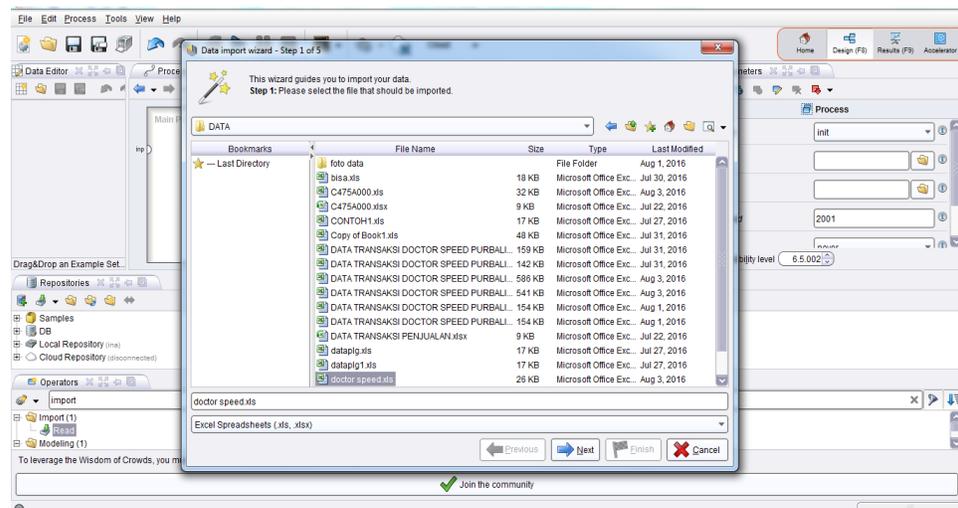
Tahap *modelling* dengan menggunakan *software rapid miner* dapat dilakukan dengan cara sebagai berikut:

- a. Untuk membentuk model diperlukan data *training*, *rapid miner* dapat membaca data dengan format .csv, .xls, .mdb, dan lain-lain. Data yang digunakan dalam pembuatan model pada penelitian ini adalah dalam format .xls
- b. Buka *software rapid miner*, masukan data *training* yang telah disiapkan pada tahap data *preparation*, pada tab operators pilih Import – Data – Read Excel tarik ke area *main process*. Pada area parameters klik Import Configuration Wizard.



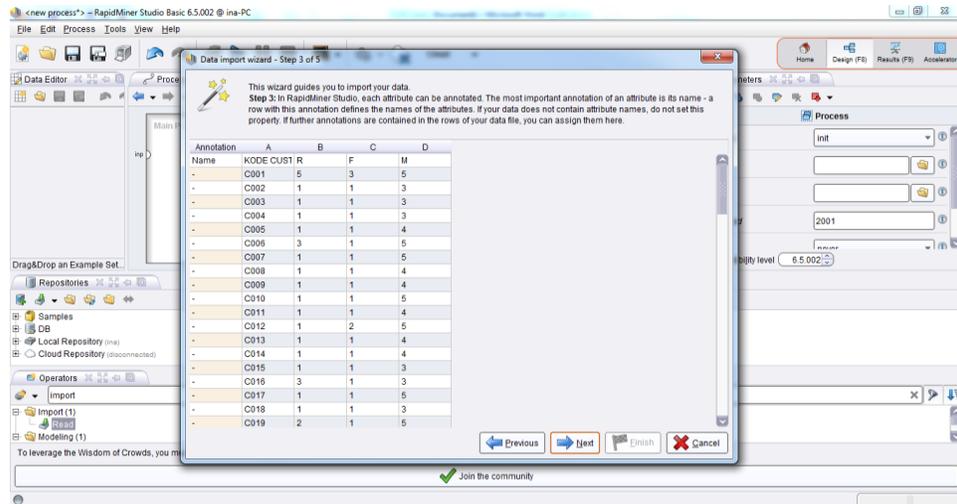
Gambar III. 2 Import Data Set Pada Lembar Kerja Rapid Miner

- c. Tampil jendela *Import Configuration Wizard*, pada langkah ini tentukan data berformat .xls yang akan digunakan sebagai *data training* kemudian klik tombol *next*.



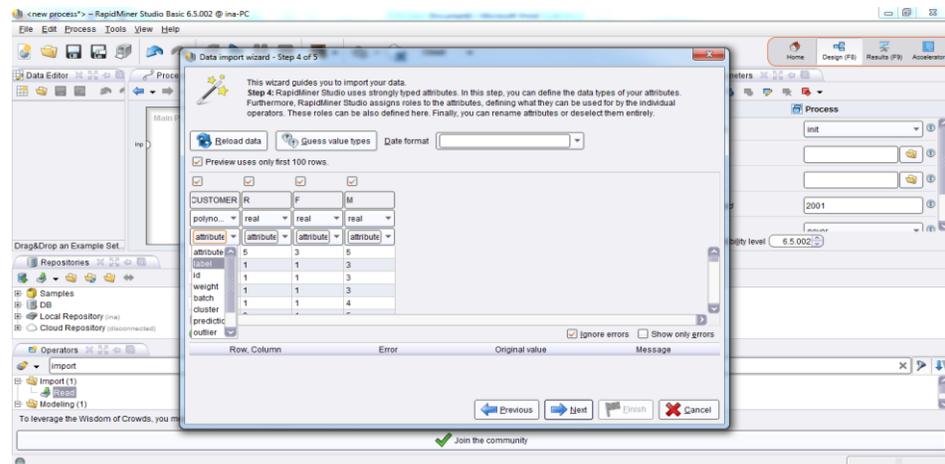
Gambar III.3 Data Set berupa excel yang akan diimpor

- d. Langkah ke dua yaitu memilih *sheet* yang akan digunakan sebagai data *training*



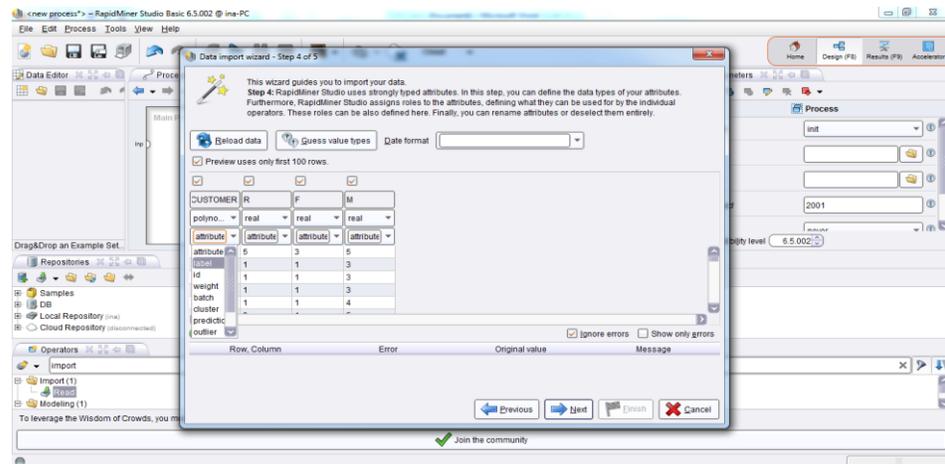
Gambar III.4 Tampilan sheet data

- e. langkah ke tiga menentukan anotasi dari setiap *tupel*, langkah ini dapat diabaikan. Kemudian klik *Next*.



