PROGRAM STUDI TEKNIK INFORMATIKA

PENERAPAN *DATA MINING* UNTUK MENENTUKAN KRITERIA CALON NASABAH POTENSIAL PADA AJB BUMIPUTERA 1912 PALEMBANG

M. KHOIRIL AMRI
09142239

Skripsi ini diajukan sebagai syarat memperoleh gelar Sarjana Komputer

FAKULTAS ILMU KOMPUTER
UNIVERSITAS BINA DARMA
2013

PENERAPAN *DATA MINING* UNTUK MENENTUKAN KRITERIA CALON NASABAH POTENSIAL PADA AJB BUMIPUTERA 1912 PALEMBANG

M. KHOIRIL AMRI
09142239

Skripsi ini diajukan sebagai syarat memperoleh gelar Sarjana Komputer

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BINA DARMA
2013

HALAMAN PENGESAHAN

PENERAPAN *DATA MINING* UNTUK MENENTUKAN KRITERIA CALON NASABAH POTENSIAL PADA AJB BUMIPUTERA 1912 PALEMBANG

M. KHOIRIL AMRI
09142239

Telah diterima sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer Pada Program Studi Teknik Informatika

Disetujui Oleh :

Palembang, September 2013
Program Studi Teknik Informatika
Fakultas Ilmu Komputer
Universitas Bina Darma Palembang

Dosen Pembimbing I **Dekan**

(PH. Saksono, S.T.,M.Sc.,Ph.D.) **(M. Izman Herdiansyah, S.T. MM, Ph.D.)**

Dosen Pembimbing II

(Eka Puji Agustina, S.Kom., M.M.)

HALAMAN PERSETUJUAN

Skripsi Berjudul **"PENERAPAN DA*TA* *MINING* UNTUK MENENTUKAN KRITERIA CALON NASABAH POTENSIAL PADA AJB BUMIPUTERA 1912 PALEMBANG"** Oleh **"M. Khoiril Amri (09142239)"** telah dipertahankan didepan komisi penguji pada hari Rabu, 31 Juli 2013.

Komisi Penguji

1. Ketua tim penguji PH. Saksono, S.T.,M.Sc.,Ph.D. (…….……)
2. Sekretaris tim penguji Eka Puji Agustina, S.Kom.,M.M. (…….……..…)
3. Anggota tim penguji Muhammad Nasir, S.Kom.,M.M. (.....……...….)
4. Anggota tim penguji Suyanto, S.Kom.,M.M. (…….……..…)

Mengetahui,
Program Studi Teknik Informatika
Fakultas Ilmu Komputer
Universitas Bina Darma
Ketua

(Syahril Rizal, S.T.,M.M.,M.Kom.)

PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : M. Khoiril Amri
Nim : 09142239

Dengan ini menyatakan bahwa:

1. Karya tulis Saya (skripsi) ini adalah asli dan belum pernah di ajukan untuk mendapatkan gelar akademik baik (sarjana) di Universitas Bina Darma atau perguruan tinggi lain;
2. Karya tulis ini murni gagasan, rumusan dan penelitian Saya sendiri dengan arahan tim pembimbing;
3. Di dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukkan ke dalam daftar rujukan;
4. Saya bersedia, skripsi yang Saya hasilkan dicek keasliannya menggunakan *plagiarism checker* serta diunggah ke internet, sehingga dapat diakses publik secara daring;
5. Surat pernyataan ini Saya tulis dengan sungguh-sungguh dan apabila terbukti melakukan penyimpanan atau ketidakbenaran dalam pernyataan ini, maka Saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini Saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, September 2013
Yang Membuat Pernyataan,

Materai

Rp. 6000,00

M. Khoiril Amri
NIM: 09142239

MOTTO DAN PERSEMBAHAN

Motto :

- *Sesungguhnya sesudah kesulitan itu ada kemudahan. Maka apabila kamu Telah selesai (dari sesuatu urusan), kerjakanlah dengan sungguh – sungguh (urusan) yang lain. Dan Hanya kepada Tuhanmulah hendaknya kamu berharap. (Q.S Alam Nasyrah : 6,7,8).*
- *Kesabaran dan ketekunan adalah kunci keberhasilan.*
- *Saya datang, saya pulang, saya bimbingan, saya ujian, saya revisi dan saya maju untuk menang !!*
- *Hidup memberi pilihan dan hidup juga memberi kesempatan, maka yang pantas untuk dilakukan adalah sebuah kepastian dalam memilih dan memanfaatkan kesempatan yang ada dengan sebaik-baiknya.*

Kami persembahkan kepada :

- *Kedua Orang Tua yang senantiasa mendoa'kan keberhasilan kami.*
- *Seluruh keluarga yang kami sayangi.*
- *Teman-teman kami di Universitas Bina Darma yang selalu bekerja sama dalam menuntut ilmu dan orang special yang selalu memberikan dukungan sampai skripsi ini selesai.*
- *Almameter kami yang senantiasa mendampingi kami selama kuliah.*

ABSTRAK

Sekarang ini bisnis asuransi semakin berkembang karena semakin tingginya kesadaran masyarakat untuk mengasuransikan dan memberikan perlindungan terhadap berbagai aspek kehidupannya. Perusahaan AJB Bumiputera 1912 adalah salah satu perusahaan jasa asuransi yang nasabahnya merupakan *client* atau *partner* kerja yang sangat penting, sehingga peningkatan kualitas pelayanan kepada para nasabah sangat diperhatikan. Kendala yang dihadapi saat ini yaitu perusahaan AJB Bumiputera 1912 mengalami kesulitan dalam menentukan nasabah potensial. Apabila perusahaan bisa mengidentifikasi tingkatan-tingkatan untuk menentukan nilai potensi nasabah maka pelayanan nasabah bisa lebih tepat. Adapun beberapa kriteria yang dianggap penting dalam menentukan nasabah potensial yaitu Penghasilan dan Umur. Untuk itu akan dikembangkan sebuah penerapan *data mining* yang berfungsi untuk menentukan kriteria nasabah. Teknik *data mining* yang diterapkan adalah Klasifikasi sedangkan metode klasifikasi yang digunakan adalah *Decision Tree* (pohon keputusan). Algoritma yang dipakai adalah algoritma C4.5 dan DTREG sebagai perangkat lunak untuk menghasilkan pohon keputusan.

Kata Kunci : Asuransi, *Data Mining*, *Decision Tree*, Algoritma C4.5

KATA PENGANTAR

بسم الله الرحمن الرحيم

Puji syukur kehadirat Allah SWT karena berkat rahmat dan karunia-Nya jualah, proposal penelitian ini dapat diselesaikan guna memenuhi salah satu syarat untuk diteruskan menjadi skripsi sebagai proses akhir dalam menyelesaikan pendidikan dibangku kuliah.

Dalam penulisan proposal ini, tentunya masih jauh dari sempurna. Hal ini dikarenakan keterbatasannya pengetahuan yang dimiliki. Oleh karena itu dalam rangka melengkapi kesempurnaan dari penulisan proposal ini diharapkan adanya saran dan kritik yang diberikan bersifat membangun.

Pada kesempatan yang baik ini, tak lupa penulis menghaturkan terima kasih kepada semua pihak yang telah memberikan bimbingan, pengarahan, nasehat dan pemikiran dalam penulisan skripsi ini, terutama kepada :

1. Prof. Ir. H. Bochari Rahman, M.Sc. selaku Rektor Universitas Bina Darma Palembang.
2. M. Izman Herdiansyah, S.T. MM, PhD selaku Dekan Fakultas Ilmu Komputer.
3. Syahril Rizal, S.T.,MM, M.Kom selaku Ketua Program Studi Teknik Informatika.
4. PH. Saksono, S.T.,M.Sc.,PhD selaku Pembimbing I yang telah memberikan bimbingan dan bantuannya.

5. Eka Puji Agustini, S.Kom.,MM selaku Pembimbing II yang telah memberikan bimbingan dan bantuannya.
6. Staf Karyawan dan dosen pengajar Universitas Bina Darma Palembang yang telah banyak memberikan ilmu pengetahuan dan bimbingan selama penulis menuntut ilmu di Universitas Bina Darma Palembang
7. Kedua orang tuaku tercinta yang selama ini telah membimbingku hingga aku dewasa dan keluargaku yang telah memberikan dorongan hingga aku tumbuh jadi orang yang berkarakter baik.
8. Teman-teman di Program Studi Teknik Informatika yang telah banyak membantu.

Palembang, September 2013

Penulis

DAFTAR ISI

Halaman

DAFTAR GAMBAR

Halaman

DAFTAR TABEL

Halaman

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi saat ini sangat berarti bagi semua kalangan masyarakat. Saat ini teknologi informasi telah menjadi salah satu kebutuhan dalam kehidupan sehari-hari. Pemanfaatan teknologi informasi terbukti dapat mempermudah kinerja manusia. Hal inilah yang menyebabkan teknologi informasi diterapkan dalam beragam bidang yang ada, tidak terkecuali dalam dunia bisnis.

Salah satu perusahaan asuransi jiwa di Indonesia adalah AJB Bumiputera 1912 yang sudah berpengalaman dalam perasuransian Indonesia. AJB Bumiputera 1912 memiliki tiga divisi jaringan operasional yaitu divisi asuransi jiwa perorangan atau individu, kelompok, dan syariah. Asuransi jiwa perorangan merupakan program proteksi yang diberikan oleh perusahaan untuk melindungi jiwa seseorang secara individu. Perusahaan berharap dengan adanya ketiga divisi tersebut dapat menambah pelayanan bagi masyarakat untuk mendapatkan perlindungan secara tak langsung.

Persaingan yang terjadi dalam dunia bisnis asuransi memaksa para pelakunya untuk selalu memikirkan strategi–strategi dan terobosan yang dapat menjamin kelangsungan dari bisnis asuransi yang mereka jalankan. Data bisnis dalam jumlah yang besar merupakan salah satu aset berharga yang dimiliki sebuah perusahaan asuransi. Sebagai salah satu perusahaan yang bergerak di bidang bisnis asuransi, AJB Bumiputera 1912 haruslah memikirkan strategi dalam pemasaran untuk mempertahankan nasabah lama dan menarik perhatian bagi calon nasabah baru. Jenis asuransi yang di tawarkan saat ini sangatlah bervariatif, seperti asuransi jiwa, kesehatan, dan pendidikan. Berdasarkan sumber media Tribunnews.com yang di *publish* pada 2 April 2013 menyebutkan bahwa hingga akhir tahun 2011, AJB Bumiputera sudah memiliki sebanyak 5,2 juta nasabah yang tersebar di seluruh Indonesia.

Hal ini melahirkan suatu kebutuhan terhadap teknologi yang dapat memanfaatkannya dalam menggali pengetahuan–pengetahuan baru, yang dapat membantu dalam penerapan strategi bisnis asuransi. Dengan memanfaatkan jumlah data yang sangat besar pihak perusahaan tentunya dapat menemukan beragam informasi. Salah satu informasi yang dapat dihasilkan yaitu berupa informasi mengenai kriteria nasabah terhadap jenis asuransi yang dipilihnya. Informasi yang dihasilkan sangat penting bagi suatu perusahaan asuransi, dimana dengan adanya informasi kriteria nasabah perusahaan asuransi dapat mengambil keputusan dalam menerapkan strategi yang tepat untuk menawarkan produk kepada calon nasabah berdasarkan kriteria nasabah yang dulu.

Data Mining merupakan teknologi yang sangat berguna untuk membantu perusahaan-perusahaan menemukan informasi yang sangat penting dari *gudang*

data mereka. *Data mining* meramalkan *trend* dan sifat-sifat perilaku bisnis yang sangat berguna untuk mendukung pengambilan keputusan penting. Analisis yang diotomatisasi yang dilakukan oleh *data mining* melebihi yang dilakukan oleh sistem pendukung keputusan tradisional yang sudah banyak digunakan. *Data Mining* mengeksplorasi basis data untuk menemukan pola-pola yang tersembunyi, mencari informasi pemrediksi yang mungkin saja terlupakan oleh para pelaku bisnis karena terletak di luar ekspektasi mereka. (Sentosa, 2002).

Berdasarkan latar belakang diatas, penulis mengambil kesimpulan untuk mengatasi masalah yang ada pada AJB Bumiputera 1912 Palembang perlu adanya **"Penerapan *Data Mining* Untuk Menentukan Kriteria Calon Nasabah Potensial pada AJB Bumiputera 1912 Palembang".**

1.2 Perumusan Masalah

Berdasarkan uraian latar belakang diatas, maka permasalahan yang akan dirumuskan dalam penelitian ini adalah "Bagaimana Menerapkan *Data Mining* untuk Menentukan Kriteria Calon Nasabah Potensial pada AJB Bumiputera 1912 Palembang?"

1.3 Batasan Masalah

Untuk menghindari pembahasan yang meluas, maka penulis hanya membatasi pembahasan permasalahan hanya pada:

a. Tahapan *Knowledge Discovery in Database (KDD)* diproses menggunakan *DBMS Microsoft Access 2007*

b. Penerapan *data mining* menggunakan teknik *Decision Tree* dan algoritma C4.5 mengikuti tahapan *Knowledge Discovery in Database (KDD).*

c. Menampilkan informasi profil kriteria data nasabah lama untuk menentukan calon nasabah potensial.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Tujuan dari penelitian ini adalah menerapkan *data mining* untuk menentukan calon nasabah potensial berdasarkan data nasabah lama berupa pola data *(data pattern)* yang terdapat pada *database* nasabah AJB Bumiputera 1912 Palembang.

1.4.2 Manfaat Penelitian

Adapun manfaat penelitian yang diambil penulis dalam penelitian ini adalah :

1. **Bagi AJB Bumiputera 1912 Palembang**

   Dengan adanya penerapan *data mining* untuk menentukan kriteria calon nasabah potensial pada AJB Bumiputera 1912 Palembang ini dapat membantu menyediakan pengetahuan dan informasi yang mendukung untuk pengambilan keputusan yang tepat dalam menentukan calon nasabah potensial berdasarkan data nasabah terdahulu.

2. **Bagi Penulis**

   Diharapkan dapat memberikan gambaran bahwa teknologi komputer dapat memberikan banyak keuntungan serta kemudahan khususnya bagi perusahaan dan bisnis lainnya dalam penyimpanan data, pengolahan data dan sistem prediksi yang akurat dalam pengambilan keputusan.

1.5 Metodologi Penelitian

1.5.1 Metode Penelitian

Dalam penelitian ini, penulis menggunakan metode deskriptif karena permasalahan yang sedang diteliti saat ini berdasarkan fakta-fakta yang ada mengenai data polis dan data nasabah pada AJB Bumiputera 1912 Palembang.

