

**ANALISIS KOMPARASI  
ALGORITMA KLASIFIKASI DATA MINING  
PADA AKSEPTASI DATA FAKULTATIF  
REASURANSI JiWA**

**TESIS**

**Diajukan Sebagai Salah Satu Syarat Untuk Menyelesaikan  
Program Strata Dua (S2) Magister Komputer**



**OLEH:**

**FERRY FEBRIAN**

**341208749**

**PROGRAM STUDI TEKNIK INFORMATIKA  
PROGRAM PASCA SARJANA (S2) MAGISTER KOMPUTER  
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER ERESHA  
JAKARTA  
2011**

## PERSETUJUAN TESIS

Nama : Ferry Febrian  
NPM : 341208749  
Konsentrasi : Software Engineering  
Judul Tesis : Analisis Komparasi Algoritma Klasifikasi *Data Mining*  
Pada Akseptasi Data Fakultatif Reasuransi Jiwa

Telah disetujui untuk disidangkan pada Sidang Tesis pada Program Pasca Sarjana (S2) Magister Komputer, Program Studi Teknik Informatika Sekolah Tinggi Manajemen Informatika dan Komputer Eresha.

Jakarta, 30 September 2011

Pembimbing Utama

Pembimbing Pendamping

(Romi Satria Wahono, B.Eng., M.Eng.)

(Dr. Rufman Iman Akbar E., MM, M.Kom.)

Mengetahui :

Ketua

Direktur

Program Pasca Sarjana

(Ir. Damsiruddin Siregar, MMT.)

(Dr. Rufman Iman Akbar E., MM., M.Kom.)

## PENGESAHAN TESIS

Nama : Ferry Febrian  
NPM : 341208749  
Konsentrasi : Software Engineering  
Judul Tesis : Analisis Komparasi Algoritma Klasifikasi *Data Mining*  
Pada Akseptasi Data Fakultatif Reasuransi Jiwa

Telah disidangkan dan dinyatakan Lulus Sidang Tesis pada Program Pasca Sarjana (S2) Magister Komputer, Program Studi Teknik Informatika Sekolah Tinggi Manajemen Informatika dan Komputer Eresha pada tanggal 30 September 2011.

Nama Penguji

Tanda Tangan

Romi Satria Wahono, B.Eng., M.Eng  
(Ketua)

-----

Didik Setiyadi, M.Kom  
(Sekretaris)

-----

Bobby Reza, S.Kom, MM  
(Anggota)

-----

Mengetahui:

Direktur

Program Pasca Sarjana

(Dr. Rufman Iman Akbar E., MM., M.Kom.)

## PERNYATAAN KEASLIAN TESIS

Nama : Ferry Febrian  
NPM : 341208749  
Konsentrasi : Software Engineering  
Judul Tesis : Analisis Komparasi Algoritma Klasifikasi *Data Mining*  
Pada Akseptasi Data Fakultatif Reasuransi Jiwa

Dengan ini saya menyatakan bahwa dalam Tesis ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Jakarta, 30 September 2011

(Ferry Febrian)

Ferry Febrian, 341208749

**Analisis Komparasi Algoritma Klasifikasi *Data Mining* Pada Akseptasi Data Fakultatif Reasuransi Jiwa; dibawah bimbingan Romi Satria Wahono, B.Eng, M.Eng dan Dr. Rufman Iman Akbar E, MM, M.Kom.**

**145+ xii hal / 34 tabel / 31 gambar / 3 lampiran / 18 pustaka (2002-2010)**

## **ABSTRAK**

Bisnis fakultatif merupakan bisnis yang memiliki tingkat resiko yang cukup tinggi. Resiko ini bisa terjadi dengan meningkatnya jumlah klaim dengan nilai yang cukup besar sehingga akan berdampak pada terganggunya kestabilan finansial perusahaan. Perusahaan berkeinginan untuk dapat menghindari atau mengurangi nilai-nilai klaim yang semestinya tidak perlu terjadi. Untuk itu dibutuhkan suatu metode yang akurat untuk dapat seminimal mungkin menerima polis-polis asuransi yang diprediksi memiliki resiko klaim yang besar.

Untuk mewujudkan hal tersebut akan dilakukan evaluasi pada akseptasi data fakultatif reasuransi jiwa. Dalam melakukan evaluasi ini akan menggunakan salah satu teknik dalam *Data Mining* yaitu teknik klasifikasi. Teknik ini dipilih karena menyesuaikan dengan tujuan dari kasus tersebut yaitu memutuskan apakah menerima atau menolak suatu data polis fakultatif yang ditawarkan perusahaan asuransi.