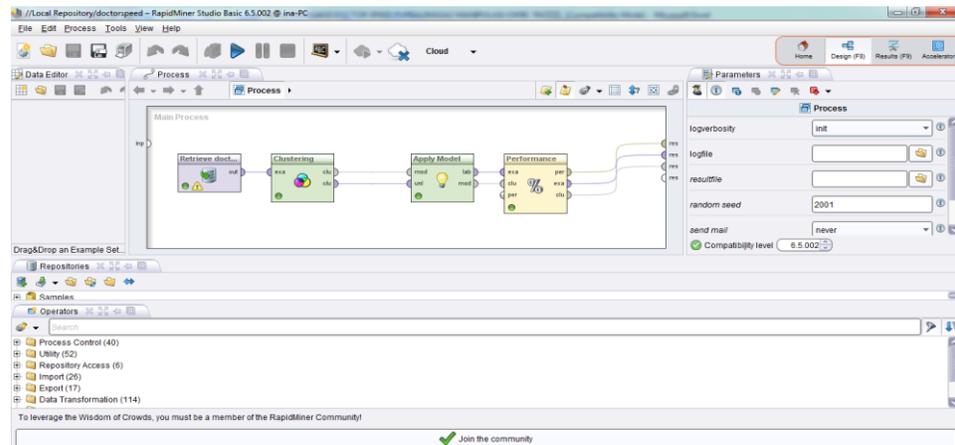
Gambar III.5 Menentukan Anotasi Dari Setiap *Tupel*

- f. Langkah ke empat yaitu menentukan nama atribut dan tipe data dari masing-masing atribut. Pada atribut *Customer* ubah tipe data menjadi label. Klik *finish*



Gambar III.6 Atribut *Customer*

- g. Setelah data training disiapkan, proses pembuatan model dapat dilakukan dengan cara pilih tab *Operators - Modelling - Clustering and Segmentation - K-means*. Geser *K-Means* ke area *main process* dan hubungkan seperti terlihat pada gambar berikut:



Gambar III.7 Pembentukan Model pada Rapid Miner

## 5. Tahap *Evaluation*

Pada tahap ini, model sudah terbentuk dan diharapkan memiliki kualitas baik jika dilihat dari sudut pandang analisa data. Kunci dari tahap ini adalah menentukan apakah ada masalah bisnis yang belum dipertimbangkan. Pada proses ini masih dilakukan klasifikasi untuk mengetahui jenis dan karakteristik masing-masing

pelanggan yaitu menggunakan *Decision Tree*. Di akhir dari tahap ini harus ditentukan penggunaan hasil proses *data mining*.

#### 6. Tahap *Deployment*

Pada tahap ini akan dilakukan pembuatan aplikasi data mining, untuk menghasilkan laporan data mining. Pada penelitian ini proses *deployment* adalah melakukan proses mining pada data baru, yaitu menggunakan program aplikasi pemetaan pelanggan, dengan menggunakan aplikasi tersebut akan sangat membantu perusahaan dalam mengetahui profil pelanggan dimasa yang akan datang.

### 3.2 Jadwal Penelitian

Jadwal penelitian dalam penyusunan tesis ini tercantum dalam tabel III.3.berikut :

Tabel III.3. Jadwal Penelitian

No	Kegiatan	April	Mei	Juni	Juli	Agust
1	Pencarian dan Pemilihan Obyek Penelitian					
2	Pengambilan data					
3	Persiapan data					
4	Pembuatan model					
5	Pengujian model					
6	Pembuatan laporan tesis					

## BAB IV HASIL PENELITIAN DAN PEMBAHASAN

### 4.1. Hasil Penelitian

Pada bagian hasil dan pembahasan akan diuraikan hasil-hasil dari tahapan penelitian dan dilakukan pula pembahasan dari setiap hasil tersebut. Data yang digunakan dalam penelitian ini adalah data transaksi pelanggan dari tahun 2014 sampai 2016. Dalam penelitian ini dilakukan analisa terhadap hasil *clustering* dan segmentasi data transaksi dalam periode 3 tahun. Periode 3 tahun dipilih pada proses *data mining* ini atas pertimbangan bahwa 3 tahun merupakan periode yang paling tepat untuk menggambarkan perilaku konsumen. Hasil *clustering* untuk algoritma *K-Means* diuji dengan *Davies Bouldin Index*, untuk menguji *validitas cluster*.

#### 4.1.1. Persiapan Data Mining

##### 1. Himpunan Data Transaksi

Pada Industri Knalpot terdapat banyak Himpunan Data Transaksi yang belum dimanfaatkan dengan baik, contoh seperti Gambar IV.1. dibawah ini:



Gambar IV.1 Nota Transaksi Penjualan Gambar IV.2 Attribute yang dihimpun

## 2. *Data Preparation*

Pada tahap persiapan data, peneliti melakukan *query* data untuk mendapatkan data yang siap untuk di-*mining*. Tabel IV.1 adalah hasil *query* data dengan Microsoft Excel (Secara lengkap lihat lampiran)

**Tabel IV.1 Data set Transaksi Penjualan**

KODE CUSTOMER	Recency (Terakhir Pembelian)	Frequency	Monetary (Total Bayar)
C001	27/07/2016	26	Rp 139.185.000
C002	04/02/2015	9	Rp 9.625.000
C003	13/06/2015	5	Rp 17.725.000
C004	27/05/2015	12	Rp 39.630.000
C005	14/03/2015	8	Rp 24.550.000

## 3. *Data Modelling*

Tahap ini adalah melakukan proses *clustering* dengan menggunakan metode *Clustering* yaitu *K-Means*. Sebelum melakukan proses *clustering*, perlu ditentukan kriteria pembobotannya terlebih dahulu. Untuk pembobotan diambil dari data *recency*, *frekuensi*, *Monetary* harus dibagi ke dalam 5 skala/score. Tabel IV.2 menunjukkan skala *score recency*, *frekuensi*, dan *Monetary* untuk data penjualan tahun 2014-2016.

Berikut proses menentukan skala/bobot yang digunakan

- a. Hitunglah *mean* (nilai rata-rata) dari seluruh data dengan rumus :

$$\bar{x} = \frac{\sum x}{n} \dots\dots\dots (1)$$

Keterangan :

$\bar{x}$  : Rata-rata

$\sum x$  : Hasil penjumlahan data

n : Banyak data

- b. Hitunglah standar deviasi dari seluruh data dengan rumus :

$$S = \sqrt{\sum \frac{(x-\bar{x})^2}{n-1}} \dots\dots\dots (2)$$

Keterangan :

S : Standar Deviasi

X : Nilai data

$\bar{x}$  : Nilai rata-rata

n : Banyak data

berdasarkan rumus standar nilai mean dan standar deviasi yang diperoleh maka dihasilkan tabel pembobotan *Recency*, *Frequency* dan *Monetary* secara berturut-turut pada Gambar IV.2. – IV.5:

**Tabel IV.2. Kategori *Recency***

Score	R	
5	Sangat Pendek	$R < \bar{x} - S$
4	Pendek	$\bar{x} - S < R < \bar{x}$
3	Biasa	$\bar{x} < R < S$
2	Jauh	$S < R < \bar{x} + S$
1	Sangat Jauh	$R > \bar{x} + S$

**Tabel IV.3 Kategori *Frequensi***

Score	F	
5	Sangat Tinggi	$F > \bar{x} + S$
4	Tinggi	$S < F < \bar{x} + S$
3	Biasa	$\bar{x} < F < S$
2	Rendah	$\bar{x} - S < F < \bar{x}$
1	Sangat Rendah	$F < \bar{x} - S$

**Tabel IV.4. Kategori Monetary**

Score	M	
5	Sangat Banyak	$M > \bar{x} + S$
4	Banyak	$S < M < \bar{x} + S$
3	Normal	$\bar{x} < M < S$
2	Sedikit	$\bar{x} - S < M < \bar{x}$
1	Sangat Sedikit	$M < \bar{x} - S$

**Tabel IV.5 Pembobotan RFM**

Score	R		F		M	
5	Sangat Pendek	< 3 Bulan	Sangat Tinggi	> 20 Kali	Sangat Banyak	>80 Juta
4	Pendek	3 - 10 Bulan	Tinggi	10 - 20 Kali	Banyak	40 - 80 Juta
3	Biasa	10 - 2 Tahun	Biasa	5 - 10 Kali	Normal	20 - 40 Juta
2	Jauh	2 - 3 Tahun	Rendah	2 - 5 Kali	Sedikit	500 - 20 Juta
1	Sangat Jauh	> 3 Tahun	Sangat Rendah	< 2 Kali	Sangat Sedikit	< 500 Ribu

Tabel IV.6 menunjukkan transformasi data setelah dilakukan *score recency*, *frekuensi*, dan *Monetary* data penjualan Industri Knalpot Motor dan Mobil Pada Doctor Speed Purbalingga dan sudah siap untuk dilakukan proses *data mining*.