1.5.2 Metode Pengumpulan Data

Metode pengumpulan data yang tepat yaitu dengan mempertimbangkan penggunaannya berdasarkan jenis data dan sumbernya. Data yang objektif dan relevan dengan pokok permasalahan penelitian merupakan indikator keberhasilan penelitian. Pengumpulan data penelitian ini dilakukan dengan cara sebagai berikut :

1. *Observasi,* merupakan teknik pengumpulan data dengan cara mengadakan pengamatan secara langsung kepada objek penelitian mengenai data Nasabah pada AJB Bumiputera 1912 Palembang.
2. Wawancara, Merupakan metode pengumpulan data dengan cara mengadakan Tanya Jawab langsung kepada bagian pengolahan data, bagian IT dan bagian pemasaran pada AJB Bumiputera 1912 Palembang.

3. Studi Pustaka, Mengumpulkan data dengan mempelajari masalah yang berhubungan dengan objek yang diteliti serta bersumber dari buku- buku pedoman, literatur yang disusun oleh para ahli untuk melengkapi data yang diperlukan dalam penelitian.

1.5.3 Metode Analisis Data

Adapun untuk menganalisis data dalam penerapan *data mining* ini menggunakan tahapan *Knowledge Discovery in Databases (KDD)* yang terdiri dari beberapa tahapan, yaitu *data selection, preprocessing, transformation, data mining,* dan *evaluation*.

1.6 Sistematika Penulisan

Sistematika ini secara garis besar dapat memberikan gambaran isi, yang berupa susunan bab dari penelitian.

BAB I PENDAHULUAN

Pada bab ini penulis menguraikan latar belakang, perumusan masalah, batasan masalah, tujuan dan manfaat, metodologi penulisan laporan, serta sistematika penulis.

BAB II LANDASAN TEORI

Bab ini membahas tentang pengertian, istilah, dan teori–teori pendukung yang digunakan untuk menguraikan dan menjelaskan mengenai penerapan *data mining* yang dilakukan penulis.

BAB III OBJEK PENELITIAN

Bab ini membahas secara singkat mengenai Sejarah, Visi dan Misi, Profil dan Struktur Organisasi AJB Bumiputera 1912 Palembang.

BAB IV PROSES *DATA MINING*

Bab ini membahas tahapan awal dari proses *Knowledge Discovery in Databases (KDD)* untuk menganalisa data dengan menggunakan *DBMS Microsoft Access 2007* yang meliputi tahapan data *selection*, *preprocessing* dan *transformation* data kedalam bentuk data yang sesuai terhadap teknik dan algoritma yang digunakan.

BAB V HASIL DAN PEMBAHASAN

Bab ini membahas dan menjelaskan hasil dari proses *data mining* yang dilakukan dengan menguraikan teknik dan algoritma *data mining* yang digunakan dalam penelitian, serta menampilkan hasil *data mining* menggunakan *software data mining* DTREG.

BAB VI KESIMPULAN DAN SARAN

Pada bab terakhir ini penulis akan membuat dan mengambil kesimpulan dari pembahasan sebelumnya dan mencoba untuk mengutarakan saran yang mungkin dapat dijadikan bahan pertimbangan bagi AJB Bumiputera 1912 Palembang dalam pengambilan keputusan.

BAB II

TINJAUAN PUSTAKA

2.1 Data Mining

Data Mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data Mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Luthfi, 2009).

Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data (Pramudiono, 2003).

Kemajuan luar biasa yang terus berlanjut dalam bidang *data mining* didorong oleh beberapa faktor, antara lain (Larose, 2005) :

1. Pertumbuhan yang cepat dalam kumpulan data.
2. Penyimpanan data dalam data warehouse, sehingga seluruh perusahaan memiliki akses ke dalam database yang andal.
3. Adanya peningkatan akses data melalui navigasi web dan intranet.

4. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
5. Perkembangan teknologi perangkat lunak untuk *data mining.*
6. Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

2.2 Pengelompokan *Data Mining*

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Luthfi, 2009).:

1. Deskripsi, terkadang peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.
2. Estimasi, estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah *numeric* dari pada ke arah kataegori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai variabel target sebagai nilai prediksi.
3. Prediksi, prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang.
4. Klasifikasi, dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.
5. Pengklusteran, pengklusteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki

kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam kluster lain.

6. Asosiasi, asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

2.3 Klasifikasi

Klasifikasi adalah suatu *fungsionalitas data mining* yang akan menghasilkan model untuk memprediksi kelas atau kategori dari objek-objek di dalam basis data. Klasifikasi merupakan proses yang terdiri dari dua tahap, yaitu tahap pembelajaran dan tahap pengklasifikasian.

Pada tahap pembelajaran, sebuah algoritma klasifikasi akan membangun sebuah model klasifikasi dengan cara mengalisis training data. Tahap pembelajaran dapat juga dipandang sebagai tahap pembentukan fungsi atau pemetaan Y=F(X) diman Y adalah kelas hasil prediksi dan X adalah *tuple* yang ingin diprediksi kelasnya. Selanjutnya pada tahap pengklasifikasian, model yang telah dihasilkan akan digunakan untuk melakukan klasifikasi.

2.4 Pohon Keputusan *Decision Tree*

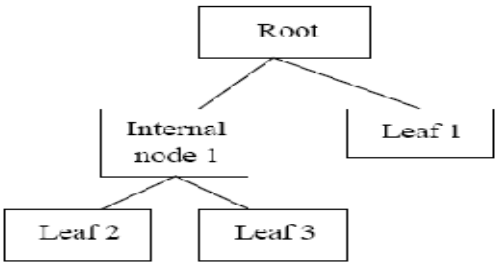
Pohon *(tree)* adalah sebuah struktur data yang terdiri dari simpul *(node)* dan rusuk *(edge)*. Simpul pada sebuah pohon dibedakan menjadi tiga, yaitu simpul akar *(root node)*, simpul percabangan/ internal *(branch/ internal node)* dan simpul daun *(leaf node)*, (Hermawati, 2013).

Pohon keputusan merupakan representasi sederhana dari teknik klasifikasi untuk sejumlah kelas berhingga, dimana simpul internal maupun simpul akar ditandai dengan nama atribut, rusuk-rusuknya diberi label nilai atribut yang mungkin dan simpul daun ditandai dengan kelas-kelas yang berbeda (Hermawati, 2013).

Gambar 2.1 Konsep Pohon Keputusan

Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule*. Manfaat utama dari penggunaan pohon keputusan adalah kemampuannya untuk *membreak down* proses pengambilan keputusan yang *kompleks* menjadi lebih simpel sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan. Pohon Keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target.

Pohon keputusan merupakan himpunan aturan *IF...THEN*. Setiap *path* dalam *tree* dihubungkan dengan sebuah aturan, di mana premis terdiri atas sekumpulan *node-node* yang ditemui, dan kesimpulan dari aturam terdiri atas kelas yang terhubung dengan *leaf* dari *path* (Wibowo, 2011).



Gambar 2.2 Konsep Dasar Pohon Keputusan

Bagian awal dari pohon keputusan ini adalah titik akar *(root)*, sedangkan setiap cabang dari pohon keputusan merupakan pembagian berdasarkan hasil uji, dan titik akhir *(leaf)* merupakan pembagian kelas yang dihasilkan.

Pohon keputusan banyak mengalami perkembangan, beberapa *algoritma* yang populer dan sering dipakai adalah ID3, C4.5 dan *CART*.

Tabel 2.1 Frekuensi Penggunaan Algoritma Pohon Keputusan

Algoritma Pohon Keputusan	**Frekuensi**
ID3	68 %
C4.5	54.55 %
CART	40.9 %
SPRINT	31.84 %
SLIQ	27.27 %
PUBLIC	13.6 %
C5.0	9 %
CLS	9 %
RANDOM FOREST	9 %
RANDOM *TREE*	4.5 %
ID3+	4.5 %
OCI	4.5 %
CLOUDS	4.5 %

2.5 Algoritma C4.5

Algoritma C4.5 adalah algoritma klasifikasi data dengan teknik pohon keputusan yang memiliki kelebihan-kelebihan. Kelebihan ini misalnya dapat mengolah data numerik *(kontinyu)* dan *diskret*, dapat menangani nilai atribut yang

hilang, menghasilkan aturan-aturan yang mudah diintrepetasikan dan tercepat diantara algoritma-algoritma yang lain (Luthfi. 2009).

Keakuratan prediksi yaitu kemampuan model untuk dapat memprediksi label kelas terhadap data baru atau yang belum diketahui sebelumnya dengan baik. Dalam hal kecepatan atau efisiensi waktu komputasi yang diperlukan untuk membuat dan menggunakan model. Kemampuan model untuk memprediksi dengan benar walaupun data ada nilai dari atribut yang hilang. Dan juga skalabilitas yaitu kemampuan untuk membangun model secara *efisien* untuk data berjumlah besar (aspek ini akan mendapatkan penekanan). Terakhir *interpretabilitas* yaitu model yang dihasilkan mudah dipahami.

Dalam algoritma C4.5 untuk membangun pohon keputusan hal pertama yang dilakukan yaitu memilih atribut sebagai akar. Kemudian dibuat cabang untuk tiap-tiap nilai didalam akar tersebut. Langkah berikutnya yaitu membagi kasus dalam cabang. Kemudian ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut dengan akar, didasarkan pada nilai *gain* tertinggi dari atribut-tribut yang ada. Untuk menghitung gain digunakan rumus sebagai berikut (Luthfi. 2009):

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy(S)$$

Keterangan:

S : Himpunan kasus

A : Atribut

N : Jumlah partisi atribut A

| Si | : Jumlah kasus pada partisi ke-i

| S | : Jumlah kasus dalam S

Sehingga akan diperoleh nilai gain dari atribut yang paling tertinggi. Gain adalah salah satu *atribute selection measure* yang digunakan untuk memilih *test atribute* tiap *node* pada *tree*. Atribut dengan *information gain* tertinggi dipilih sebagai *test atribute* dari suatu *node*.

Sementara itu, penghitungan nilai entropi dapat dilihat pada persamaan :

$$Entropy(S) = \sum_{i=1}^{n} -pi * log_2 pi$$

Keterangan :

S : Himpunan kasus

A : Atribut

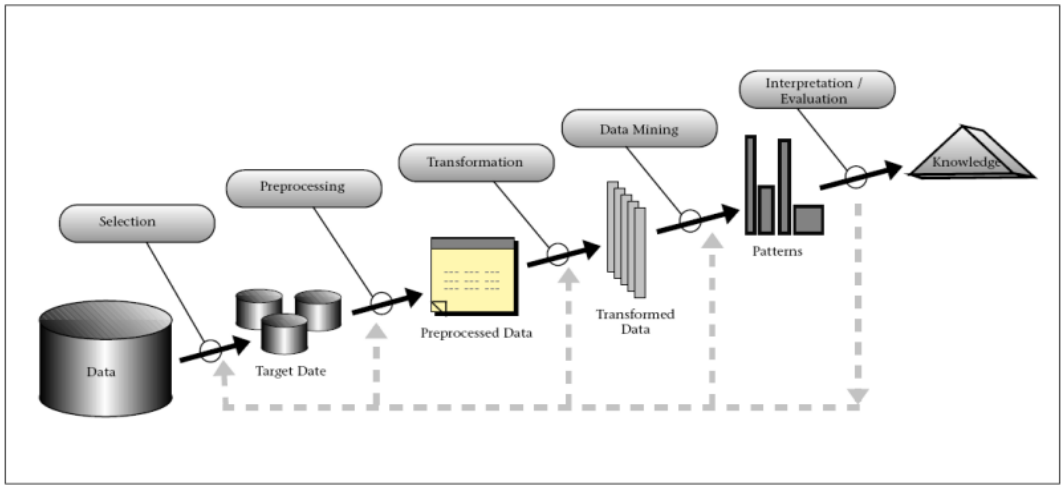
N : Jumlah partisi S

Pi : Proporsi dari Si terhadap S

2.6 Metode *Knowledge Discovery in Databases* (*KDD)*

Proses KDD adalah proses menggunakan metode *data mining* untuk mengekstrak pengetahuan apa yang dianggap sesuai dengan spesifikasi ukuran dan batas, menggunakan database bersama dengan *preprocessing* yang diperlukan, pengambilan sampel dan transformasi dari database (Azevedo, 2008).

Istilah *data mining* dan *knowledge discovery in databases* (KDD) seringkali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah *data mining.* Proses KDD secara garis besar dapat dijelaskan sebagai berikut (Kusrini, 2009).



Gambar 2.3 Tahapan *Knowledge Discovery in Databases*

1. *Data Selection,* pemilihan (seleksi) 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,* sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahn cetak.

3. *Transformation, coding* adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining.* Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada sejenis atau pola informasi yang akan dicari dalam basis data.
4. *Data Mining, data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.
5. *Interpretation/Evaluation,* pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation.* Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.7 DTREG

DTREG adalah *software* analisis statistik yang menghasilkan klasifikasi dan pohon keputusan bahwa model regresi data dapat digunakan untuk memprediksi nilai. *Single-pohon, pohon TreeBoost* dan Model Keputusan dapat dibuat dalam waktu yang singkat (Eka, 2010).

DTREG juga dapat melakukan *analisis deret* waktu dan peramalan. DTREG mencakup Korelasi, Analisis Faktor dan Analisis Komponen Dasar Proses penggalian informasi yang berguna dari satu set nilai data disebut sebagai

"data mining". Data ini dapat digunakan untuk membuat model untuk membuat prediksi teknik.

Banyak telah dikembangkan untuk pemodelan prediktif, dan ada seni untuk memilih dan menerapkan metode terbaik untuk situasi tertentu. DTREG menerapkan metode pemodelan prediksi paling kuat yang telah dikembangkan seperti penggunaaan pohon keputusan berdasarkan *metode TreeBoost* dan Pohon Keputusan Hutan serta *Neural Networks, Support Vector Machine, Gene Expression Programming, simbolis Regresi, K-Means Clustering, Linear Diskriminan Analisis, Regresi Linier model* dan *Regresi Logistik model*.