Agar didapatkan algoritma yang paling akurat akan dilakukan komparasi beberapa algoritma yang memiliki perbedaan karakteristik antara lain adalah *Logistic Regression*, *Decision Trees*, *K-Nearest Neighbors*, *Naïve Bayes* dan *Support Vector Machines*.

Hasil analisis komparasi yang diperoleh dengan menggunakan pengujian *Confusion Matrix*, *ROC Curve* dan *Paired T-Test* dapat disimpulkan bahwa metode atau algoritma yang paling akurat adalah algoritma *Decision Trees*, kemudian algoritma *Support Vector Machines* dengan nilai uji T-Test yang dominan dan tingkat *accuracy* yang cukup tinggi. Sedangkan algoritma *Logistic Regression* cukup baik dengan nilai AUC yang tinggi, *K-Nearest Neighbors* dan *Naïve Bayes* memiliki kinerja yang kurang memuaskan.

**Kata kunci:** Fakultatif, Komparasi, Klasifikasi, *Data Mining*, *Logistic Regression*, *Decision Trees*, *K-Nearest Neighbors*, *Naïve Bayes*, *Support Vector Machines*.

Ferry Febrian, 341208749

Analysis Comparison of Classification Algorithms *Data Mining* In Data Acceptance  
Facultative Life Reinsurance, Under the guidance of Romi Satria Wahono, B.Eng,  
M.Eng and Dr. Rufman Iman Akbar E, MM, M.Kom.

145 + xii page / 34 table / 31 picture / 3 enclosure / 18 reference (2001-2010)

## ABSTRACT

Facultative business is a business that has a fairly high level of risk. This risk can occur with the increasing number of claims with a value large enough so it will have an impact on the disruption of the company's financial stability. The company wanted to be able to avoid or reduce the values of claims that should not have happened. That requires an accurate method to be as minimum as possible receive the insurance policies that were predicted risk of large claims.

To achieve this will be evaluated on the data acceptance facultative life reinsurance. In conducting this evaluation will use one of the techniques in *Data Mining* classification techniques. This technique was chosen because it fit with the objectives of the case is to decide whether to accept or reject a policy of data facultative insurance company has to offer.

In order to obtain the most accurate algorithm will do a comparison few algorithms that have different characteristics, among others, are *Logistic Regression*, *Decision Trees*, *K-Nearest Neighbors*, *Naïve Bayes* dan *Support Vector Machines*.

The results of analysis comparison obtained by using test *Confusion Matrix*, ROC Curve and *Paired T-Test* can be concluded that the method or the most accurate algorithm is *Decision Trees*, then the algorithm *Support Vector Machines* with a test value T-Test is dominant and relatively high level of accuracy. While the *Logistic Regression* algorithm is quite good with a high AUC value, *K-Nearest Neighbors* and Naive Bayes has a less satisfactory performance.

**keyword:** Facultative, Comparasion, Classification, *Data Mining*, *Logistic Regression*, *Decision Trees*, *K-Nearest Neighbors*, *Naïve Bayes*, *Support Vector Machines*.

## KATA PENGANTAR

Puji syukur penulis ucapkan kepada Tuhan Yang Maha Esa yang telah melimpahkan segala rahmat dan hidayah-NYA, sehingga tersusunlah Tesis yang berjudul **“Analisis Komparasi Algoritma Klasifikasi Data Mining Pada Akseptasi Data Fakultatif Reasuransi Jiwa”**.

Tesis ini melengkapi salah satu persyaratan dalam rangka menyelesaikan perkuliahan untuk memperoleh gelar Magister Komputer (M. Kom.) pada Program Pasca Sarjana (S2), Program Studi Teknik Informatika Sekolah Tinggi Manajemen Informatika Eresha Jakarta.

Penulis sangat menyadari bahwa penulisan Tesis ini tidak akan bisa terwujud tanpa adanya dukungan dan bantuan dari berbagai pihak. Maka dari itu dalam kesempatan ini penulis ingin mengucapkan terima kasih dan penghargaan yang sebesar-besarnya kepada yang terhormat:

1. Bapak Ir. Damsiruddin Siregar, MMT, selaku Ketua Pasca Sarjana STMIK Eresha, Jakarta, yang telah menyediakan wadah untuk dapat menimba ilmu dalam program Pasca Sarjana di kampus tercinta ini.
2. Bapak Romi Satria Wahono, B.Eng, M.Eng, sebagai Pembimbing Utama, yang telah banyak mengarahkan, membimbing dan memberikan materi-materi pengajaran dalam menyelesaikan tesis ini.
3. Bapak Dr. Rufman Iman Akbar E, MM, M.Kom, sebagai Direktur Program Pasca Sarjana STMIK Eresha, Jakarta, dan juga sebagai Pembimbing Pendamping yang telah banyak memberikan masukan dalam menyempurnakan penyusunan tesis ini.
4. Istri, anak dan orang tuaku tercinta yang telah memberikan kesempatan, waktu luang dan do'a serta menjadi inspirator dan motivator dalam penyelesaian tesis ini.
5. Rekan-rekan kerja di PT. ReINDO yang telah memberikan masukan tentang masalah perasuransian yang menjadi bahan penelitian.
6. Rekan-rekan Angkatan 34 yang telah banyak memberikan motivasi belajar serta dukungan moril dalam menyelesaikan Tesis ini.

7. Seluruh dosen penguji (proposal, kolokium dan tesis) serta para staf di STMIK Eresha yang memberikan masukan dan saran serta kemudahan dalam proses pembelajaran dan administrasi akademik.

Dengan segala keterbatasan penulis maka akan terjadi banyak kesalahan dan kekurangan dalam penelitian ini. Oleh karena itu kritik, saran dan masukan yang membangun akan sangat membantu penulis dalam memperbaiki penelitian ini.

Penulis



# DAFTAR ISI

	Hal.
PERSETUJUAN TESIS.....	i
PENGESAHAN TESIS.....	ii
PERNYATAAN KEASLIAN TESIS.....	iii
ABSTRAK.....	iv
ABSTRACT.....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Permasalahan Penelitian.....	3
1.2.1 Identifikasi Masalah.....	3
1.2.2 Ruang Lingkup Masalah.....	3
1.2.3 Rumusan Masalah.....	4
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	4
1.5 Sistematika Penulisan.....	5
BAB II LANDASAN TEORI DAN KERANGKA PEMIKIRAN.....	6
2.1 Tinjauan Studi.....	6
2.2 Tinjauan Pustaka.....	7
2.2.1 Reasuransi.....	7
2.2.1.1 Underwriting.....	8
2.2.1.2 Fakultatif.....	10
2.3 <i>Data Mining</i> .....	10
2.3.1 Algoritma <i>Logistic Regression</i> .....	11
2.3.2 Algoritma <i>Decision Tree</i> .....	15
2.3.3 Algoritma <i>k-Nearest Neighbors</i> .....	22
2.3.4 Algoritma <i>Naïve Bayes</i> .....	27
2.3.5 Algoritma <i>Support Vector Machine</i> .....	29

2.4 Evaluasi Algoritma Klasifikasi <i>Data Mining</i> .....	30
2.4.1 Confusion Matrix.....	31
2.4.2 ROC Curve.....	32
2.4.3 Paired T-Test.....	34
2.5 Kerangka Pemikiran.....	35
BAB III METODE PENELITIAN.....	36
3.1 Business Understanding.....	38
3.2 Data Understanding.....	39
3.3 Data Preparation.....	40
3.4 Modelling.....	42
3.4.1 K-Fold Cross Validation.....	45
3.4.2 Analisis Pengujian Menggunakan <i>Logistic Regression</i> .....	46
3.4.3 Analisis Pengujian Menggunakan <i>Decision Tree</i> .....	48
3.4.4 Analisis Pengujian Menggunakan K-Nearest Neighbor.....	52
3.4.5 Analisis Pengujian Menggunakan <i>Naïve Bayes</i> .....	59
3.4.6 Analisis Pengujian Menggunakan <i>Support Vector Machine</i> .....	63
BAB IV HASIL DAN PEMBAHASAN.....	66
4.1 Evaluasi dan Validasi Hasil Penelitian.....	66
4.1.1 Pengujian Model <i>Logistic Regression</i> .....	67
4.1.2 Pengujian Model <i>Decision Tree</i> .....	68
4.1.3 Pengujian Model <i>K-Nearest Neighbors</i> .....	69
4.1.4 Pengujian Model <i>Naïve Bayes</i> .....	70
4.1.5 Pengujian Model <i>Support Vector Machine</i> .....	71
4.2 Performance Komparasi.....	72
4.3 Analisis Hasil Komparasi.....	74
4.4 Implikasi Penelitian .....	77
BAB V KESIMPULAN DAN SARAN.....	79
5.1 Kesimpulan.....	79
5.2 Saran.....	80
DAFTAR PUSTAKA.....	81
DAFTAR RIWAYAT HIDUP.....	83
LAMPIRAN-LAMPIRAN.....	84