**Tabel IV.6 Data Penjualan Score RFM**

KODE CUSTOMER	R	F	M
C001	5	4	5
C002	1	2	3
C003	1	2	4
C004	1	3	5
C005	1	2	4
C006	2	4	5
C007	1	1	3
C008	3	4	5
....	.....	.....	.....
....	.....	.....	.....
C326	1	3	5

Setelah semua data transaksi ditransformasi ke dalam bentuk angka, maka data-data tersebut telah dapat dikelompokkan dengan menggunakan algoritma *K-Means Clustering*. Untuk dapat melakukan pengelompokan data-data tersebut menjadi beberapa *cluster* perlu dilakukan beberapa langkah, yaitu:

1. Tentukan jumlah cluster yang diinginkan. Dalam penelitian ini data-data yang ada akan dikelompokkan mejadi empat *cluster*.
2. Tentukan titik pusat awal dari setiap cluster. Dalam penelitian ini titik pusat awal ditentukan secara random dan didapat titik pusat dari setiap cluster dapat dilihat pada tabel IV.7

**Tabel IV.7. Titik Pusat Awal Setiap Cluster**

TITIK PUSAT CLUSTER	KODE CUSTOMER	R	F	M
CLUSTER 1	C021	5	2	2
CLUSTER 2	C087	3	4	3
CLUSTER 3	C244	4	3	3
CLUSTER 4	C030	4	3	4

Tempatkan setiap data pada *cluster*. Dalam penelitian ini digunakan metode *k-means* untuk mengalokasikan setiap data ke dalam suatu *cluster*, sehingga data akan dimasukan dalam suatu *cluster* yang memiliki jarak paling dekat dengan titik pusat dari setiap *cluster*. Untuk mengetahui *cluster* mana yang paling dekat dengan data, maka perlu dihitung jarak setiap data dengan titik pusat setiap *cluster*. Sebagai contoh, akan dihitung jarak dari data transaksi customer pertama ke pusat *cluster* pertama:

$$D(i,j) = \sqrt{(X1i - X1j)^2 + (X2i - X2j)^2 + \dots + (XKi - XKj)^2}$$

$$D(1,1) = \sqrt{(5 - 5)^2 + (5 - 2)^2 + (5 - 2)^2}$$

Dari hasil perhitungan di atas di dapatkan hasil bahwa jarak data transaksi customer pertama dengan pusat *cluster* pertama adalah 4,2426407.

Jarak data mahasiswa pertama ke pusat *cluster* kedua:

$$D(1,2) = \sqrt{(5 - 3)^2 + (5 - 4)^2 + (5 - 3)^2}$$

Dari hasil perhitungan di atas didapatkan hasil bahwa jarak data transaksi customer pertama dengan pusat *cluster* kedua adalah 3,0000000.

Jarak data transaksi pertama ke pusat *cluster* ketiga:

$$D(1,3) = \sqrt{(5 - 4)^2 + (5 - 3)^2 + (5 - 3)^2}$$

Dari hasil perhitungan di atas didapatkan hasil bahwa jarak data transaksi customer pertama dengan pusat *cluster* ketiga adalah 3,0000000.

$$D(1,4) = \sqrt{(5 - 4)^2 + (5 - 3)^2 + (5 - 4)^2}$$

Dari hasil perhitungan di atas didapatkan hasil bahwa jarak data transaksi customer pertama dengan pusat *cluster* ketiga adalah 2,4494897.

Berdasarkan hasil ketiga perhitungan di atas dapat disimpulkan bahwa jarak data transaksi customer pertama yang paling dekat adalah dengan cluster 4, sehingga data transaksi pertama dimasukkan ke dalam cluster 4. Hasil perhitungan selengkapnya dapat di lihat pada tabel IV.8.

**Tabel IV.8. Contoh Hasil Perhitungan Setiap Data ke Setiap Cluster**

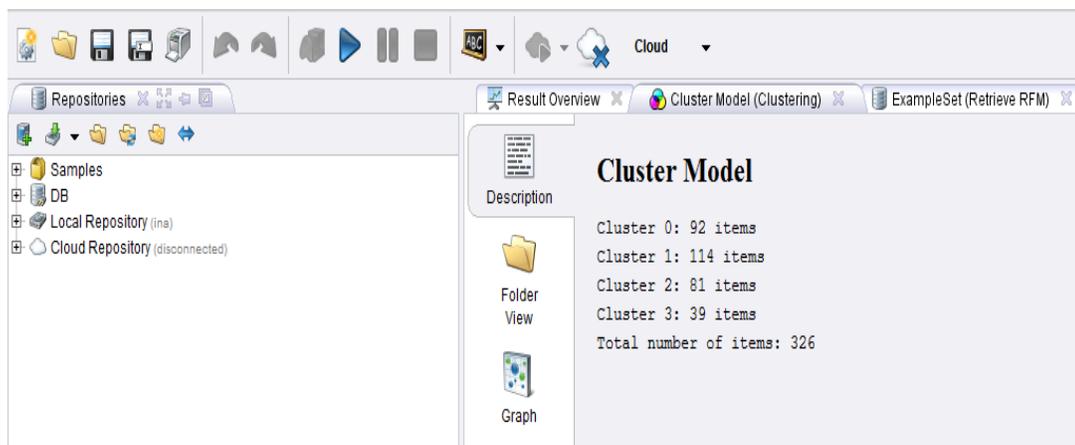
Data Ke-i	KODE CUSTOMER	R	F	M	Jarak ke centroid				Terdekat	Cluster Yang Diikuti
					C1	C2	C3	C4		
1	C001	5	5	5	4,243	3,000	3,000	2,449	2,449	4
2	C002	3	3	2	2,236	1,414	1,414	2,236	1,414	2
3	C003	3	3	2	2,236	1,414	1,414	2,236	1,414	2
4	C004	3	4	3	3,000	0,000	1,414	1,732	0,000	2

4. Setelah semua data ditempatkan ke dalam *cluster* yang terdekat, kemudian hitung kembali pusat *cluster* yang baru berdasarkan rata-rata anggota yang ada pada *cluster* tersebut.
5. Setelah didapatkan titik pusat yang baru dari setiap *cluster*, lakukan kembali dari langkah ketiga hingga titik pusat dari setiap *cluster* tidak berubah lagi dan tidak ada lagi data yang berpindah dari satu *cluster* ke *cluster* yang lain

## 4.2. Pembahasan

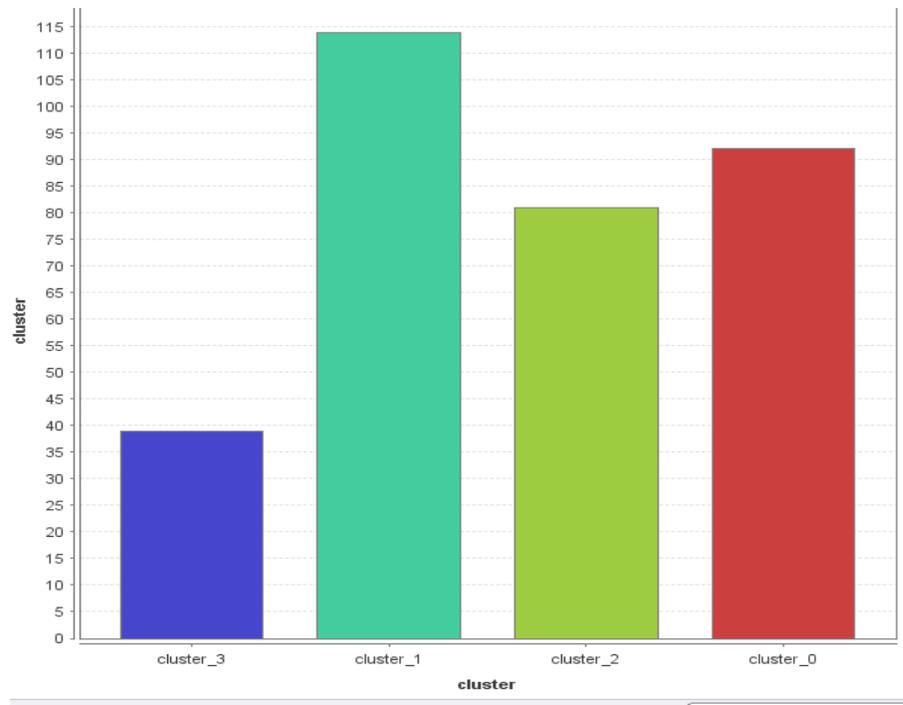
### 4.2.1. Hasil Penghitungan *Cluster*

Dari hasil pengolahan data yang dilakukan berdasarkan *dataset* transaksi pelanggan menggunakan *K-Means* melalui 4 iterasi terbentuk *cluster* seperti pada Gambar IV.3 menunjukkan hasil clustering didapatkan 92 anggota *cluster* 1, 114 anggota *cluster* 2, 81 anggota *cluster* 3 dan 39 anggota *cluster* 4.