Tingkatan yang dicapai telah menunjukkan metode ini sangat efektif untuk menganalisa dan banyak jenis pemodelan data. DTREG juga merupakan aplikasi yang diinstal dengan mudah pada sistem *Windows*. DTREG dapat membaca file data dengan format *Comma Separated Value* (CSV). Setelah membuat file data, inputkan data tersebut ke DTREG, dan biarkan DTREG yang melakukan semua pekerjaan membuat pohon keputusan ke ukuran yang optimal. Bahkan analisis kompleks dapat diatur dalam beberapa menit. DTREG dapat membangun Klasifikasi Pohon di mana variabel target yang diperkirakan adalah kategoris dan Regresi Pohon adalah variabel *kontinu* seperti target pendapatan atau volume penjualan. Hanya dengan beberapa langkah saja, DTREG dapat membangun sebuah model tunggal-pohon klasik, model *TreeBoost* terdiri dari serangkaian pohon atau model Hutan Pohon Keputusan. DTREG dapat menampilkan pohon keputusan yang dihasilkan pada layar, kemudian dijadikan format jpg atau png file disk dan mencetaknya.

BAB III

OBJEK PENELITIAN

3.1 AJB Bumiputera 1912

AJB Bumiputera 1912 merupakan perusahaan asuransi jiwa nasional yang pertama dan tertua di Indonesia. Perusahaan asuransi ini mulai terbentuk pada tanggal ***12 Februari 1912, di Magelang, Jawa Tengah***, dengan nama ***Onderlinge Levensverzekering Maatschapij PGHB*** (bahasa Belanda) disingkat dengan ***O.L Mij. PGHB*** atau lebih dikenal dengan bahasa Inggrisnya ***Mutual Life Insurance*** (Asuransi Jiwa Bersama). Dengan badan usaha yang seperti ini, maka seorang pemilik perusahaan adalah Para Pemegang Polis.

O.L Mij PGHB didirikan berdasarkan keputusan dalam sidang pada Kongres Perserikatan Guru-guru Hindia Belanda yang pertama di Magelang, saat itu pesertanya hanya terbatas pada kalangan guru-guru saja. Para peserta kongres pun menyambut positif. Jumlah peserta yang terdaftar sebagai anggota O.L Mij. PGHB, baru 5 orang.

Karena perusahaan ini dibentuk oleh para guru, maka pengurusnya pun untuk pertama kali, hanya terdiri dari tiga orang Pengurus PGHB, yang terdiri dari:

1. ***Mas Ngabehi (M.Ng) Dwidjosewojo, sebagai Presiden Komisaris.***

2. ***Mas Karto Hadi (M.K.H) Soebroto, sebagai Direktur.***

3. ***Mas Maryoto Soedibyo (M.) Soebroto, sebagai Bendahara***

Pada mulanya perusahaan hanya melayani para guru sekolah Hindia Belanda, kemudian perusahaan memperluas jaringan pelayanannya ke masyarakat umum. Dengan bertambahnya anggota, maka para pengurus sepakat untuk mengubah nama perusahaannya. Berdasarkan Rapat Anggota/Pemegang Polis di Semarang, November 1914, nama O.L Mij. PGHB diubah menjadi ***O.L Mij. Boemi Poetra***.

Pada tahun 1942 ketika Jepang berada di Indonesia, nama O.L Mij. Boemi Poetra yang menggunakan bahasa asing segera diganti. Maka pada tahun 1943 O.L Mij. Boemi Poetra kembali diubah namanya menjadi Perseroan Pertanggungan Djiwa (PTD) Boemi Poetra, yang merupakan satu-satunya perusahaan asuransi jiwa nasional yg tetap bertahan. Namun karena dirasa kurang memiliki rasa kebersamaan, maka pada tahun 1953 PTD Boemi Poetra dihapuskan. Dan, hingga sekarang terkenal dengan nama Asuransi Jiwa Bersama (AJB) di depan nama Bumiputera 1912 yang merupakan bentuk badan hukum.

Pada tahun 1921, perusahaan pindah ke Yogyakarta. Pada tahun 1934 perusahaan melebarkan sayapnya dengan membuka cabang-cabang di Bandung, Jakarta, Surabaya, Palembang, Medan, Pontianak, Banjarmasin, dan Ujung Padang. Dengan demikian semakin berkembang, maka tahun 1958 secara bertahap kantor pusat dipindahkan ke Jakarta, dan pada tahun 1959 secara resmi kantor pusat AJB Bumiputera berdomisili di Jakarta.

Selama lebih Sembilan dasawarsa, Bumiputera telah berhasil melewati berbagai rintangan yang amat sulit, antara lain pada masa penjajahan, masa revolusi, dan masa-masa krisis ekonomi seperti sanering di tahun 1965 dan krisis moneter yang dimulai pada pertengahan tahun 1997.

Salah satu kekuatan Bumiputera adalah kepemilikan dan bentuk perusahaannya yang unik, dimana Bumiputera adalah satu-satunya perusahaan di Indonesia yang berbentuk mutual atau usaha bersama, artinya pemilik perusahaan adalah pemegang polis bukan pemegang saham. Jadi perusahaan tidak berbentuk PT atau Koperasi. Hal ini dikarenakan premi yang diberikan kepada perusahaan sekaligus dianggap modal. Badan perwakilan pera pemegang polis ikut serta menentukan garis-garis besar haluan perusahaan, memilih dan mengangkat direksi, dan ikut serta mengawasi jalannya perusahaan.

AJB Bumiputera 1912 memulai usahanya dengan modal awal nol sen. Dengan demikian, perusahaan asuransi ini berbentuk onderling atau mutual (Usaha Bersama), karena perusahaan dapat didirikan tanpa harus menyediakan modal lebih dahulu. Uang yang diterima perusahaan untuk pertama kalinya berasal dari kelima peserta kongres PGHB yang menjadi O.L Mij. PGHB. Syarat utamanya dalah bahwa ganti rugi tidak akan diberikan kepada ahli waris pemegang polis yang meninggal sebelum polisnya berjalan selama tiga tahun penuh. Perusahaan ini hanya mengutamakan pembayaran premi sebagai modal kerjanya dan tidak mendapatkan honorarium bagi para pengurusnya, sehingga mereka bekerja dengan sukarela.

3.2 Visi dan Misi

Visi

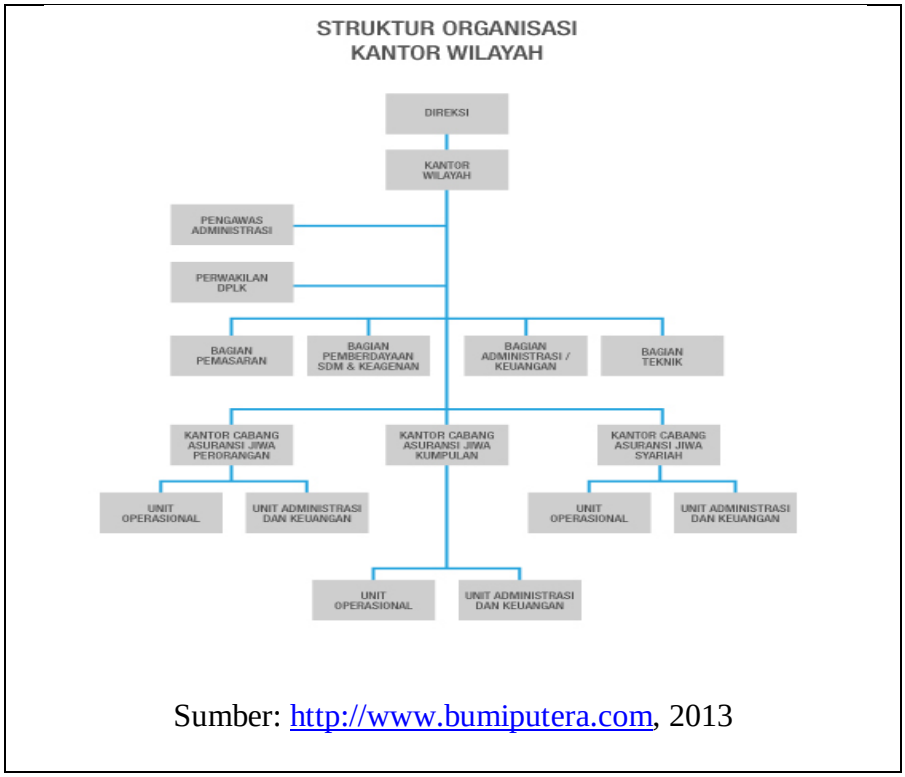
1. Menjadikan AJB Bumiputera 1912 sebagai Perusahaan Asuransi Jiwa Nasional yang kuat, modern dan menguntungkan.
2. Didukung oleh sumber daya manusia (SDM) profesional yang menjunjung tinggi nilai -nilai idealisme serta mutualisme.

Misi

1. AJB Bumiputera 1912 menyediakan pelayanan dan produk jasa asuransi jiwa berkualitas se bagai wujud partisipasi dalam pembangunan nasional melalui peningkatan kesejahteraan masyarakat Indonesia.
2. AJB Bumiputera 1912 senantiasa mengadakan pendidikan dan pelatihan untuk menjamin pertumbuhan kompetensi karyawan, peningkatan kesejahteraan, dalam rangka peningkatan kualitas pelayanan perusahaan kepada pemegang polis.
3. AJB Bumiputera 1912 mendorong terciptanya iklim kerja yang motivasif dan inovatif untuk mendukung proses bisnis internal perusahaan yang efektif dan efisien.

3.3 Struktur Organisasi Kantor Wilayah AJB Bumiputera 1912

Struktur organisasi merupakan gambaran mengenai fungsional unit kerja dalam suatu organisasi. Berikut ini merupakan struktur organisasi yang berada pada setiap kantor wilayah AJB Bumiputera 1912 dapat dilihat pada gambar 3.1.



Sumber: http://www.bumiputera.com, 2013

Gambar 3.1 Struktur Organisasi AJB Bumiputera 1912 Palembang

3.4 Profile AJB Bumiputera 1912

AJB Bumiputera 1912 adalah perusahaan asuransi terkemuka di Indonesia.

Didirikan seabad yang lalu untuk memenuhikebutuhan spesifik masyarakat Indonesia. AJB Bumiputera 1912 telah berkembang untuk mengikuti perubahan kebutuhan masyarakat. Pendekatan modern, produk yang beragam, serta teknologi mutakhir yang ditawarkan didukung oleh nlai-nilai tradisional yang melandasi pendirian AB Bumiputera 1912.

AJB Bumiputera 1912 telah merintis industri asuransi jiwa di Indonesia dan hingga saat ini tetap menjadi perusahaan asuransi jiwa nasional terbesar di Indonesia.

AJB Bumiputera 1912 menyadari pentingnya hubungan personal antara nasbah dan penasehat finansial mereka, serta menyediakan akses yang mudah untuk mendapatkan solusi khusus yang mudah untuk memenuhi semua kebutuhan asuransi nasabah.

AJB Bumiputera 1912 dimiliki oleh masyarakat indonesia dari berbagai latar belakang dan kelompok umur serta menyediakan berbagai produk dan layanan yang setara dengan produk asuransi terbaik dunia, namun tetap menjaga keuntungannya di Indonesia bagi para pemegang polisnya.

AJB Bumiputera 1912 adalah aset nasional pelopor asuransi di Indonesia.

3.5 Gambaran Umum Pendataan Nasabah Kantor Wilayah AJB Bumiputera 1912 Palembang

Data yang dimiliki AJB Bumiputera 1912 Palembang disimpan dalam dua jenis data yaitu data yang berupa berkas lembaran-lembaran dan data yang telah disimpan secara komputerisasi kedalam *Database Managemen Sistem (DBMS).* Pada tahapan awal pendataan data nasabah menggunakan aplikasi pendataan nasabah yang digunakan agen dalam mendata calon nasabah dimana data tersimpan dalam *server* lokal dan bersifat sementara. Selanjutnya setelah penerbitan polis *database* disimpan menggunakan aplikasi pemasaran yang digunakan bagian pemasaran dimana data disimpan dalam *server* pusat.

BAB IV

PROSES *DATA MINING*

Data yang akan di-*mining* diproses melalui tahapan *knowledge discovery in databases (KDD)* dengan menggunakan *DBMS Microsoft Access 2007,* berikut tahapan-tahapan KDD :

4.1 Data Selection

Data yang digunakan dalam penelitian ini berasal dari perusahaan AJB Bumiputera 1912 Palembang yaitu data transaksi wilayah palembang pada tahun 2010, terdiri dari beberapa tabel antara lain tabel nasabah dan tabel polis. Tabel nasabah berisi tentang informasi data nasabah dan tabel polis berisi tentang informasi data polis terbit. Jumlah *dataset* asli pada data polis sebanyak 1621 *record* atau selama 12 bulan. Dari semua atribut yang ada pada tabel nasabah dan polis terdapat 4 atribut yang akan digunakan dalam proses *knowledge discovery in databases (KDD)*. Atribut tersebut yaitu:

1. NOPOLIS merupakan atribut yang terdapat pada tabel nasabah dan tabel polis yang berperan sebagai *primary key* dalam menghubungkan tabel nasabah dan tabel polis.
2. TGLLAHIR merupakan atribut yang terdapat pada tabel nasabah yang berisi informasi mengenai tanggal lahir nasabah. Atribut ini digunakan untuk

menghitung umur yang nantinya akan digunakan untuk menentukan kriteria nasabah.

3. INCOME merupakan atribut yang terdapat pada tabel nasabah yang berisi informasi mengenai penghasilan nasabah. Atribut ini juga digunakan untuk menentukan kriteria nasabah.
4. PLAN merupakan atribut yang terdapat pada tabel polis yang berisi mengenai jenis asuransi yang dipilih oleh nasabah. PLAN merupakan kode dari jenis asuransi yang digunakan perusahaan dalam mengidentifikai jenis asuransi mereka.

4.2 Preprocessing

Pada tahapan *preprocessing* ini akan dilakukan proses integrasi data untuk menghubungkan tabel nasabah dan tabel polis, selanjutnya dilakukan proses data *cleaning* untuk menghasilkan *dataset* yang bersih sehingga dapat digunakan dalam tahap berikutnya yaitu *mining* dengan tujuan memperoleh pola mengenai kriteria nasabah untuk menentukan kriteria nasabah baru. Berikut merupakan penjelasan dari kedua proses di atas:

1. Integrasi Data, tahap integrasi data adalah tahap penggabungan data dari beberapa sumber. Pada tahapan ini dilakukan penggabungan dua tabel yaitu tabel polis dan nasabah. Proses penggabungan dilakukan dengan merelasikan tabel polis dan nasabah dengan *query* seperti yang terlihat pada gambar 4.1.

```
SELECT polis.NOPOLIS, nasabah.TGLLAHIR, nasabah.INCOME, polis.PLAN
FROM nasabah LEFT JOIN polis ON nasabah.NOPOLIS = polis.NOPOLIS;
```

Gambar 4.1 *Query* Integrasi Data

Query pada gambar 4.1 merupakan *query join* yang digunakan pada *DBMS Microsoft Access 2007*. "Select nasabah.NOPOLIS, nasabah.TGLLAHIR, nasabah.INCOME, polis.PLAN" menjelaskan bahwa atribut yang ditampilkan yaitu NOPOLIS, TGLLAHIR, INCOME dari tabel nasabah dan PLAN dari tabel polis. Selanjutnya "from nasabah left join polis nasabah.NOPOLIS = polis.NOPOLIS" menjelaskan bahwa atribut tersebut diambil dari penggabungan tabel polis dan tabel nasabah dimana atribut penghubungnya adalah NOPOLIS. Dari proses tersebut di dapatlah sebuah tabel baru yang saya beri nama tabel integrasi seperti gambar 4.2.