## DAFTAR TABEL

	Hal.
Tabel 2.1 Data pengadopsi layanan telepon system baru.....	13
Tabel 2.2 Data berformat diskrit.....	14
Tabel 2.3 Perbandingan data aktual dan prediksi.....	14
Tabel 2.4 Penggunaan sampel data yang berbeda.....	15
Tabel 2.5 Keputusan Main Tennis.....	17
Tabel 2.6 Perhitungan Node 1.....	18
Tabel 2.7 Perhitungan Node 1.1.....	20
Tabel 2.8 Perhitungan Node 1.1.2.....	21
Tabel 2.9 Tabel Kasus Nesrest Neighbor.....	24
Tabel 2.10 Bobot Atribut.....	24
Tabel 2.11 Kedekatan Nilai Atribut jenis Kelamin.....	25
Tabel 2.12 Kedekatan Nilai Atribut Pendidikan.....	25
Tabel 2.13 Kedekatan Nilai Atribut Agama.....	25
Tabel 2.14 Tabel konsumen AllElectronics.....	28
Tabel 2.15 Tabel <i>Confusion Matrix</i> .....	31
Tabel 3.1 Atribut dan data fakultatif.....	39
Tabel 3.2 Ilustrasi anomali dan inkonsisten data.....	40
Tabel 3.3 Ilustrasi anomali data.....	40
Tabel 3.4 Ilustrasi inkonsisten data.....	40
Tabel 3.5 Hasil <i>reduction</i> dan <i>cleaning</i> data.....	41
Tabel 3.6 Pemetaan Data.....	41
Tabel 3.7 Pemetaan data dengan <i>discretize by frequency</i> .....	43
Tabel 3.8 Data hasil <i>discretize by frequency</i> .....	43
Tabel 3.9 Pengkategorisasian atribut.....	44
Tabel 3.10 Data pengujian menggunakan <i>Logistic Regression</i> .....	47
Tabel 3.11 Perhitungan entropy <i>Decision Tree</i> .....	48
Tabel 3.12 Data pengujian menggunakan <i>K-Nearest Neighbors</i> .....	53
Tabel 3.13 Data pengujian menggunakan <i>Naïve Bayes</i> .....	61
Tabel 3.14 Tabel probabilitas atribut <i>Naïve Bayes</i> .....	61
Tabel 3.15 Data pengujian menggunakan <i>Support Vector Machine</i> .....	64
Tabel 4.1 Perbandingan <i>Performance</i> Algoritma.....	72

Tabel 4.2 Perbandingan <i>Performance</i> untuk <i>Excellent Classification</i> .....	74
Tabel 4.3 Uji Statistik T-Test.....	75
Tabel 4.4 Hasil Perbandingan Seluruh Pengujian.....	77

## DAFTAR GAMBAR

	Hal.
Gambar 1.1 Grafik resiko klaim fakultatif reasuransi jiwa.....	2
Gambar 2.1 Perbedaan Regresi Logistik dan Regresi Linier.....	11
Gambar 2.2 Proses <i>Decision Tree</i> .....	15
Gambar 2.3 Bentuk simpul <i>Decision Tree</i> .....	16
Gambar 2.4 Pohon Keputusan Node 1.....	20
Gambar 2.5. Pohon Keputusan Hasil Perhitungan Node 1.1.....	21
Gambar 2.6 Pohon Keputusan Hasil Perhitungan Node 1.1.2.....	22
Gambar 2.7 Ilustrasi tetangga terdekat.....	23
Gambar 2.8a Bidang pemisah .....	30
Gambar 2.8b Bidang pemisah terbaik.....	30
Gambar 2.9 Kurva ROC.....	33
Gambar 3.1 Proses CRIPSP-DM .....	38
Gambar 3.2 Grafik peningkatan klaim.....	38
Gambar 3.3 Ilustrasi 10-Fold <i>Cross Validation</i> .....	46
Gambar 3.4 Desain model <i>Logistic Regression</i> .....	47
Gambar 3.5 Desain model <i>Decision Tree</i> .....	50
Gambar 3.6 Pohon hasil perhitungan <i>Decision Tree</i> .....	50
Gambar 3.7 Desain model <i>Rule Decision Tree</i> .....	51
Gambar 3.8 Desain model <i>K-Nearest Neighbors</i> .....	53
Gambar 3.9 Desain model <i>Naïve Bayes</i> .....	60
Gambar 3.10 Desain model <i>Support Vector Machine</i> .....	64
Gambar 4.1 Desain model validasi.....	66
Gambar 4.2 Grafik Area Under Curve <i>Logistic Regresion</i> .....	67
Gambar 4.3 Grafik Area Under Curve <i>Decision Tree</i> .....	68
Gambar 4.4 Grafik Area Under Curve <i>k-Nearest Neighbors</i> .....	69
Gambar 4.5 Grafik Area Under Curve <i>Naïve Bayes</i> .....	70
Gambar 4.6 Grafik Area Under Curve <i>Support Vector Machine</i> .....	71
Gambar 4.7 Desain model komparasi menggunakan <i>ROC's Curve</i> .....	72
Gambar 4.8 Modul algoritma klasifikasi pada modul <i>Compare ROCs</i> .....	73
Gambar 4.9 Grafik komparasi <i>ROC Curve</i> .....	73
Gambar 4.10 Model T-Test pada <i>Logistic Regression</i> dan <i>Decision Trees</i> ...	75