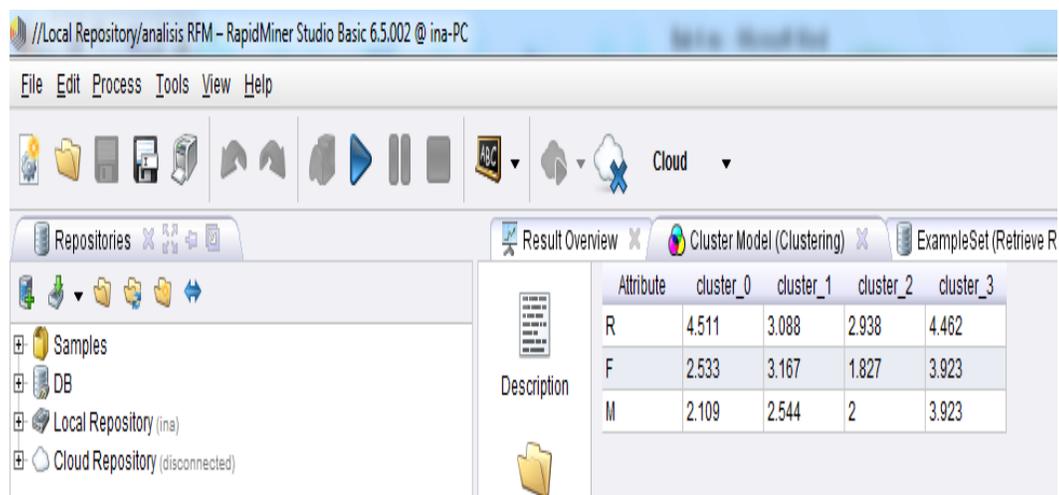
**Gambar IV.3 Hasil Proses Clustering**

### Tampilan Grafik Hasil *Clustering* menggunakan *K-Mean*



**Gambar IV.4. Grafik Hasil *Clustering***

Untuk *centroid* seperti Gambar IV.5.



**Gambar IV.5. *Centroid*/Pusat *Cluster* yang terbentuk**

Selanjutnya perlu dilakukan uji *validitas cluster* dengan menghitung *Davies Bouldin Index* untuk memastikan apakah jumlah cluster yang terbentuk sudah optimal atau belum

#### 4.2.2. Hasil Cluster Optimal

Suatu *cluster* dinyatakan *konvergen* apabila sudah tidak terjadi adanya perubahan atau perpindahan anggota dari satu *cluster* ke *cluster* lain. Selain itu, cluster yang telah *konvergen* juga ditandai oleh tidak adanya perubahan nilai *DB-index*. Dari kelima skenario ujicoba yang telah dilakukan, akan dibuat suatu peringkat berdasarkan nilai *DB-index* terbaik. Tingkat optimal suatu *cluster* dapat diukur Nilai *DB Index. Davies dan Bouldien* (1979) menyebutkan bahwa jumlah *cluster* yang terbaik adalah yang memiliki index terkecil diantara *cluster* lainnya.

**Tabel IV.9 Perbandingan Performa Cluster**

Jumlah Cluster	DB-index
2	0,401
3	0,323
4	0,301
5	0,313

Hasilnya, untuk jumlah klaster 4 memiliki nilai *Davies Bouldien Index* (DI) yang paling kecil diantara jumlah cluster yang lain yaitu 0, 301 tercantum dalam Tabel IV.9.

#### 4.2.3 Penentuan Karakteristik Pelanggan

Setelah menemukan segmentasi pelanggan selanjutnya akan dilakukan analisis setiap segmen tersebut untuk menentukan karakteristik dari pelanggan. Penentuan karakteristik Pelanggan diawali dengan melakukan klasifikasi terhadap data transaksi pelanggan yang diimplementasikan dalam WEKA. Tujuan dari klasifikasi adalah menggali informasi penting yang terdapat dalam data transaksi pelanggan yaitu berdasarkan *Recency, Frequency, Monetary* dan *Cluster* Tabel IV.10

menunjukkan secara lengkap jumlah data transaksi Pelanggan. Pada proses ini dipilih sebanyak 326 *Instances* yang terdiri dari 4 *cluster*.

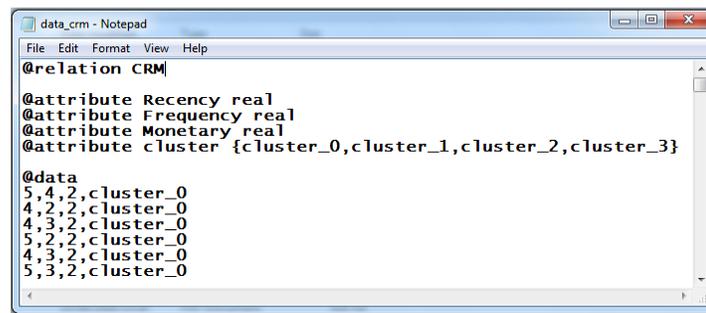
**Tabel IV.10. Data set transaksi Pelanggan**

No	Cluster	Number Of Instances
1	Cluster 0	92
2	Cluster 1	114
3	Cluster 2	81
4	Cluster 3	39

#### 4.2.4. Model Klasifikasi

Pembentukan Model klasifikasi dilakukan dengan implementasi data ke dalam GUI *Explorer* WEKA dengan klasifikasi J48. Format.arff yang berjumlah 326 data menghasilkan model *tree* dan *rule* penentuan karakteristik Pelanggan

**Tabel IV.11. Format .arff Data Cluster**



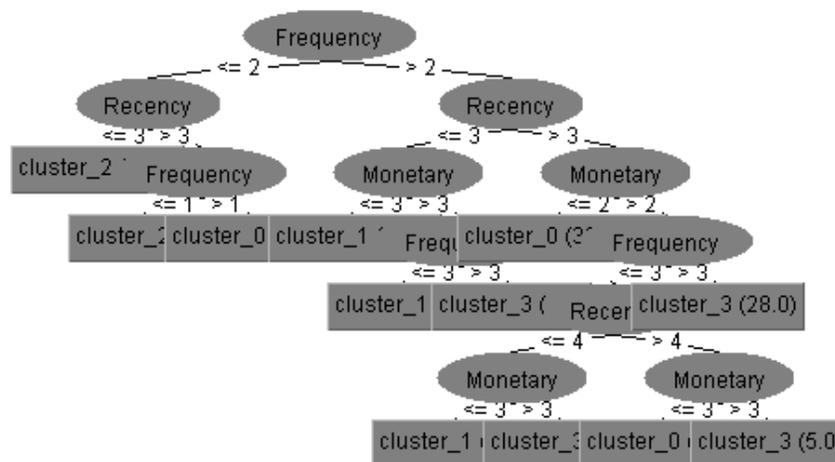
```

@relation CRM
@attribute Recency real
@attribute Frequency real
@attribute Monetary real
@attribute cluster {cluster_0,cluster_1,cluster_2,cluster_3}

@data
5,4,2,cluster_0
4,2,2,cluster_0
4,3,2,cluster_0
5,2,2,cluster_0
4,3,2,cluster_0
5,3,2,cluster_0

```

Model klasifikasi yang dihasilkan dapat dilihat berupa *tree* pada gambar IV. 3. Dan *rule* dari *tree* yang disajikan pada gambar IV.4. Model klasifikasi ini melibatkan 4 atribut yaitu *Recency*, *Frequency*, *Monetary* dan *Cluster*.