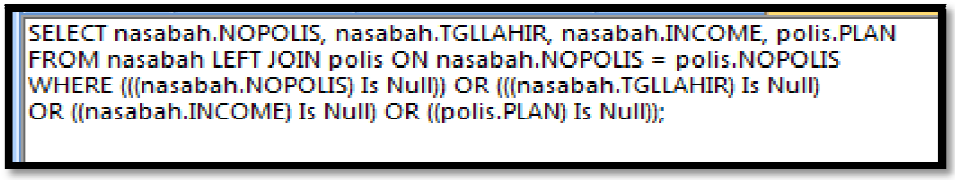
polis | nasabah | Relationships | Integrasi

NOPOLIS	TGLLAHIR	INCOME	PLAN
20100001	13/08/1990	<5jt	AG55
20100002	09/08/1990	<5jt	AG57
20100003	19/07/1990	<5jt	AG57
20100004	13/10/1989	<5jt	AG57
20100005	25/09/1989	<5jt	AG57
20100006	15/09/1989	<5jt	AG57
20100007	09/09/1989	<5jt	AG55
20100008	05/04/1989	<5jt	AG57
20100009	15/11/1988	<5jt	AG57
20100010	28/03/1988	<5jt	AG55
20100011	15/02/1987	<5jt	AG55
20100012	25/12/1986	<5jt	AG55
20100013	06/12/1986	<5jt	AG31
20100014	25/11/1986	<5jt	AG57
20100015	14/04/1986	<5jt	AG57
20100016	07/10/1985	<5jt	AG57
20100017	17/07/1985	<5jt	AG55
20100018	23/06/1984	<5jt	AG55
20100019	17/04/1984	<5jt	AG57
20100020	05/03/1984	<5jt	AG31
20100021	20/12/1983	<5jt	AG55
20100022	20/12/1983	<5jt	AG55
20100023	08/04/1983	<5jt	AG57

Gambar 4.2 Sebagian *dataset* hasil integrasi data

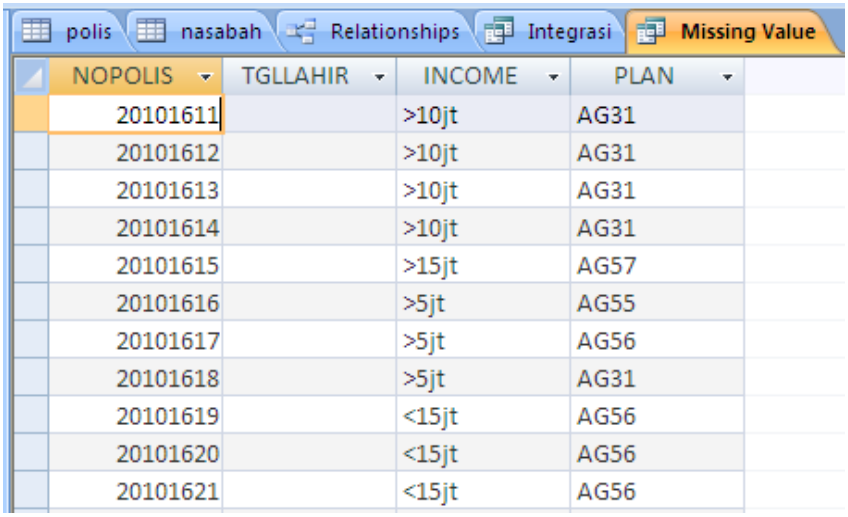
2. Data *Cleaning,* tahap data *cleaning* merupakan tahap awal dari proses KDD. Pada tahapan ini data yang yang tidak *relevan*, *missing value*, dan *redundant* harus di bersihkan. Hal ini dikarenakan data yang *relevan*, tidak *missing value*, dan tidak *redundant* merupakan syarat awal dalam melakukan *data mining*. Suatu data dikatakan *missing value* jika terdapat atribut dalam *dataset* yang tidak berisi nilai atau kosong, sedangkan data dikatakan *redundant* jika dalam satu *dataset* terdapat lebih dari satu *record* yang berisi nilai yang sama.

   Untuk menampilkan data *missing value* dapat menggunakan *query* seperti yang terlihat pada gambar 4.3.



Gambar 4.3 *Query* menampilkan data *missing value*

Query pada gambar 4.3 hampir sama dengan *query* pada gambar 4.1 hanya saja *query* pada gambar 4.3 ditambahkan perintah *WHERE* untuk melakukan *filter* terhadap data yang akan di tampilkan yaitu "WHERE (((nasabah.NOPOLIS) Is Null)) OR (((nasabah.TGLLAHIR) Is Null) OR ((nasabah.INCOME) Is Null) OR ((polis.PLAN) Is Null)" yang artinya data yang akan ditampilkan adalah data yang dimana kondisi salah satu dari atribut NOPOLIS,TGLLAHIR,INCOME, atau PLAN terdapat data yang *NULL* atau kosong. Dari *query* tersebut didapatlah hasil seperti pada gambar 4.4.



polis | nasabah | Relationships | Integrasi | Missing Value

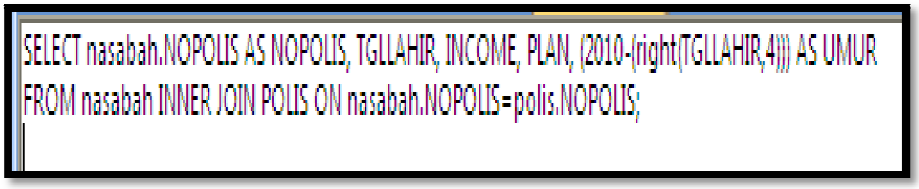
NOPOLIS	TGLLAHIR	INCOME	PLAN
20101611		>10jt	AG31
20101612		>10jt	AG31
20101613		>10jt	AG31
20101614		>10jt	AG31
20101615		>15jt	AG57
20101616		>5jt	AG55
20101617		>5jt	AG56
20101618		>5jt	AG31
20101619		<15jt	AG56
20101620		<15jt	AG56
20101621		<15jt	AG56

Gambar 4.4 Data *Missing Value*

Pada gambar 4.4 ditemukan 11 record yang memiliki data kosong sehingga data tersebut harus dibersihkan dengan cara dihapus. Sehingga jumlah *dataset* yang awalnya 1621 menjadi 1610.

4.3 Transformation

Tahapan *transformation* data merupakan tahap merubah data ke dalam bentuk yang sesuai untuk di-*mining*. Perubahan awal yang dilakukan yaitu menambahkan 1 atribut baru yaitu atribut UMUR. Atribut UMUR merupakan atribut yang berisi umur nasabah yang diperoleh dari perhitungan TGLLAHIR nasabah. Atribut UMUR dibuat untuk mempermudah dalam melakukan perhitungan dibandingkan TGLLAHIR yang memiliki tipe *date*. *Query* untuk mengisi atribut umur dapat dilihat pada gambar 4.5.



```
SELECT nasabah.NOPOLIS AS NOPOLIS, TGLLAHIR, INCOME, PLAN, (2010-(right(TGLLAHIR,4))) AS UMUR
FROM nasabah INNER JOIN POLIS ON nasabah.NOPOLIS=polis.NOPOLIS;
```

Gambar 4.5 *Query* menghitung umur

Atribut UMUR disini diisi dengan cara perhitungan 2010 sebagai tahun terbit polis – (right(TGLLAHIR,4)) yaitu 4 karakter dari kanan pada atribut TGLLAHIR. Jadi misalnya TGLLAHIR=13/08/1990 yang merupakan 4 karakter dari kanan yaitu 1990 maka dilakukan perhitungan 2010-1990=20, sehingga diperoleh 20 yang merupakan umur nasabah seperti gambar 4.6.

NOPOLIS	TGLLAHIR	INCOME	PLAN	UMUR
20100001	13/08/1990	<5jt	AG55	20
20100002	09/08/1990	<5jt	AG57	20
20100003	19/07/1990	<5jt	AG57	20
20100004	13/10/1989	<5jt	AG57	21
20100005	25/09/1989	<5jt	AG57	21
20100006	15/09/1989	<5jt	AG57	21
20100007	09/09/1989	<5jt	AG55	21
20100008	05/04/1989	<5jt	AG57	21
20100009	15/11/1988	<5jt	AG57	22
20100010	28/03/1988	<5jt	AG55	22
20100011	15/02/1987	<5jt	AG55	23
20100012	25/12/1986	<5jt	AG55	24
20100013	06/12/1986	<5jt	AG31	24
20100014	25/11/1986	<5jt	AG57	24
20100015	14/04/1986	<5jt	AG57	24
20100016	07/10/1985	<5jt	AG57	25
20100017	17/07/1985	<5jt	AG55	25
20100018	23/06/1984	<5jt	AG55	26
20100019	17/04/1984	<5jt	AG57	26
20100020	05/03/1984	<5jt	AG31	26
20100021	20/12/1983	<5jt	AG55	27
20100022	20/12/1983	<5jt	AG55	27
20100023	08/04/1983	<5jt	AG57	27

Gambar 4.6 hasil perhitungan umur

Dari hasil yang didapat maka proses selanjutnya akan dilakukan *sorting* untuk mengelompokkan data berdasarkan UMUR dan INCOME agar mempermudah mengetahui tingkatan UMUR dengan jumlah INCOME sebagai kriteria nasabah.

Dari penjelasan yang diuraikan di atas maka diperoleh hasilnya seperti pada gambar 4.7.

NOPOLIS	UMUR	INCOME	PLAN
20100669	17	<5jt	AG57
20100102	18	<5jt	AG31
20101373	18	<5jt	AG57
20100983	19	<5jt	AG57
20100531	20	<5jt	AG57
20101196	20	<5jt	AG57
20100324	20	<5jt	AG57
20101293	20	<5jt	AG57
20101199	20	<5jt	AG57
20100858	20	<5jt	AG57
20101060	20	<5jt	AG57
20101300	20	<5jt	AG31
20100906	20	<5jt	AG55
20100583	20	<5jt	AG56
20101334	20	<5jt	AG57
20101445	20	<5jt	AG57
20101360	20	<5jt	AG57
20101124	20	<5jt	AG55
20101299	20	<5jt	AG58
20101136	21	<5jt	AG55
20100104	21	<5jt	AG57
20100994	21	<5jt	AG57
20100867	21	<5jt	AG57

Gambar 4.7 Sebagian *dataset* hasil *transformasion*

Atribut UMUR dan INCOME merupakan atribut yang berisikan kriteria dari nasabah dimana pengelompokkan hasil *sorting* dapat dilihat pada tabel 4.1.

Tabel 4.1 Kriteria nasabah

UMUR	INCOME
<30 >30 <40	<5JT >5jt >10JT
40 – 50 40 – 50 40 – 50	>5JT <10JT <5JT
>50 >50 >50	>5JT <10JT >10JT

Berdasarkan kriteria pada tabel 4.1 di atas maka didapatlah pengelompokkan atribut UMUR dan INCOME sebagai KRITERIA nasabah yang akan menjadi target dalam proses *mining.*

Karena *software* yang digunakan untuk data *mining* merupakan DTREG maka *dataset* di atas terlebih dahulu di-*export* kedalam *format* (.xlsx) kemudian dari *format* (.xlsx) akan dirubah lagi menjadi *format* "csv" *(Comma Delimited).* Berikut *dataset* yang telah di-*export* ke dalam *format* (.xlsx) dapat dilihat pada gambar 4.8.

	A	B	C	D
1	NOPOLIS	UMUR	INCOME	PLAN
2	20101124	20	<5jt	AG55
3	20101445	20	<5jt	AG57
4	20101360	20	<5jt	AG57
5	20100994	21	<5jt	AG57
6	20100867	21	<5jt	AG57
7	20100486	21	<5jt	AG57
8	20100539	21	<5jt	AG55
9	20100556	21	<5jt	AG57
10	20100234	22	<5jt	AG57
11	20101570	22	<5jt	AG55
12	20100175	23	<5jt	AG55
13	20101318	24	<5jt	AG55
14	20100275	24	<5jt	AG31
15	20100616	24	<5jt	AG57
16	20100106	24	<5jt	AG57
17	20100587	25	<5jt	AG57
18	20100173	25	<5jt	AG55
19	20100996	26	<5jt	AG55
20	20100233	26	<5jt	AG57
21	20101428	26	<5jt	AG31
22	20100217	27	<5jt	AG55
23	20101064	27	<5jt	AG55
24	20101222	27	<5jt	AG57
25	20101545	27	<5jt	AG55
26	20100910	28	<5jt	AG60
27	20100529	28	<5jt	AG57
28	20101095	28	<5jt	AG57
29	20101378	28	<5jt	AG55
30	20101189	28	<5jt	AG55

Gambar 4.8 *Dataset* dengan *format* xlsx

Dataset dengan *format* (.xlsx) akan di-*export* lagi kedalam *format* "csv" *(Comma Delimited)* karena Software yang digunakan hanya dapat membaca file data dengan *format* "csv" *(Comma Delimited).* Berikut *dataset* yang telah sesuai untuk proses *mining* di-*export* ke dalam *format* "csv" dapat dilihat pada gambar 4.9.

	A	B	C	D
1	NOPOLIS	UMUR	INCOME	PLAN
2	20101124	20	<5jt	AG55
3	20101445	20	<5jt	AG57
4	20101360	20	<5jt	AG57
5	20100994	21	<5jt	AG57
6	20100867	21	<5jt	AG57
7	20100486	21	<5jt	AG57
8	20100539	21	<5jt	AG55
9	20100556	21	<5jt	AG57
10	20100234	22	<5jt	AG57
11	20101570	22	<5jt	AG55
12	20100175	23	<5jt	AG55
13	20101318	24	<5jt	AG55
14	20100275	24	<5jt	AG31
15	20100616	24	<5jt	AG57
16	20100106	24	<5jt	AG57
17	20100587	25	<5jt	AG57
18	20100173	25	<5jt	AG55
19	20100996	26	<5jt	AG55
20	20100233	26	<5jt	AG57
21	20101428	26	<5jt	AG31
22	20100217	27	<5jt	AG55
23	20101064	27	<5jt	AG55
24	20101222	27	<5jt	AG57
25	20101545	27	<5jt	AG55

Gambar 4.9 *Dataset* dengan *format* "csv" *(Comma Delimited)*

Dataset pada gambar 4.9 telah memiliki bentuk yang sesuai untuk tahap data *mining* menggunakan teknik *Decision Tree.*

BAB V

HASIL DAN PEMBAHASAN

5.1 *Data Mining*

Data mining merupakan tahapan untuk menemukan pola atau informasi dalam sekumpulan data dengan menggunakan teknik dan algoritma tertentu. Pemilihan teknik dan algoritma yang tepat sangat bergantung pada proses KDD secara keseluruhan. Pada penelitian ini penerapan data mining menggunakan teknik *decision tree* dan algoritma *C4.5* untuk menemukan informasi mengenai kriteria nasabah.

5.1.1 Penerapan Klasifikasi *Decission Tree* dengan Algoritma *C4.5*

Setelah melakukan proses transformasi data kedalam bentuk data yang sesuai untuk penerapan data *mining* dengan teknik *decision tree* maka tahapan ini dapat dilakukan. *Decission Tree* adalah mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan dan juga dapat diekspresikan dalam bentuk basis data seperti *Structured Query Language* untuk mencari *record* pada kategori tertentu.

Dalam tahapan penemuan aturan *decision tree* ini, langkah awal yang akan dilakukan adalah mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target.

Selanjutnya menyiapkan data awal yang telah siap untuk di-*mining* menggunakan *software decision tree* dengan *format* "csv" *(Comma Delimited).*

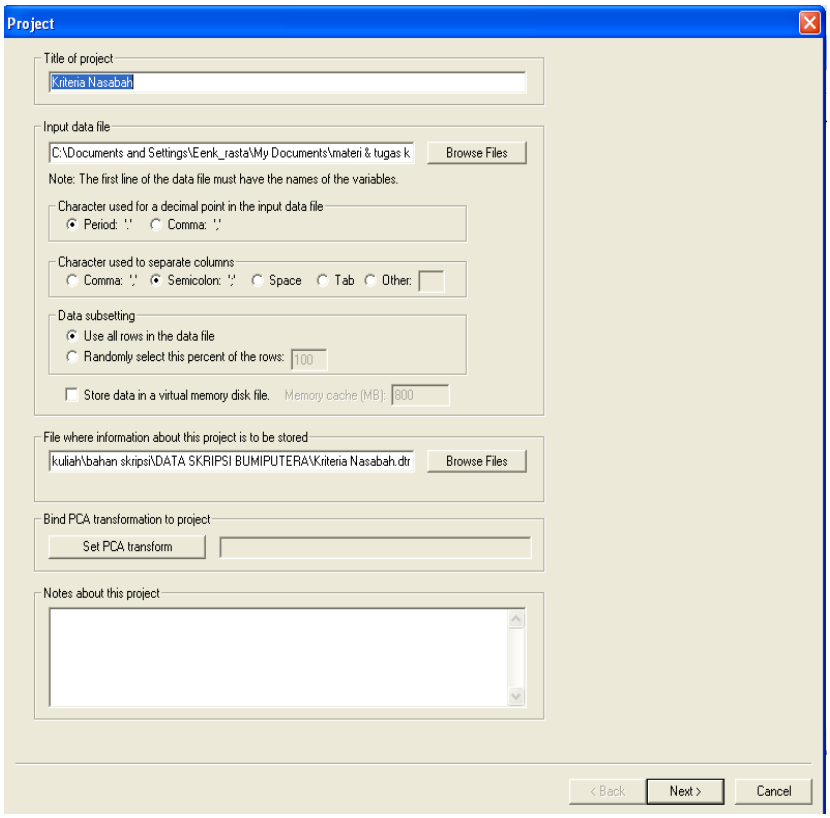
Berikut merupakan proses *mining* untuk menemukan informasi mengenai kriteria nasabah berdasarkan jenis asuransi yang dipilih.

1. Data Nasabah, pada tahapan ini data siap di inputkan untuk proses analisis, seperti pada gambar 5.1.

	A	B	C	D
1	NOPOLIS	UMUR	INCOME	PLAN
2	20101124	20	<5jt	AG55
3	20101445	20	<5jt	AG57
4	20101360	20	<5jt	AG57
5	20100994	21	<5jt	AG57
6	20100867	21	<5jt	AG57
7	20100486	21	<5jt	AG57
8	20100539	21	<5jt	AG55
9	20100556	21	<5jt	AG57
10	20100234	22	<5jt	AG57
11	20101570	22	<5jt	AG55
12	20100175	23	<5jt	AG55
13	20101318	24	<5jt	AG55
14	20100275	24	<5jt	AG31
15	20100616	24	<5jt	AG57
16	20100106	24	<5jt	AG57
17	20100587	25	<5jt	AG57
18	20100173	25	<5jt	AG55
19	20100996	26	<5jt	AG55
20	20100233	26	<5jt	AG57
21	20101428	26	<5jt	AG31
22	20100217	27	<5jt	AG55
23	20101064	27	<5jt	AG55
24	20101222	27	<5jt	AG57
25	20101545	27	<5jt	AG55

Gambar 5.1 Kriteria Nasabah dengan *format "csv"*

2. Input Data, setelah semua pola data sudah siap maka data akan di inputkan, seperti gambar 5.2.



Gambar 5.2 *input* data

5.1.2 Algoritma ID3 dan Algoritma C4.5

Sebelum membahas algoritma C4.5 perlu dijelaskan terlebih dahulu algoritma ID3 karena C4.5 adalah ekstensi dari algoritma *decision-tree* ID3. Algoritma ID3/C4.5 ini secara rekursif membuat sebuah *decision tree* berdasarkan *training* data yang telah disiapkan. Algoritma ini mempunyai inputan berupa *training samples* dan *samples*. *Training samples* berupa data contoh yang akan digunakan untuk membangun sebuah *tree* yang telah diuji kebenaranya. Sedangkan *samples* merupakan *field-field* data yang nantinya akan kita gunakan sebagai parameter dalam melakukan klasifikasi data. Berikut adalah algoritma dasar dari ID3 dan C4.5

5.1.2.1 Algoritma ID3

Algoritma Dasar ID3

Input : *Training samples, samples*
Output : *Decision tree*
Method :

(1) *Create node N;*
(2) ***If*** *samples are all of the same class, C* ***then***
(3) *Return N as a leaf node labeled with the class C;*
(4) ***if*** *atribute-list is empty* ***then***
(5) *Return N as a leaf node labeled with the most common class in samples; // majority voting*
(6) *select test-atribute, atribute among atribute-list with the highest information gain;*
(7) *label node N with test-atribute;*
(8) *for each known value ai of test-atribute // partition the samples*
(9) *grow a branch from node N for the condition test-atribute = ai;*
(10) *let si be the set of samples in samples for which test-atribute = ai; // a partition*
(11) ***if*** *si is empty then*
(12) *attach a leaf labeled with the ,most common class in samples;*
(13) ***else*** *attach the node returned by Generate_decision_tree(si, attribute-list-test-atribute);*

5.1.2.2 Algoritma C4.5

Algoritma Dasar C4.5

1. *Build the decision tree form the training set (conventional ID3).*
2. *Convert the resulting tree into an equivalent set of rules. The number of rules is equivalent to the number of possible paths from the root to a leaf node.*
3. *Prune (generalize) each rule by removing preconditions that increase classification accuracy.*
4. *Sort pruned rules by their accuracy, and use them in this order when classifying future test examples.*

5.1.2.3. Information Gain

Information gain adalah salah satu a*tribute selection measure* yang digunakan untuk memilih *test* atribut tiap *node* pada *tree*. Atribut dengan *information gain* tertinggi dipilih sebagai *test* atribut dari suatu *node*. Ada 2 kasus berbeda pada saat penghitungan *Information Gain*, pertama untuk kasus penghitungan atribut tanpa *missing value* dan kedua, penghitungan atribut dengan *missing value.* Disini hanya dijelaskan perhitungan atribut tanpa *missing value* karena data yang akan digunakan sudah melalui tahapan KDD yaitu *cleaning* yang artinya data sudah bersih dan siap untuk di-*mining,* berikut perhitungan tanpa *missing value* :

Misalkan S berisi *s* data *samples*. Anggap atribut untuk class memiliki *m* nilai yang berbeda, C_i (untuk i = 1, …,I). anggap S_i menjadi jumlah *samples* S pada class C_i . Maka besar *information*-nya dapat dihitung dengan :

$$I(S_1, S_2 \ldots, S_m) = -\sum_{i=1}^{m} P_i * \log_2(P_i)$$

Dimana $P_i = \frac{si}{s}$ adalah probabilitas dari *sample* yang mempunyai *class* C_i.

Misalkan atribut A mempunyai *v* nilai yang berbeda, $\{a_1, a_2 \ldots, a_v\}$. Atribut A dapat digunakan untuk mempartisi S menjadi *v* subset, $\{S_1, S_2 \ldots, S_v\}$, dimana S_j berisi *samples* pada S yang mempunyai nilai a_j dari A. Jika A terpilih menjadi *test* atribut (yaitu, *best* atribut untuk *splitting*), maka subset-subset akan berhubungan dengan pertumbuhan *node-node* cabang yang berisi S. Anggap s_{ij}

sebagai jumlah *samples* class C_i pada subset S_j. *Entropy,* atau nilai *information* dari subset A adalah :

$$E(A) = \sum_{j=1}^{v} = \frac{S_{1j}+\cdots+S_{mj}}{S} \; I\,(S_1,\; S_2 \ldots, S_m)$$

$\frac{S_{1j}+\cdots+S_{mj}}{S}$ adalah bobot dari subset *j*th dan jumlah *samples* pada subset (yang mempunyai nilai a_j dari A) dibagi dengan jumlah total samples pada S. Untuk subset S_j,

$$I\,(S_{1j},\; S_{2j} \ldots, S_{mj}) \;=\; -\sum_{i=1}^{m} P_{ij} * \log_2(P_{ij})$$

Dimana $P_{ij} = \frac{S_{ij}}{|S_j|}$ adalah probabilitas sample S_j yang mempunyai class C_i .

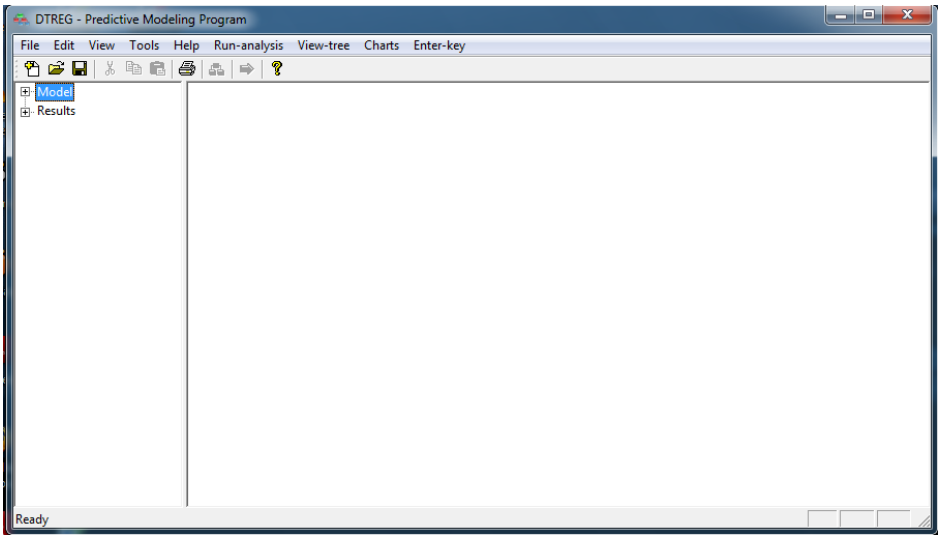
Maka nilai information gain atribut A pada subset S adalah :

$$Gain\ (A)\ (S_1,\; S_2 \ldots, S_m) \;-\; E\ (A)$$

5.2 Proses *Data Mining* Menggunakan DTREG

Setelah dijelaskan proses penerapan data mining dengan teknik *decision tree* secara teoritis pada penjelasan di atas, maka kali ini akan di jelaskan proses data *mining* secara aplikatif dimana proses data *mining* yang akan dilakukan menggunakan *software* data *mining* DTREG.

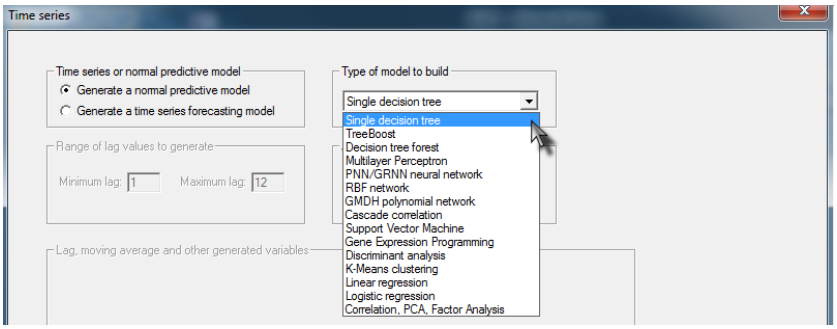
Seperti yang telah dijelaskan pada bab sebelumnya Dtreg adalah software analisis program statistik yang menghasilkan klasifikasi dan pohon keputusan bahwa model regresi data dapat digunakan untuk memprediksi nilai. Dtreg hanya dapat membaca *file* data dengan *format* "csv" *(Comma Delimited)*, kemudian data dapat diinputkan ke DTREG dan akan diproses untuk membuat pohon keputusan ke ukuran yang optimal.



Gambar 5.3 Tampilan Awal Program DTREG

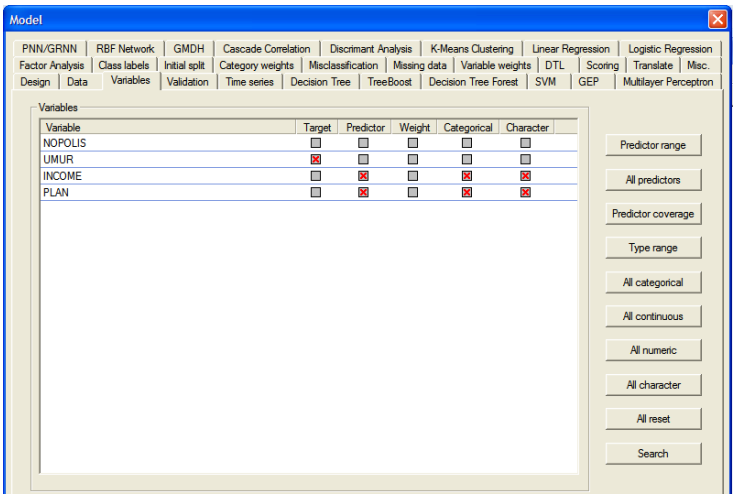
Pada gambar 5.3 terlihat tampilan awal program DREG, data yang sudah diinputkan seperti penjelasan diatas tidak dapat dilihat tetapi data tersebut sudah masuk kedalam program DTREG dan akan dilakukan proses selanjutnya.