**Gambar IV.6. Model *Tree* untuk Penentuan Karakteristik Pelanggan**

```

Classifier output
Frequency <= 2
| Recency <= 3: cluster_2 (75.0)
| Recency > 3
| | Frequency <= 1: cluster_2 (6.0)
| | Frequency > 1: cluster_0 (44.0)
Frequency > 2
| Recency <= 3
| | Monetary <= 3: cluster_1 (91.0)
| | Monetary > 3
| | | Frequency <= 3: cluster_1 (3.0)
| | | Frequency > 3: cluster_3 (4.0)
| Recency > 3
| | Monetary <= 2: cluster_0 (39.0/1.0)
| | Monetary > 2
| | | Frequency <= 3
| | | | Recency <= 4
| | | | | Monetary <= 3: cluster_1 (19.0)
| | | | | Monetary > 3: cluster_3 (2.0)
| | | | | Recency > 4
| | | | | Monetary <= 3: cluster_0 (10.0)
| | | | | Monetary > 3: cluster_3 (5.0)
| | | | Frequency > 3: cluster_3 (28.0)
  
```

**Gambar IV.7. Rule *Fitur* untuk Penentuan Karakteristik Pelanggan**

Hasil prediksi dari *rule* penentuan karakteristik Pelanggan pada 326 data *training* menghasilkan prediksi **98,1595%** *correctly classified intances* dan **1,8405%** *incorectly classified intances*. Prediksi ini sangat baik karena hanya **1,8405%** menghasilkan prediksi yang salah.

```

Classifier output

Time taken to build model: 0.06 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      320          98.1595 %
Incorrectly Classified Instances    6            1.8405 %
Kappa statistic                     0.9746
Mean absolute error                 0.01
Root mean squared error            0.0874
Relative absolute error             2.7554 %
Root relative squared error        20.5652 %
Total Number of Instances          326

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
          0.989    0        1          0.989  0.995     0.994    cluster_0
          0.974    0.009  0.982     0.974  0.978     0.995    cluster_1
          1        0        1          1        1         1        cluster_2
          0.949    0.014  0.902     0.949  0.925     0.998    cluster_3
Weighted Avg.  0.982    0.005  0.982     0.982  0.982     0.996

=== Confusion Matrix ===

 a  b  c  d  <-- classified as
91  0  0  1 |  a = cluster_0
0 111 0  3 |  b = cluster_1
0  0 81  0 |  c = cluster_2
0  2  0 37 |  d = cluster_3

```

**Gambar IV.8. Hasil Prediksi Data *Training* Fitur Penentuan Karakteristik Pelanggan.**

Berdasarkan *rule* yang diolah menggunakan *Algoritma Decision Tree J48* dan Menurut Hill (1996) tentang Jenis-jenis Pelanggan dapat diputuskan karakteristik Pelanggan pada tabel IV.12

**Tabel IV.12. Karakteristik Pelanggan Setiap Cluster**

<b>Jenis Pelanggan</b>	<b>Cluster</b>	<b>Rule</b>	<b>Karakteristik Pelanggan</b>
<i>Advocates</i>	4	<ul style="list-style-type: none"> <li>- <i>Frequency</i> (tinggi) AND <i>Recency</i> (biasa) AND <i>Monetary</i> (banyak)</li> <li>- <i>Frequency</i> (biasa) (pendek) AND <i>Recency</i> (pendek) AND <i>Monetary</i> (banyak)</li> <li>- <i>Frequency</i> (biasa) (pendek) AND <i>Recency</i> (pendek) AND <i>Monetary</i> (normal)</li> <li>- <i>Frequency</i> (tinggi) (pendek) AND <i>Recency</i> (pendek) AND <i>Monetary</i> (normal)</li> </ul>	<p><i>Pelanggan dalam Cluster 4 diberi nama jenis (advocates) ini merupakan pelanggan yang tingkat kedatangannya tinggi (antara 10-20 kali), terakhir pembelian biasa (10 bulan-2 tahun yang lalu) namun memiliki jumlah pembelian banyak (40-80 Juta). Masih dalam Cluster 4 yaitu pelanggan dengan tingkat pembelian biasa (antara 5-10 kali), jarak pembelian pendek (3-10 bulan) dan jumlah pembelian banyak (40-80 juta). Masih pada cluster ini untuk jumlah pembelian normal (antara 20-40 juta) dan yang terakhir pelanggan dengan tingkat</i></p>

			<p><i>intensitas pembelian tinggi 10-20 kali, dengan jarak pembelian yang pendek (3 bulan-10 bulan) dan jumlah pembelian berada pada kisaran normal (20-40 juta). Selain itu pelanggan pada tipe <b>advocates</b> ini melakukan pembelian secara teratur hal tersebut sangat menguntungkan bagi perusahaan. Tipe pelanggan ini juga dapat dimanfaatkan sebagai market perusahaan karena pada tipe pelanggan ini melakukan pemasaran untuk perusahaan dengan cara mendorong teman-teman mereka yang lain agar membeli barang/ jasa pada perusahaan.</i></p>
<b>Client</b>	2	<ul style="list-style-type: none"> <li>- <i>Frequency</i> (biasa) AND <i>Recency</i> (biasa) AND <i>Monetary</i> (normal)</li> <li>- <i>Frequency</i> (biasa) AND <i>Recency</i> (biasa) AND</li> </ul>	<p><i>Pelanggan pada Cluster 2 diberi nama jenis pelanggan <b>Client</b>. memiliki karakteristik frekuensi pembelian biasa yaitu sebanyak 5-10 kali, kemudian kedatangannya</i></p>

		<p><i>Monetary</i> (banyak)</p> <p>- <i>Frequency</i> (biasa) AND <i>Recency</i> (pendek) AND <i>Monetary</i> (normal)</p>	<p>antara 10 bulan-2 tahun yang lalu, dan jumlah pembelian normal antara 20-40 juta, selain itu pembelian dalam jumlah banyak yaitu antara 40-80 juta juga masuk dalam cluster ini. mereka membeli secara teratur, hubungan dengan jenis pelanggan ini sudah kuat dan berlangsung lama yang membuat mereka tidak terpengaruh oleh tarikan pesaing produk lain.</p>
<b><i>Repeat Customer</i></b>	1	<p>- <i>Frequency</i> (rendah) AND <i>Recency</i> (pendek)</p> <p>- <i>Frequency</i> (biasa) AND <i>Recency</i> (pendek) AND <i>Monetary</i> (sedikit)</p> <p>- <i>Frequency</i> (biasa) AND <i>Recency</i> (sangat pendek) AND <i>Monetary</i> (normal)</p>	<p>Pelanggan pada Cluster 1 diberi nama <b><i>repeat customer</i></b> yaitu pelanggan yang telah melakukan pembelian suatu produk sebanyak 2 kali atau lebih. memiliki karakteristik frekuensi pembelian rendah (dalam range skala 2-5 kali) dalam jangka waktu yg pendek antara 3 sampai 10 bulan yang lalu Selain itu pelanggan dengan frekuensi pembelian biasa</p>

			<p>(5-10 kali) dan jumlah monetary sedikit (antara 500 ribu sampai 20 juta). Termasuk dalam cluster ini juga jenis pelanggan dengan frekuensi pembelian biasa (5-10 kali) dan recency sangat pendek (kuang dari 3 bulan) dengan monetary normal antara Rp20.000.000 sampai Rp. 40.000.000.</p>
<b><i>Firs Time Customer</i></b>	3	<ul style="list-style-type: none"> <li>- Frequency (rendah) AND Recency (biasa)</li> <li>- Frequency (sangat rendah) AND Recency (pendek)</li> </ul>	<p>Pelanggan pada Cluster 3 memiliki karakteristik tingkat intensitas yang masih rendah (yaitu antara 2-5 kali) dengan kedatangan antara 10 bulan-2 tahun, mereka adalah pelanggan baru yang tingka kedatangannya sangat rendah (dalam sekala kurang dari 2 kali). Pelanggan jenis ini tergolong pelanggan membeli untuk yang pertama kalinya, mereka masih menjadi pelanggan yang baru.</p>