Dalam proses data mining menggunakan DTREG ini data yang digunakan merupakan data yang telah di transformasi kedalam *format Microsoft Excel* 2007 (csv). Selanjutnya proses data mining dilakukan dengan menggunakan menu *Single Decision Tree* → pada DTREG, seperti pada gambar 5.4.



Gambar 5.4 Menu DTREG

Setelah memilih menu *single decision tree* seperti pada gambar 5.4, maka langkah yang dilakukan selanjutnya yaitu memilih variabel yang akan menjadi target dan variabel predictor untuk hasil yang diinginkan seperti pada gambar 5.5.

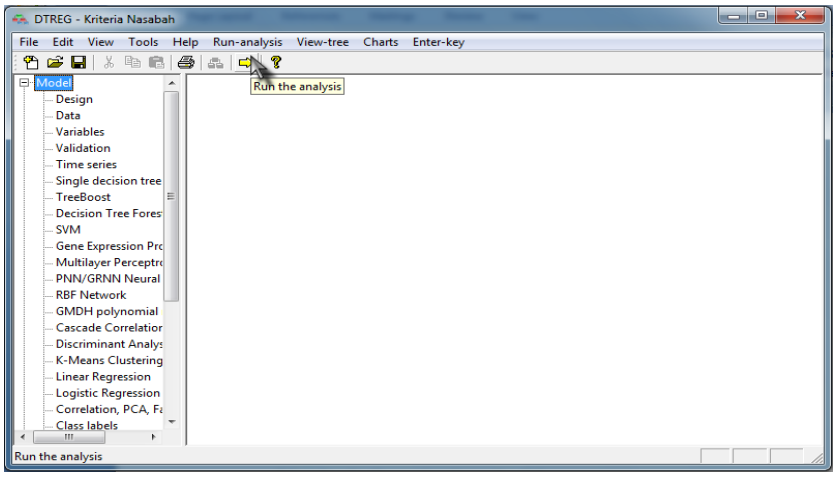


Gambar 5.5 Pemilihan Variabel

Pada gambar 5.5 merupakan tampilan proses penentuan *field* yang akan menjadi target untuk menentukan hasil yang diharapkan. Seperti yang terlihat

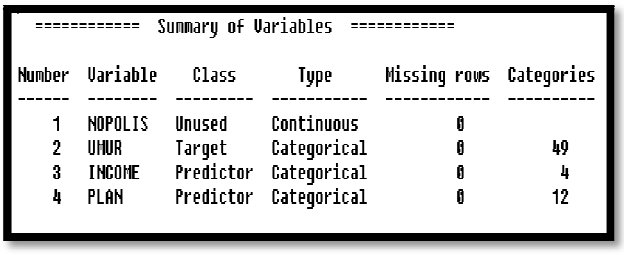
pada gambar 5.5 tersebut, *field* UMUR dipilih sebagai target karena *field* tersebut memiliki nilai *gain* tertinggi dari *field-field* yang lainnya.

Proses selanjutnya data siap dilakukan RUN untuk melihat hasil dari proses DTREG yang akan mengubah data dengan ukuran besar kedalam bentuk pohon keputusan yang optimal seperti gambar 5.6.



Gambar 5.6 Proses Run Analysis

Dari proses yang akan dilakukan pada gambar 5.6 dimana dari 1610 data akan di-*run* untuk menghasilkan analysis data yang kemudian akan menghasikan data dalam bentuk pohon. Dari proses yang dilakukan pada gambar 5.6 didapatkan hasil seperti pada gambar 5.7.



========== Summary of Variables ==========

Number	Variable	Class	Type	Missing rows	Categories
1	NOPOLIS	Unused	Continuous	0	
2	UMUR	Target	Categorical	0	49
3	INCOME	Predictor	Categorical	0	4
4	PLAN	Predictor	Categorical	0	12

Gambar 5.7 Variabel hasil *Analysis*

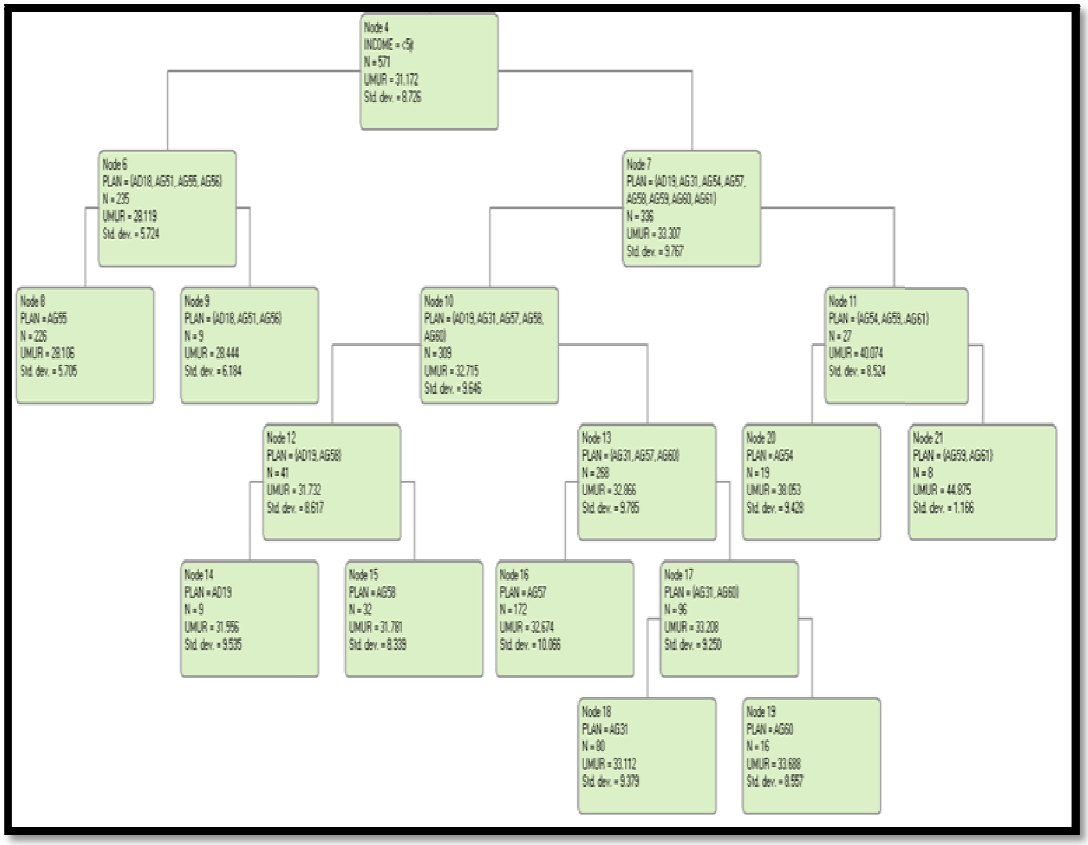
Hasil dari proses data mining menggunakan DTREG terdapat *Results* dengan bagian – bagian tertentu hasil *analysis* diantaranya seperti gambar 5.7 diatas bagian dari *results analysis report* yaitu *summary of variable* menunjukkan bahwa *field* yang diinputkan menghasilkan variabel *class* yang memiliki 4 *class* dan hanya 2 *class* yang digunakan yaitu *class target, dan predictor* variabel UMUR menjadi *class target* karena memiliki nilai *gain* teringgi dengan *categories* nilai 49, sedangkan INCOME dan PLAN menjadi *class predictor* karna nilai *gain*-nya dibawah UMUR.

Seperti penjelasan diatas maka yang akan menjadi titik akar *(root)* adalah UMUR, sedangkan INCOME dan PLAN akan menjadi cabang dari pohon keputusan yang merupakan pembagian berdasarkan hasil uji, sedangkan titik akhir *(leaf)* merupakan pembagian kelas yang dihasilkan. Karena data yang di-*run* dalam jumlah besar maka hasil *run* akan di pisah menjadi 3 bagian, berikut gambar pertama dari hasil *run* dengan kriteria INCOME = {<5jt,>5jt} yang akan memperlihatkan *decision tree* (Pohon Keputusan) seperti pada gambar 5.8.

Gambar 5.8 *root* bagian dari INCOME = {<5jt, >5jt}

Dari gambar 5.8 maka dapat dilihat bahwa jumlah nasabah dengan kriteria INCOME = <5jt, >5jt dan UMUR diantara 34 tahun sebanyak 1312 dari jumlah data asli sebanyak 1610 nasabah yang akan terbagi menjadi 2 bagian yaitu berdasarkan INCOME <5jt dan >5jt. Berikut bagian pertama dengan INCOME <5jt pada gambar 5.9.

1. INCOME <5JT



Gambar 5.9 Hasil dari INCOME <5jt

Berdasarkan pohon keputusan diatas maka didapatkan hasil dari INCOME <5jt pada *node 4* dengan jumlah nasabah sebanyak 571 dan UMUR diantara 30 tahun dalam satu tahun tepatnya pada tahun 2010.

Agar lebih spesifik dan data mudah dimengerti maka dari 571 nasabah dipecah lagi berdasarkan kriteria UMUR dan INCOME <5jt.

Node 6 merupakan cabang pertama yang menjadi *root* dari INCOME <5jt yang jumlah nasabah sebanyak 571 dan akan dibagi menjadi 2 cabang. Cabang pertama dengan jumlah nasabah sebanyak 235 dan UMUR diantara 28 tahun

memilih jenis asuransi yaitu dengan kode PLAN = {AD18, AG51, AG55, dan AG56}.

Node 8 akan menghasilkan pembagian dari *node* 6 dengan jumlah nasabah sebanyak 235 berdasarkan jenis PLAN yang sama yaitu PLAN = AG55 dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 226 nasabah dan berUMUR diantara 28 tahun.

Node 9 akan menghasilkan pembagian dari *node* 6 dengan jumlah nasabah sebanyak 235 berdasarkan jenis PLAN yang sama yaitu PLAN = {AD18, AG51, dan AG56} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 9 nasabah dan berUMUR diantara 28 tahun

Berdasarkan penjelasan diatas dari *node* 4 dengan INCOME <5jt dan jumlah nasabah sebanyak 571 yang akan dibagi menjadi 2 cabang berdasarkan INCOME <5jt dan jenis asuransi yang sama, maka *node 6, 8, dan 9* menjadi titik akhir *(leaf)*.

Node 7 merupakan cabang kedua yang menjadi *root* dari INCOME <5jt yang jumlah nasabah sebanyak 571 dan akan dibagi menjadi 2 cabang. Cabang kedua dengan jumlah nasabah sebanyak 336 dan UMUR diantara 33 tahun memilih jenis asuransi yaitu dengan kode PLAN = {AD19, AG31, AG54 AG57, AG58, AG59, AG60 dan AG61}.

Node 10 akan menghasilkan pembagian dari *node 7* dengan jumlah nasabah sebanyak 336 berdasarkan jenis PLAN yang sama yaitu PLAN = {AD19, AG31, AG57, dan AG58} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 309 nasabah dan berUMUR diantara 32 tahun.

Node 12 akan menghasilkan pembagian dari *node 10* dengan jumlah nasabah sebanyak 309 berdasarkan jenis PLAN yang sama yaitu PLAN = {AD19 dan AG58} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 41 nasabah dan berUMUR diantara 31 tahun.

Node 14 akan menghasilkan pembagian dari *node 12* dengan jumlah nasabah sebanyak 41 berdasarkan jenis PLAN yang sama yaitu PLAN = AD19 dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 9 nasabah dan berUMUR diantara 31 tahun yang menjadi titik akhir *(leaf)* dari pembagian cabang *node 12.*

Node 15 akan menghasilkan pembagian dari *node 12* dengan jumlah nasabah sebanyak 41 berdasarkan jenis PLAN yang sama yaitu PLAN = AG58 dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 32 nasabah dan berUMUR diantara 31 tahun yang menjadi titik akhir *(leaf)* dari pembagian cabang *node 12.*

Node 13 akan menghasilkan pembagian dari *node 10* dengan jumlah nasabah sebanyak 309 berdasarkan jenis PLAN yang sama yaitu PLAN = {AG31, AG57, dan AG60} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 268 nasabah dan berUMUR diantara 32 tahun.

Node 16 akan menghasilkan pembagian dari *node 13* dengan jumlah nasabah sebanyak 268 berdasarkan jenis PLAN yang sama yaitu PLAN = AG57 dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 172 nasabah dan berUMUR diantara 32 tahun yang menjadi titik akhir *(leaf)* dari pembagian cabang *node 13.*

Node 17 akan menghasilkan pembagian dari *node 13* dengan jumlah nasabah sebanyak 268 berdasarkan jenis PLAN yang sama yaitu PLAN = {AG31 dan AG60} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 96 nasabah dan berUMUR diantara 33 tahun.