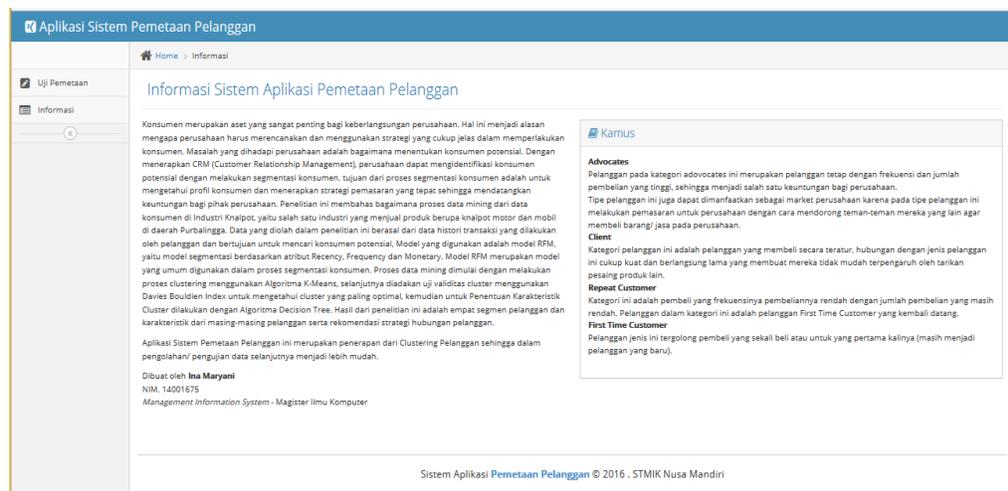
Dengan *rule* yang ada pada tabel IV.12. akan mempermudah perusahaan dalam melakukan penggolongan pelanggan di masa yang akan datang.

Setelah proses terbentuk tabel karakteristik pelanggan berikutnya dilakukan uji empiris terhadap 5 pelanggan baru yang dimiliki perusahaan pada tahun 2016 dan berhasil menempatkan kelima pelanggan baru ini dengan tepat, 3 berada pada segmen *First Time Customer* sementara 2 sisanya berada pada segmen *Repeat Customer*.

#### 4.2.5. Deployment

Berdasarkan proses clustering dan classification maka didapatkan hasil yang dijadikan sebagai *rule* untuk pembuatan interface yang dapat memudahkan perusahaan dalam memetakan pelanggannya untuk mengetahui profil dan karakteristik pelanggan. *Interface* yang digunakan dalam penelitian ini dibuat dengan menggunakan *PHP* dan *HTML*.

##### 1. Tampilan Halaman Depan



Gambar IV.9. Tampilan Halaman Depan

## 2. Tampilan Halaman Form Input Pemetaan

Aplikasi Sistem Pemetaan Pelanggan

Home > Uji Pemetaan

Uji Pemetaan Pelanggan

Form Uji

1 Input Variabel 2 Hasil 3 Saran Tindakan

Masukkan Variabel yang Dibutuhkan

Nama Pelanggan:

Alamat Pelanggan:

Tanggal Pembelian Terakhir adalah tanggal terakhir pelanggan melakukan transaksi pada periode uji

Frekuensi Pembelian adalah jumlah transaksi yang dilakukan dalam periode uji

Total Pembelian adalah Total Pemasukan (dalam rupiah) yang dikeluarkan oleh pelanggan dalam periode uji

Tanggal Pembelian Terakhir:

Frekuensi Pembelian:  kali

Total Pembelian:  rupiah

Tidak ada isian yang kosong

Previ Next

Sistem Aplikasi Pemetaan Pelanggan © 2016 . STMIK Nusa Mandiri

Gambar IV.10. Tampilan Halaman Form Input Pemetaan

## 3. Tampilan Halaman Setelah Data di Input

Aplikasi Sistem Pemetaan Pelanggan

Home > Uji Pemetaan

Uji Pemetaan Pelanggan

Form Uji

1 Input Variabel 2 Hasil 3 Saran Tindakan

Masukkan Variabel yang Dibutuhkan

Nama Pelanggan:

Alamat Pelanggan:

Tanggal Pembelian Terakhir adalah tanggal terakhir pelanggan melakukan transaksi pada periode uji

Frekuensi Pembelian adalah jumlah transaksi yang dilakukan dalam periode uji

Total Pembelian adalah Total Pemasukan (dalam rupiah) yang dikeluarkan oleh pelanggan dalam periode uji

Tanggal Pembelian Terakhir:

Frekuensi Pembelian:  kali

Total Pembelian:  rupiah

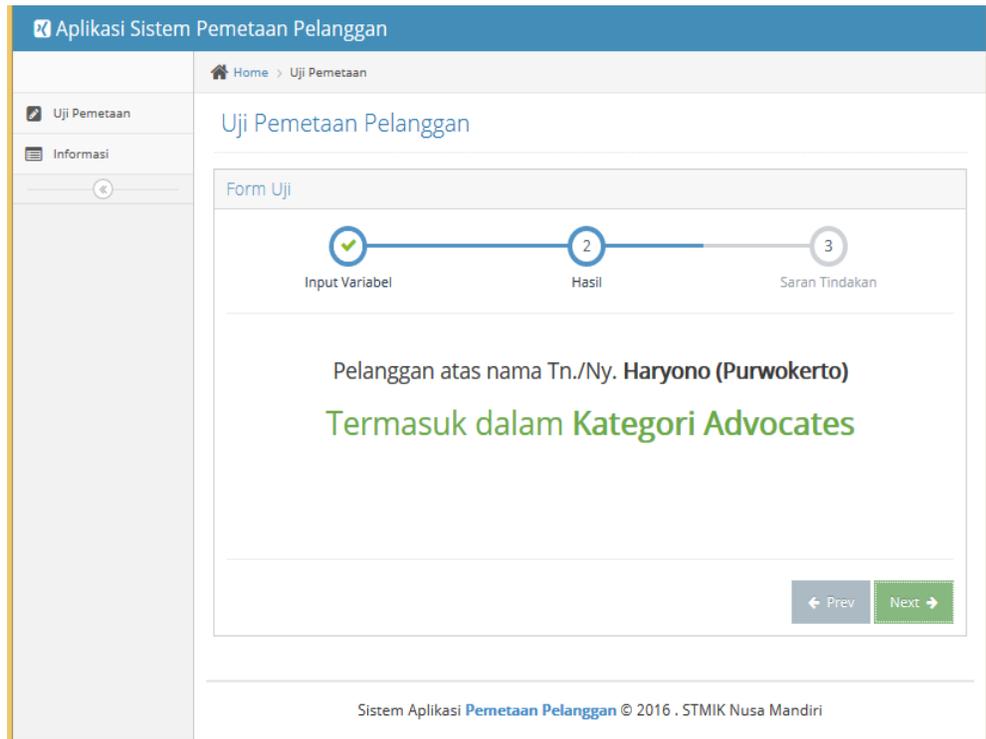
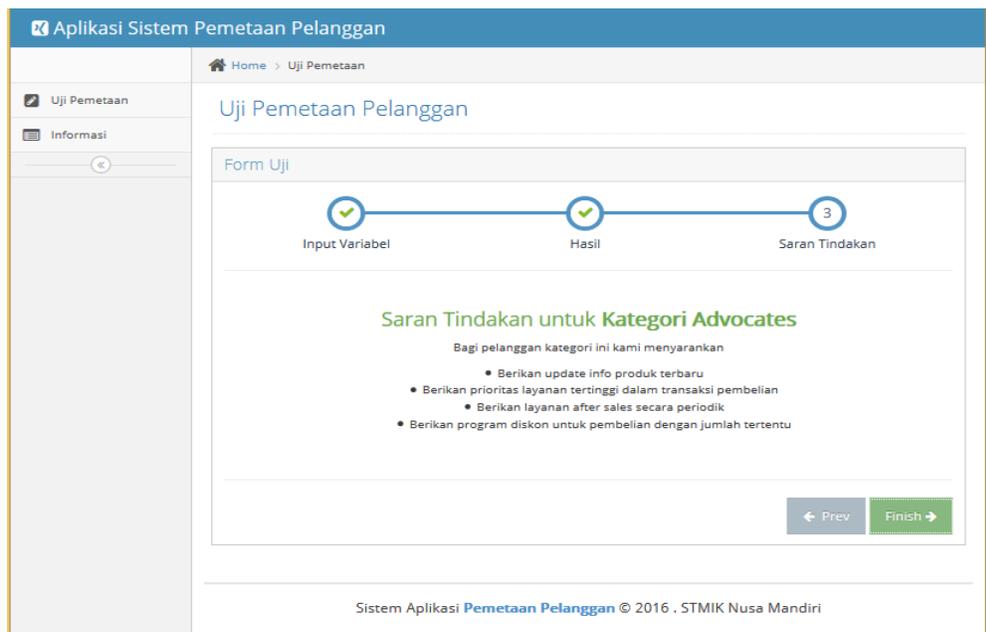
Tidak ada isian yang kosong