Node 18 akan menghasilkan pembagian dari *node 17* dengan jumlah nasabah sebanyak 96 berdasarkan jenis PLAN yang sama yaitu PLAN = AG31 dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 80 nasabah dan berUMUR diantara 33 tahun yang menjadi titik akhir *(leaf)* dari pembagian cabang *node 17.*

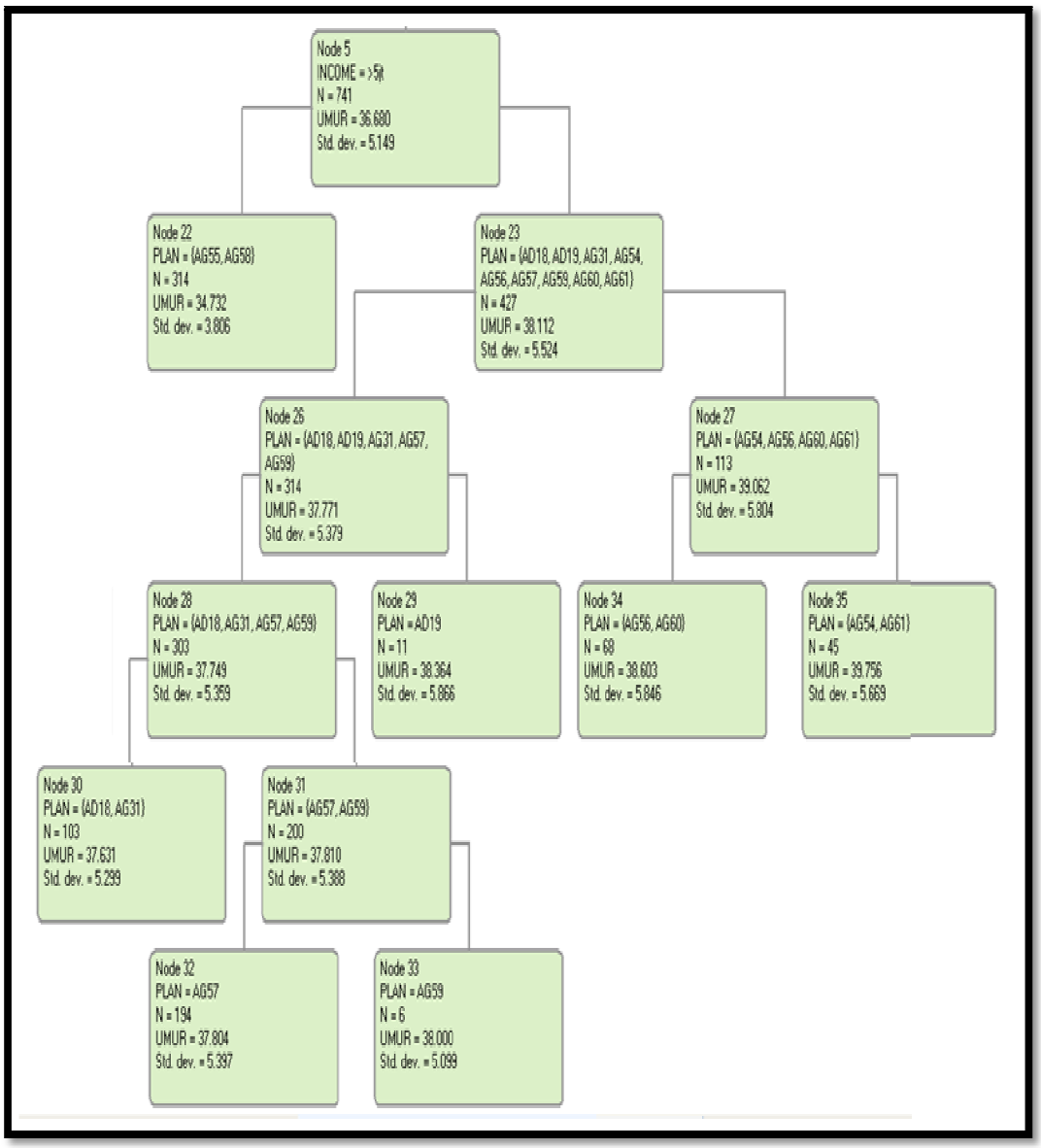
Node 19 akan menghasilkan pembagian dari *node 17* dengan jumlah nasabah sebanyak 96 berdasarkan jenis PLAN yang sama yaitu PLAN = AG60 dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 16 nasabah dan berUMUR diantara 33 tahun yang menjadi titik akhir *(leaf)* dari pembagian cabang *node 17.*

Node 11 akan menghasilkan pembagian dari *node 7* dengan jumlah nasabah sebanyak 336 berdasarkan jenis PLAN yang sama yaitu PLAN = {AG54, AG59, dan AG61} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 27 nasabah dan berUMUR diantara 40 tahun.

Node 20 akan menghasilkan pembagian dari *node 11* dengan jumlah nasabah sebanyak 27 berdasarkan jenis PLAN yang sama yaitu PLAN = AG54 dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 19 nasabah dan berUMUR diantara 38 tahun yang menjadi titik akhir *(leaf)* dari pembagian cabang *node 11.*

Node 21 akan menghasilkan pembagian dari *node 11* dengan jumlah nasabah sebanyak 27 berdasarkan jenis PLAN yang sama yaitu PLAN = {AG59 dan AG61} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 8 nasabah dan berUMUR diantara 44 tahun yang menjadi titik akhir *(leaf)* dari pembagian cabang *node 11*.

2. INCOME >5JT



Gambar 5.10 Hasil dari INCOME >5jt

Berdasarkan pohon keputusan diatas maka didapatkan hasil dari INCOME >5jt pada *node 5* dengan jumlah nasabah sebanyak 741 dan UMUR diantara 36 tahun dalam satu tahun tepatnya pada tahun 2010.

Agar lebih spesifik dan data mudah dimengerti maka dari 741 nasabah dipecah lagi berdasarkan kriteria UMUR dan INCOME >5jt.

Node 22 merupakan cabang pertama dari *node 5* dengan jumlah nasabah sebanyak 741 dan menjadi titik akhir dari *node 5* dengan INCOME >5jt yang jumlah nasabah sebanyak 314 dan UMUR diantara 34 tahun memilih jenis asuransi yaitu dengan kode PLAN = {AD55 dan AG58}.

Node 23 merupakan cabang kedua dari *node 5* dengan jumlah nasabah sebanyak 741 nasabah yang akan menghasilkan pembagian ke *node 23* dengan jumlah nasabah sebanyak 427 berdasarkan jenis PLAN yaitu = {AD18, AD19, AG31, AG54, AG56, AG57, AG59, AG60, dan AG61} berUMUR diantara 38 tahun.

Node 26 akan menghasilkan pembagian dari *node 23* dengan jumlah nasabah sebanyak 427 berdasarkan jenis PLAN yang sama yaitu PLAN = {AD18, AD19, AG31, AG57 dan AG59} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 314 nasabah dan berUMUR diantara 37 tahun.

Node 28 akan menghasilkan pembagian dari *node 26* dengan jumlah nasabah sebanyak 314 berdasarkan jenis PLAN yang sama yaitu PLAN = {AD18, AG31, AG57, dan AG59} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 303 nasabah dan berUMUR diantara 37 tahun.

Node 30 akan menghasilkan pembagian dari *node 28* dengan jumlah nasabah sebanyak 303 berdasarkan jenis PLAN yang sama yaitu PLAN = {AD18, dan AG31} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 103 nasabah dan berUMUR diantara 37 tahun yang menjadi titik terakhir *(leaf)* pada pembagian cabang *node 28*.

Node 31 akan menghasilkan pembagian dari *node 28* dengan jumlah nasabah sebanyak 303 berdasarkan jenis PLAN yang sama yaitu PLAN = {AG57 dan AG59} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 200 nasabah dan berUMUR diantara 37 tahun.

Node 32 akan menghasilkan pembagian dari *node 31* dengan jumlah nasabah sebanyak 200 berdasarkan jenis PLAN yang sama yaitu PLAN = AG57 dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 194 nasabah dan berUMUR diantara 37 tahun yang menjadi titik akhir *(leaf)* dari pembagian cabang *node 31*.

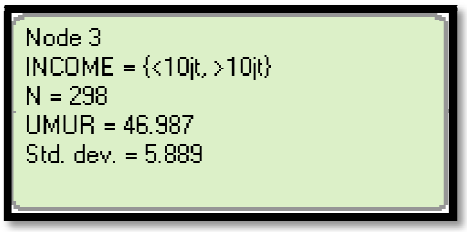
Node 33 akan menghasilkan pembagian dari *node 31* dengan jumlah nasabah sebanyak 200 berdasarkan jenis PLAN yang sama yaitu PLAN = AG59 dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 27 nasabah dan berUMUR diantara 38 tahun yang menjadi titik akhir *(leaf)* dari pembagian cabang *node 31*.

Node 27 akan menghasilkan pembagian dari *node 23* dengan jumlah nasabah sebanyak 427 berdasarkan jenis PLAN yang sama yaitu PLAN = {AG54, AG56, AG60, AG61} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 113 nasabah dan berUMUR diantara 39 tahun.

Node 34 akan menghasilkan pembagian dari *node 27* dengan jumlah nasabah sebanyak 113 berdasarkan jenis PLAN yang sama yaitu PLAN = {AG56 dan AG60} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 68 nasabah dan berUMUR diantara 38 tahun yang menjadi titik akhir *(leaf)* dari pembagian cabang *node 27.*

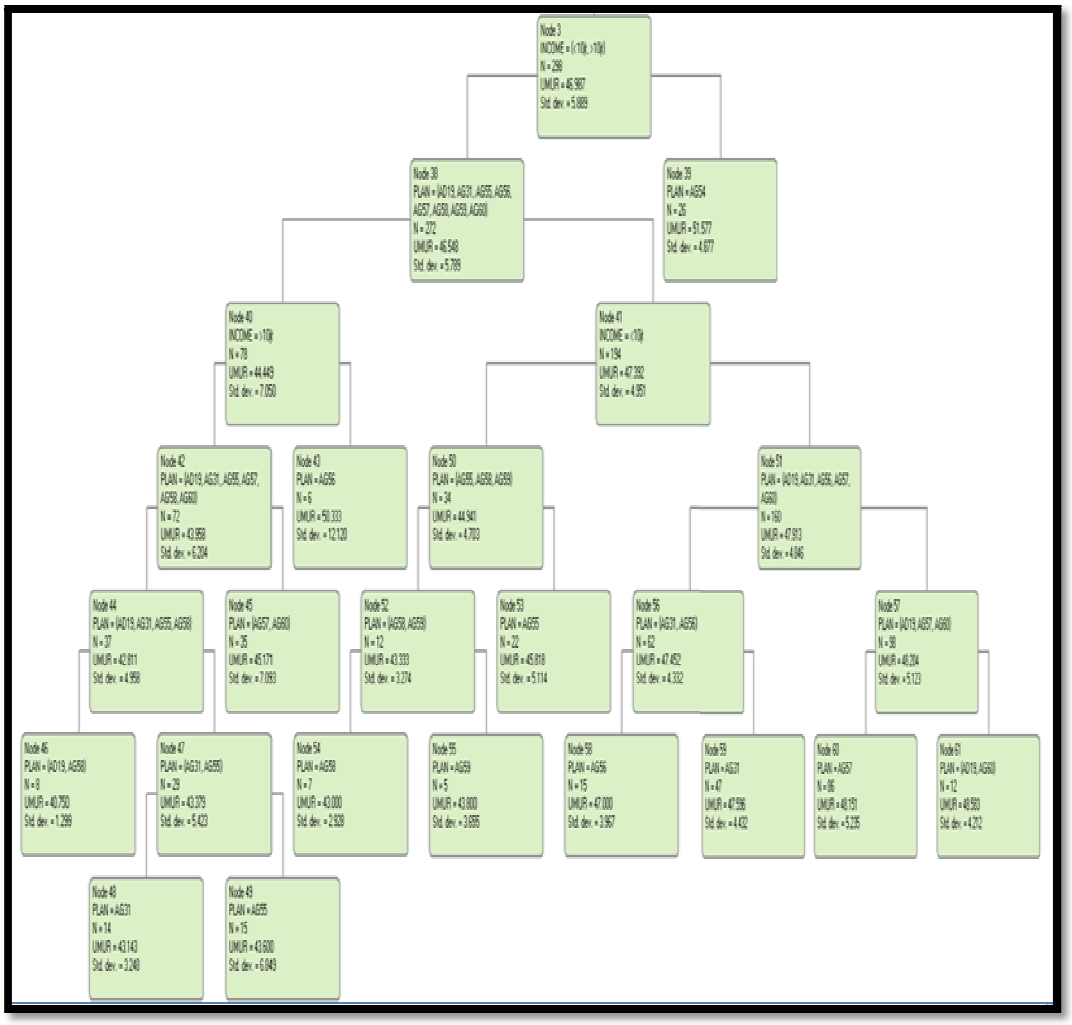
Node 35 akan menghasilkan pembagian dari *node 27* dengan jumlah nasabah sebanyak 113 berdasarkan jenis PLAN yang sama yaitu PLAN = {AG54 dan AG61} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 45 nasabah dan berUMUR diantara 39 tahun yang menjadi titik akhir *(leaf)* dari pembagian cabang *node 27.*

3. INCOME <10JT, >10JT



Gambar 5.11 *root* bagian dari INCOME = {<10jt, >10jt}

Dari gambar 5.11 maka dapat dilihat bahwa jumlah nasabah dengan kriteria INCOME = <10jt, >10jt dan UMUR diantara 46 tahun sebanyak 298 nasabah dan lebih sedikit dari nasabah dengan INCOME <5jt dan >5jt sebanyak 1312 dari jumlah data asli sebanyak 1610 nasabah yang akan terbagi menjadi 2 bagian lagi yaitu berdasarkan INCOME <10jt dan >10jt. Berikut gambar 5.12 dengan INCOME <10jt dan >10jt.



Gambar 5.12 Hasil dari INCOME <10jt dan >10jt

Berdasarkan pohon keputusan diatas maka didapatkan hasil dari INCOME >10jt dan INCOME >10jt pada *node 3* dengan jumlah nasabah sebanyak 298 dan UMUR diantara 46 tahun dalam satu tahun tepatnya pada tahun 2010.

Agar lebih spesifik dan data mudah dimengerti maka dari 298 nasabah dipecah lagi berdasarkan kriteria UMUR dan INCOME <10jt dan INCOME >10jt.

Node 38 merupakan cabang pertama dari *node 3* dengan jumlah nasabah sebanyak 298 yang akan dibagi berdasarkan jenis asuransi yaitu dengan kode asuransi PLAN = {AD19, AG31, AG55, AG56, AG57, AG58, AG59 dan AG60} dengan jumlah nasabah sebanyak 272 dan UMUR diantara 46 tahun.

Node 40 merupakan cabang dari *node 38* yang menunjukkan bahwa jenis asuransi dengan kode PLAN = {AD19, AG31, AG55, AG56, AG57, AG58, AG59 dan AG60} akan dibagi berdasarkan kriteria INCOME >10jt dengan jumlah nasabah sebanyak 272 dan UMUR diantara 44 tahun.

Node 42 dengan INCOME >10jt akan menghasilkan pembagian dari *node 40* dengan jumlah nasabah sebanyak 272 berdasarkan jenis PLAN yang sama yaitu PLAN = {AD19, AG31, AG55, AG57, AG58 dan AG60} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 72 nasabah dan berUMUR diantara 43 tahun.

Node 44 dengan INCOME >10jt akan menghasilkan pembagian dari *node 42* dengan jumlah nasabah sebanyak 72 berdasarkan jenis PLAN yang sama yaitu PLAN = {AD19, AG31, AG55 dan AG58} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 37 nasabah dan berUMUR diantara 42 tahun.