Previ Next

Sistem Aplikasi Pemetaan Pelanggan © 2016 . STMIK Nusa Mandiri

Gambar IV.11. Tampilan Halaman Setelah Data di Input

Program Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri

4. Tampilan Halaman *Output/Hasil*Gambar IV.12. Tampilan Halaman *Output/Hasil*5. Tampilan Halaman *Rekomendasi Strategi*Gambar IV.11. Tampilan Halaman *Rekomendasi Strategi*

### 4.3. Implikasi Penelitian

Setelah dilakukan proses data mining dengan model *Clustering* dan *Classification* untuk melakukan proses segmentasi pelanggan, penelitian ini dapat digunakan untuk mendukung proses pengambilan keputusan dan pengembangan sistem, hasil penelitian ini memiliki implikasi pada tiga aspek, yaitu aspek sistem, aspek manajerial, dan aspek penelitian lanjutan.

#### 4.3.1. Aspek Sistem

Hasil penelitian ini dapat diterapkan dengan adanya dukungan sistem yang berjalan dengan baik, sehingga perusahaan dapat menggunakan hasil penelitian ini untuk memetakan atau melakukan segmentasi pelanggan. Diperlukan sarana dan prasana yang memadai yang terdiri dari *Hardware*, *Software* (sistem operasi dan aplikasi yang dibuat pada tahap *deployment*), dan infrastruktur lainnya guna memberikan hasil yang terbaik, spesifikasi *hardware* dan *software* yang dapat digunakan dalam penelitian ini memiliki spesifikasi minimal seperti pada tabel IV.3. sebagai berikut.

Tabel IV.13 Spesifikasi *Hardware* dan *Software*

Spesifikasi Kebutuhan	Tipe
Prosesor	Core2
Memori	1 GB
Monitor	SVGA
Sistem Operasi	Windows
Aplikasi	Aplikasi Pemetaan Pelanggan

#### 4.3.2. Aspek Manajerial

Penelitian ini diharapkan dapat membantu pihak Perusahaan untuk mengetahui jenis dan karakteristik Pelanggan sehingga membantu dalam meningkatkan pelayanan

terhadap Pelanggan atau menerapkan *Customer Relationship Management*. Hal tersebut sangat membantu keberlangsungan perusahaan.

#### **4.3.3. Aspek Penelitian Lanjutan**

Penelitian lebih lanjut dapat dilakukan dengan cara seperti berikut:

1. Hasil penelitian dapat diterapkan pada industri Knalpot atau perusahaan lainnya.
2. Penambahan jumlah *variabel attribut* seperti *geografi*, *demografi* untuk *customer profiling* dan sarana pendukungnya.
3. Variasi tentang rentang waktu transaksi untuk data yang akan di mining
4. Untuk penentuan cluster optimal dapat digunakan pendekatan lainnya tidak hanya *Davies Bouldien Index*
5. Meskipun dengan algoritma *decission tree* sudah mencapai tingkat akurasi 98, 1595% perlu dilakukan perbandingan dengan algoritma lainnya agar dapat diketahui algoritma mana yang paling tepat.

## BAB V

### PENUTUP

#### 5.1. Kesimpulan

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

1. Berdasarkan hasil *clustering* dengan metode *K-Means* jumlah segmen pelanggan yang terbentuk adalah 4 *cluster*/segmen pelanggan dari keseluruhan jumlah pelanggan sebanyak 326 Pelanggan.
2. Cluster yang terbentuk sudah optimal, terbukti dengan nilai *Davies Bouldien Index 0,301*
3. *Cluster* yang terbentuk antara lain: 92 anggota berada pada *cluster 1*, 114 anggota berada pada *cluster 2*, 81 anggota berada pada *cluster 3* dan 39 anggota berada pada *cluster 4*. Selanjutnya hasil *clustering* akan diklasifikasi untuk proses segmentasi pelanggan.
4. Setelah dilakukan klasifikasi dengan metode *algorithm decision tree* terbentuk *rule* untuk menentukan jenis dan karakteristik Pelanggan dengan prediksi **98,1595% correctly classified instances** dan **1,8405% incorrectly classified instances**. Prediksi ini sangat baik karena hanya **1,8405%** menghasilkan prediksi yang salah.
5. Dari *rule* menggunakan *decision tree* diperoleh 4 jenis dan karakteristik Pelanggan, ini sebagai penunjang keputusan *manajemen* perusahaan untuk memetakan pelanggan dimasa yang akan datang.
6. Jenis pelanggan diambil berdasarkan teori menurut Hill(1996) yaitu, *Advocates, Client, Repeat Customer dan Firs Time Customer*
7. Rekomendasi Strategi untuk Customer Relationship Management antara lain:
  - a. Prioritas pelayanan kepada pelanggan yang termasuk dalam kategori *potential customer (Advocates dan Client)*
  - b. Program diskon atau potongan harga layanan bagi pelanggan yang bertransaksi dengan intensitas cukup tinggi atau rutin. (***Repeat Customer***)
  - c. Sebagai bahan acuan bagi perusahaan untuk mengembangkan pelayanan atau melihat celah peluang usaha.

## 5.2. Saran

Saran yang dapat dipertimbangkan untuk mengembangkan penelitian ini antara lain:

1. Penambahan jumlah *variabel attribut* seperti *geografi*, *demografi* untuk *customer profiling* dan sarana pendukungnya.
2. Variasi tentang rentang waktu transaksi untuk data yang akan di mining
3. Untuk penentuan cluster optimal dapat digunakan pendekatan lainnya tidak hanya *Davies Bouldien Index*
4. Meskipun dengan algoritma *decission tree* sudah mencapai tingkat akurasi 98, 1595% perlu dilakukan perbandingan dengan algoritma lainnya agar dapat diketahui algoritma mana yang paling tepat serta dicari *algoritma* yang paling tepat untuk mengetahui produk terlaris.
5. Perlu dibuatkan *database histori* pembelian pelanggan dan *database testing evaluation*

## DAFTAR REFERENSI

- Aggelis, Vasilis dan Christodoulakis (2005), "Customer Clustering using RFM Analysis", *Proceedings of the 9th WSEAS International Conference on Computers*. (ICCOMP'05).
- Atyanto, Dhanika Budhi., Erawan, Mahendrawathi dan Soelaiman, Rully (2011), "Improving Modern Port Competitiveness by using Data Mining to Obtain Strategic Decision", *Proceeding of International Engineering Symposium (IES)2011*, Kumamoto University, Jepang, hal. E.4.1-E.4.4.
- Bramer, M. (2007). *Principles of Data Mining*. London: Springer.
- Brown, S.A. 2000. *Customer Relationship Management : A Strategic*
- Cheng, Ching-Hsue dan Chen, You-Shyang (2009), "Classifying the segmentation of customer value via RFM model and RS Theory", *Expert Systems with Applications*, Vo.36, hal. 4176-4184
- Fadli, A. 2011. *Konsep Data Mining*. Artikel Populer IlmuKomputer.com.
- Farooqi, M. R. dan K. Raza. 2011. *A Comprehensive Study of CRM through Data Mining Techniques*. Paper presented at the National Conference; NCCIST-2011. September 09, 2011.
- Golsefid, Samira M., Ghazanfari, Mehdi dan Alizadeh, Somayeh (2007), "Customer Segmentation in Foreign Trade based on Clustering Algorithms, case study: Trade Promotion Organization of Iran", *International Journal of Computer, Information, and Systems Science, and Engineering*, Vol 1, No.3, hal. 175-181
- Gorunescu, F. (2011). *Data Mining Concepts, Models and Techniques*. Berlin: Springer.
- Gupta, G. dan H. Anggarwal. 2012. Improving Customer Relationship Management Using Data Mining. *International Journal of Machine Learning and Computing*. 2 (6), 874-877.
- Han, J., & Kamber, M. (2006). *Data Mining Concepts and Techniques Second Edition*. California: Morgan Kaufmann.
- Hill, Niegel, 1996,**"*Hand Book of Customer Satisfaction Measurement*", Gower. Publishing , England.
- Hui-Huang, H. (2006). *Advanced Data Mining Technologies in Bioinformatics*. United States of America: Idea Group Publishing.
- Kohavi, Ron., dan Parekh, Rajesh (2004). "Visualizing RFM Segmentation", *SIAM*.

- Kothari, C. R. (2004). *Research Methology Methods and Techniques*. India: New Age International Limited.
- Kusrini, & Luthfi, T. E. (2009). *Algoritma Data Mining*. Yogyakarta: Penerbit Andi.
- Larose, D. T. (2006). *Data Mining Methods and Models*. Canada: John Wiley & Sons, inc.
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook Second Edition*. New York: Springer.
- Mingoti, Sueli A. dan Lima, Joab O. (2006), “Comparing SOM neural network with Fuzzy c-means, K-means, and traditional hierachical clustering algorithms”, *European Journal of Operational Research*, Vol. 174, hal. 1742-1759.
- Niyagas, Waminee., Srivihok, Anongnart., dan Kristin, Sukumal (2006), “Clustering Ebanking Customer Using Data Mining And Marketing Segmentation”, *ECTI Transactions on Computer and Information Technology*, Vol.2, No.1, Hal. 6369.
- Ngai, E.W.T., Xiu, L., Chau, D.C.K., 2008. *Application of Data Mining Techniques in Customer Relationship Management : A Literature Review and Classification*. *Expert Systems with Applications* 36 (2009) 2592–2602.
- Olson, D. L., & Delen, D. (2008). *Advanced Data Mining Techniques*. Berlin: Springer.
- Swift, R.S. 2000. *Accelerating Customer Relationships – Using CRM and Relationship Technologies*, Prentice Hall Inc.
- Tama, B.A. 2009. *Implementasi Teknik Data Mining di dalam Konsep Customer Relationship Management (CRM)*. Konferensi Nasional Sistem dan Informatika 2009, Bali, November 14, 2009. KNS&I09-011.
- Todman, C. 2001. *Designing a Data Warehouse – Supporting Customer Relationship Management*, Hewlett Packard.
- Trappey, Charles V., Trappey, Amy J.C., Chang, Ai-Che., dan Huang, Ashley Y.L. (2009), “The Analysis of Customer Service Choices and Promotion Preferences using Hierarchical Clustering”, *Journal of the Chinese Institue of Industrial Engineer*, Vol.26, No.5, hal. 367-376

- Tsiptsis, K, Chorianopoulos, A. 2009. *Data Mining Techniques in CRM*. A John Wiley and Sons, Ltd., Publication. United Kingdom.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools and Techniques*. United States: Elsevier Inc
- Xu, M., Walton, J. 2005. *Gaining Customer Knowledge Through Analytical CRM*, *Industrial Management & Data System*, Vol.105 No.7. pp 955-977.
- Yanuarita, Gemala., Hendrawan, Rully Agus dan Prananta, Danu (2011), *Segmentasi Pelanggan Berbasis Rfm Dengan Menggunakan Algoritma K-Means Untuk Mendukung Strategi Pengelolaan Pelanggan Di PT Indosat Mega Media*, Jurusan Sistem Informasi, ITS, Surabaya.
- Zumstein, D. 2007. "Customer Performance Measurement : Analysis of the Benefit of a Fuzzy Classification Approach in Customer Relationship Management" (thesis). Switzerland : University of Fribourg.



SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER  
**PROGRAM PASCA SARJANA**  
**MAGISTER ILMU KOMPUTER**

NIM : 14001675  
Nama Mahasiswa : Ina Maryani  
Dosen Pembimbing : Dr. Dwiza Riana, S.Si, MM, M.Kom  
Konsentrasi : Management Information System  
Judul Tesis : Pemanfaatan Data Mining Pada Industri  
Knalpot Dalam Menerapkan Customer  
Relationship Management



No	Tanggal Bimbingan	Pokok Bahasan	Paraf Dosen Pembimbing
1	15 April 2016	Pengajuan Judul	
2	04 Mei 2016	BAB I	
3	10 Juni 2016	ACC BAB I dan Pengajuan BAB II	
4	30 Juni 2016	Studi Literature Review	
5	15 Juli 2016	Kerangka Pemikiran, Variabel, Atribut	
6	16 Juli 2016	Olah Data	
7	19 Juli 2016	BAB III	
8	20 Juli 2016	ACC BAB III	
9	05 Agustus 2016	BAB IV	
10	12 Agustus 2016	ACC Keseluruhan	

**Bimbingan Tesis**

- Dimulai pada tanggal : 15 April 2016
- Diakhiri pada tanggal : 12 Agustus 2016
- Jumlah pertemuan bimbingan : 10 Kali

Disetujui Oleh,  
Dosen Pembimbing

Dr. Dwiza Riana, S.Si, MM, M.Kom

## DAFTAR RIWAYAT HIDUP

### A. Biodata Mahasiswa

N.I.M : 14001675  
Nama Lengkap : Ina Maryani  
Tempat & Tanggal Lahir : Purbalingga, 01 Maret 1991  
Alamat Lengkap : Sidareja Rt 10 / Rw 04 Kecamatan Kaligondang,  
Kabupaten Purbalingga 53391

### B. Riwayat Pendidikan Formal dan Non-Formal

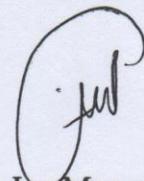
1. SD Negeri 2 Sidareja, lulus tahun 2003
2. SMP Negeri 2 Kaligondang, lulus tahun 2006
3. SMK Negeri 1 Purbalingga, lulus tahun 2009

### C. Riwayat Pengalaman Berorganisasi/Pekerjaan

1. Tenaga Administrasi SD Negeri 2 Sidareja 2009 sampai sekarang
2. Sekretaris Dewan Kesenian Kabupaten Purbalingga Periode 2013-2017
3. Instruktur AMIK BSI Purwokerto 2013 Sampai Sekarang

Purwokerto, 22 Agustus 2016

Hormat Saya,



Ina Maryani



**INDUSTRI KNALPOT MOTOR DAN MOBIL**

**DOCTOR SPEED PURBALINGGA**

*Jl. Kiswadi No.32 - Pesayangan Telp. 082226477333*

*Purbalingga 53257*

---

**SURAT KETERANGAN**

Nomor: 003/S.Ket/VI/2016

Yang bertandatangan dibawah ini:

Nama : FAHRI SAPUTRA  
Jabatan : Pemilik Industri Knalpot Motor dan Mobil Doctor  
Speed Purbalingga

Dengan ini menerangkan bahwa:

Nama : INA MARYANI  
Nim : 14001675

Adalah benar telah melaksanakan riset/penelitian di Industri knalpot motor dan mobil Doctor Speed Purbalingga sejak tanggal 15 Mei 2016 s/d 15 Juni 2016.

Demikian surat keterangan ini dibuat dengan sebenar-benarnya untuk dapat dipergunakan sebagaimana mestinya.

Purbalingga, 15 Juni 2016

**DOCTOR SPEED**  
  
**RACING PRODUCT**  
**FAHRI SAPUTRA**  
**PURBALINGGA**