Node 46 dengan INCOME >10jt akan menghasilkan pembagian dari *node 44* dengan jumlah nasabah sebanyak 37 berdasarkan jenis PLAN yang sama yaitu PLAN = {AD19 dan AG58} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 8 nasabah dan berUMUR diantara 40 tahun yang menjadi titik akhir *(leaf)* dari *node 44.*

Node 47 dengan INCOME >10jt akan menghasilkan pembagian dari *node 44* dengan jumlah nasabah sebanyak 37 berdasarkan jenis PLAN yang sama yaitu PLAN = {AG31 dan AG55} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 29 nasabah dan berUMUR diantara 43 tahun.

Node 48 dengan INCOME >10jt akan menghasilkan pembagian dari *node 47* dengan jumlah nasabah sebanyak 29 berdasarkan jenis PLAN yang sama yaitu PLAN = AG31 dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 14 nasabah dan berUMUR diantara 43 tahun yang menjadi titik akhir *(leaf)* dari *node 47*.

Node 49 dengan INCOME >10jt akan menghasilkan pembagian dari *node 47* dengan jumlah nasabah sebanyak 29 berdasarkan jenis PLAN yang sama yaitu PLAN = AG55 dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 15 nasabah dan berUMUR diantara 43 tahun yang menjadi titik akhir *(leaf)* dari *node 47*.

Node 39 dengan INCOME >10jt akan menghasilkan pembagian dari *node 3* dengan jumlah nasabah sebanyak 298 berdasarkan jenis asuransi yang dipilih yaitu PLAN = AG54 dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 26 nasabah dan berUMUR diantara 51 tahun yang menjadi titik akhir *(leaf)* dari *node 3*.

Node 41 merupakan cabang dari *node 38* yang menunjukkan bahwa jenis asuransi dengan kode PLAN = {AD19, AG31, AG55, AG56, AG57, AG58, AG59 dan AG60} akan dibagi berdasarkan kriteria INCOME <10jt dengan jumlah nasabah sebanyak 194 dan UMUR diantara 47 tahun.

Node 50 dengan INCOME <10jt akan menghasilkan pembagian dari *node 41* dengan jumlah nasabah sebanyak 194 berdasarkan jenis PLAN yang sama yaitu PLAN = { AG55, AG58 dan AG59} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 44 nasabah dan berUMUR diantara 34 tahun.

Node 52 dengan INCOME <10jt akan menghasilkan pembagian dari *node 50* dengan jumlah nasabah sebanyak 34 berdasarkan jenis PLAN yang sama yaitu PLAN = { AG58 dan AG59} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 12 nasabah dan berUMUR diantara 43 tahun.

Node 54 dengan INCOME <10jt akan menghasilkan pembagian dari *node 52* dengan jumlah nasabah sebanyak 12 berdasarkan jenis PLAN yang sama yaitu PLAN = AG58 dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 7 nasabah dan berUMUR diantara 43 tahun yang menjadi titik akhir *(leaf)* dari *node 52*.

Node 55 dengan INCOME <10jt akan menghasilkan pembagian dari *node 52* dengan jumlah nasabah sebanyak 12 berdasarkan jenis PLAN yang sama yaitu PLAN = AG59 dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 5 nasabah dan berUMUR diantara 43 tahun yang menjadi titik akhir *(leaf)* dari *node 52.*

Node 53 dengan INCOME <10jt akan menghasilkan pembagian dari *node 50* dengan jumlah nasabah sebanyak 34 berdasarkan jenis PLAN yang sama yaitu PLAN = AG55 dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 22 nasabah dan berUMUR diantara 45 tahun yang menjadi titik akhir *(leaf)* dari *node 50.*

Node 51 dengan INCOME <10jt akan menghasilkan pembagian dari *node 41* dengan jumlah nasabah sebanyak 194 berdasarkan jenis PLAN yang sama yaitu PLAN = { AD19, AD31, AG56, AG57 dan AG60} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 160 nasabah dan berUMUR diantara 47 tahun.

Node 56 dengan INCOME <10jt akan menghasilkan pembagian dari *node 51* dengan jumlah nasabah sebanyak 160 berdasarkan jenis PLAN yang sama yaitu PLAN = { AG31 dan AG56} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 62 nasabah dan berUMUR diantara 47 tahun.

Node 58 dengan INCOME <10jt akan menghasilkan pembagian dari *node 56* dengan jumlah nasabah sebanyak 62 berdasarkan jenis PLAN yang sama yaitu PLAN = AG56 dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 15 nasabah dan berUMUR diantara 47 tahun yang menjadi titik akhir *(leaf)* dari *node 56.*

Node 59 dengan INCOME <10jt akan menghasilkan pembagian dari *node 56* dengan jumlah nasabah sebanyak 62 berdasarkan jenis PLAN yang sama yaitu PLAN = AG31 dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 47 nasabah dan berUMUR diantara 47 tahun yang menjadi titik akhir *(leaf)* dari *node 56.*

Node 57 dengan INCOME <10jt akan menghasilkan pembagian dari *node 51* dengan jumlah nasabah sebanyak 160 berdasarkan jenis PLAN yang sama yaitu PLAN = { AD19, AG57 dan AG60} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 98 nasabah dan berUMUR diantara 48 tahun.

Node 60 dengan INCOME <10jt akan menghasilkan pembagian dari *node 57* dengan jumlah nasabah sebanyak 98 berdasarkan jenis PLAN yang sama yaitu PLAN = AG57 dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 86 nasabah dan berUMUR diantara 48 tahun yang menjadi titik akhir *(leaf)* dari *node 57*

Node 61 dengan INCOME <10jt akan menghasilkan pembagian dari *node 57* dengan jumlah nasabah sebanyak 98 berdasarkan jenis PLAN yang sama yaitu PLAN = {AD19 dan AG60} dengan jumlah nasabah yang memilih jenis asuransi tersebut sebanyak 12 nasabah dan berUMUR diantara 48 tahun yang menjadi titik akhir *(leaf)* dari *node 57*.

Setelah dilakukan pengujian pada hasil dan pembahasan diatas, maka didapatkan hasil pengujian untuk semua atribut pada metode yang diuji coba, hasil dari metode tersebut dapat dilihat pada tabel 5.1 sebagai berikut.

Tabel 5.1 Hasil Pengujian

INCOME	UMUR	PLAN ASURANSI	TOTAL
<5JT	≤ 30 - ≤ 47 Tahun	AG31, AG54, AG55, AG56, AG57, AG58, AG59, AG60, AG61, AD19.	571
>5JT	≥ 30 - >49 Tahun	AG31, AG54, AG55, AG57, AG58, AG60, AD19.	741
<10JT	> 39 - ≤ 45 Tahun	AG31, AG55, AG56, AG57, AG58, AG59, AG60.	118
	> 47 - ≤ 50 Tahun	AG31, AG54, AG56, AG57, AG60.	70
	> 50 - ≤ 59 Tahun	AG31, AG54, AG56, AG57, AG60.	92
>10JT	≤ 40 - > 59 Tahun	AG31, AG55, AG56, AG57.	80

Dari uraian pada tabel 5.1 maka dapat dilihat kriteria INCOME dan UMUR yang memilih jenis asuransi dengan PLAN sebagai kode asuransi tertentu untuk tujuan mempermudah dalam mengetehaui asuransi apa saja yang di ambil nasabah dengan kriteria INCOME dan UMUR sebagai kriteria untuk menentukan calon nasabah potensial yang baru, berikut analisa hasil dari tabel 5.1 :

1. Calon nasabah dengan kriteria INCOME <5jt dan UMUR ≤30 - ≤47 tahun berpotensial memilih produk asuransi dengan PLAN (AG31, AG54, AG55, AG56, AG57, AG58, AG59, AG60, AG61, AD19) berdasarkan kriteria nasabah lama seperti pada tabel diatas dengan peminat nasabah sebanyak 571.
2. Calon nasabah dengan kriteria INCOME >5jt dan UMUR ≥30 - >49 tahun berpotensial memilih produk asuransi dengan PLAN (AG31, AG54, AG55, AG57, AG58, AG60, AD19) berdasarkan kriteria nasabah lama seperti pada tabel diatas dengan peminat nasabah sebanyak 741.
3. Calon nasabah dengan kriteria INCOME <10jt dan UMUR >39 - ≤45 tahun berpotensial memilih produk asuransi dengan PLAN (AG31, AG55, AG56, AG57, AG58, AG59, AG60) berdasarkan kriteria nasabah lama seperti pada tabel diatas dengan peminat nasabah sebanyak 118. Nasabah dengan kriteria UMUR >47 - ≤50 tahun dengan INCOME yang sama yaitu <10jt memilih produk asursansi dengan PLAN (AG31, AG54, AG56, AG57, AG60) sebanyak 70 nasabah yang akan berpotensial bagi calon nasabah baru untuk menentukan produk asuransi yang akan dipilihnya dan membantu pihak asuransi dalam menawarkan produk asuransi yang sesuai untuk calon nasabah berdasarkan kriteria INCOME dan UMUR nasabah tersebut. Selanjutnya nasabah dengan kriteria UMUR >50 - ≤59 tahun dan pendapatan yang sama

yaitu INCOME <10jt memilih produk asuransi dengan PLAN (AG31, AG54, AG56, AG57, AG60) sebanyak 92 nasabah yang akan berpotensial bagi calon nasabah baru untuk menentukan produk asuransi yang akan dipilihnya dan membantu pihak asuransi dalam menawarkan produk asuransi yang sesuai untuk calon nasabah berdasarkan kriteria INCOME dan UMUR nasabah tersebut.

4. Calon nasabah dengan kriteria INCOME >10jt dan UMUR ≤40 - >59 tahun berpotensial memilih produk asuransi dengan PLAN (AG31, AG55, AG56, AG57) berdasarkan kriteria nasabah lama seperti pada tabel diatas dengan peminat nasabah sebanyak 80 orang.

Dari penjelasan diatas maka hasil dari penerapan *data mining* menggunakan metode Klasifikasi yang menghasilkan pohon keputusan dengan mengubah pola data (table) menjadi pohon dan mengubah model pohon menjadi *rule*, dan menyederhanakan *rule* yang dapat mudah dipahami. Manfaat utama dari penggunaan pohon keputusan adalah kemampuannya untuk *membreak down* proses pengambilan keputusan yang *kompleks* menjadi lebih simpel sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan. Pohon Keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target.

Pohon keputusan merupakan representasi sederhana dari teknik klasifikasi untuk sejumlah kelas berhingga, dimana simpul internal maupun simpul akar ditandai dengan nama atribut, rusuk-rusuknya diberi label nilai atribut yang mungkin dan simpul daun ditandai dengan kelas-kelas yang berbeda.

BAB VI

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Berdasarkan dari penelitian yang telah dilaksanakan dan sudah diuraikan dalam penerapan *data mining* dari data transaksi nasabah tepatnya pada tahun 2010 sebanyak 1610 nasabah untuk menentukan kriteria calon nasabah potensial pada AJB Bumiputera 1912 Palembang, maka penulis dapat menarik kesimpulan sebagai berikut :

1. Penerapan *data mining* dengan teknik *decision tree* dan algoritma *C4.5* yang dilakukan menghasilkan sebuah informasi mengenai profil nasabah dalam menentukan kriteria calon nasabah potensial pada AJB Bumiputera 1912 palembang.
2. Dalam penerapan *data mining* ini dapat memberikan informasi profil nasabah dengan kriteria income dan umur yang memilih jenis asuransi apa saja sehingga dapat memberikan informasi untuk calon nasabah berikutnya, dan suatu keputusan atau suatu pertimbangan pada AJB Bumiputera 1912 untuk kedapannya yang lebih baik lagi.
3. Perhitungan yang dilakukan secara teoritis dan aplikatif menghasilkan sebuah pohon keputusan yang ditentukan dalam penerapan *data mining.*

4. *Decision tree* yang dihasilkan telah mampu menunjukkan keterkaitan suatu jenis asuransi dengan identitas nasabah seperti yang tercatat pada rekam hasil penerapan *data mining*.
5. Pemecahan *field* income dan umur menjadi beberapa kelompok kecil membantu *user* dalam mengetahui informasi yang dihasilkan dari pohon keputusan.

6.2 Saran

Berdasarkan hasil dan kesimpulan yang telah diuraikan diatas, maka ada beberapa saran yang ingin disampaikan yaitu:

1. Dengan penerapan *data mining* yang telah dihasilkan, pihak AJB Bumiputera 1912 Palembang dapat memanfaatkan informasi dari hasil penerapan *data mining* untuk menentukan calon nasabah potensial berdasarkan kriteria nasabah lama.
2. Pada penelitian selanjutnya dapat mencoba menggunakan *dataset* yang berbeda dan dengan jumlah data yang lebih besar lagi sehingga nilai data selanjutnya yang dihasilkan dapat menghasilkan tingkat akurasi yang lebih tinggi.
3. Selain penerapan secara teoritis dan aplikatif, pada penelitian berikutnya dapat dicoba untuk membuat suatu aplikasi dengan teknik dan algoritma *data mining* yang berbeda sehingga dapat menghasilkan informasi yang pariatif.
4. Penelitian ini disarankan dapat menjadi bahan referensi yang dipergunakan dan dikembangkan untuk penenlitian selanjutnya.

DAFTAR PUSTAKA

Azevedo, A. Santos & Manuel F . (2008) , *KDD, SEMMA AND CRISP-DM: A PARALLEL OVERVIEW* , IADIS. ISBN: 978-972-8924-63-8.

DTREG. (2013). *"Software For Predictive Modeling and Forecasting"*, *http://www.dtreg.com/*, diakses 1 juni 2013

Eka Sabna, (2010). *Aplikasi Data Mining Untuk Menganalisis Track Record Penyakit Pasien Dengan Menggunakan Teknik Decision Tree,* Universitas Putra Indonesia "YPTK" Padang

Hermawati. F. Astuti. (2013). *Data Mining.* Yogyakarta: Andi Offset.

Kusrini & Luthfi. E. Taufiq. (2009). *Algoritma Data Mining.* Yogyakarta: Andi Offset.

Larose, Daniel T . (2005) , *Discovering Knowledge in Data: An Introduction to Data Mining* , John Willey & Sons, Inc.

Pramudiono, I . (2003). Pengantar *Data Mining*, *http://ikc.depsos.go.id /umum/iko-datamining.php*, diakses tgl 10 April 2013

Santosa, Budi 2007, *"Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis"*, Graha Ilmu, Yogyakarta

Wibowo, Ari. (2011). *Prediksi Nasabah Potensial Menggunakan Metode Klasifikasi Pohon Biner:* Universitas Politeknik Negri Batam.