## DAFTAR LAMPIRAN

	Hal
Lampiran 1 Data mentah (sebelum <i>reduction</i> dan <i>cleaning</i> ) .....	84
Lampiran 2 Data hasil <i>reduction</i> dan <i>cleaning</i> data .....	109
Lampiran 3 Data hasil <i>discretize by frequency</i> .....	134

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

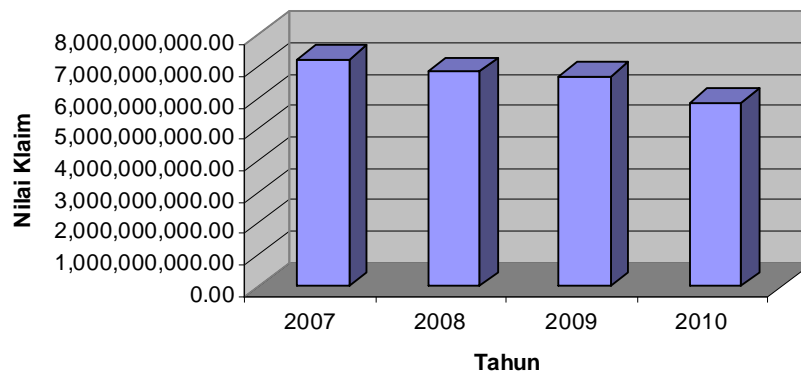
Bisnis fakultatif merupakan bisnis yang memiliki tingkat resiko yang cukup tinggi. Resiko ini bisa terjadi dengan meningkatnya jumlah klaim dengan nilai yang cukup besar sehingga akan berdampak pada terganggunya kestabilan finansial perusahaan. Perusahaan berkeinginan untuk dapat menghindari atau mengurangi nilai-nilai klaim yang semestinya tidak perlu terjadi. Untuk itu dibutuhkan suatu metode yang akurat untuk dapat seminimal mungkin menerima polis-polis asuransi yang diprediksi memiliki resiko klaim yang besar.

Fakultatif (Sula, 2004: p147) adalah salah satu metode bisnis reasuransi dalam menerima pertanggungan ulang dari pihak asuransi. Fakultatif merupakan suatu perjanjian dimana masing-masing pihak (asuransi dan reasuransi) memiliki kebebasan dalam menetapkan suatu keputusan. Pihak asuransi bebas menentukan apakah akan mereasuransikan resikonya atau tidak mereasuransikan, sedangkan pihak reasuransi bebas menentukan pilihan apakah menerima atau menolak resiko yang akan ditawarkan oleh asuransi. Pada metode fakultatif, tiap resiko ditawarkan secara individu (resiko per resiko) kepada pihak reasuransi dan pihak asuransi berkewajiban untuk melakukan *full disclosure* kepada pihak reasuransi tentang fakta-fakta material yang berkenaan dengan pokok pertanggungan, *terms and condition*, dan informasi lainnya yang dipandang perlu dalam mempertimbangkan akseptasi reasuransi itu. Dalam metode reasuransi secara fakultatif, perusahaan reasuransi dapat melakukan seleksi resiko sesuai dengan kebijakan *underwriting* (ReINDO, PT., 2002: p2) yang telah ditentukan. Hal ini dapat dipahami karena tingkat resiko dari tertanggung ini berbeda-beda. Oleh karena itu dalam prakteknya banyak pertimbangan yang digunakan untuk mengambil keputusan menerima atau menolak tawaran reasuransi fakultatif tersebut.

Menurut KUHD pasal 271, disebutkan bahwa reasuransi adalah asuransi dari asuransi atau asuransinya asuransi. Bisnis reasuransi merupakan suatu persetujuan yang dilakukan antara dua pihak yang disebut pemberi sesi

(perusahaan asuransi) dan penanggung ulang (perusahaan reasuransi) dimana perusahaan asuransi menyetujui untuk menyerahkan dan perusahaan reasuransi menyetujui untuk menerima resiko yang telah ditentukan dengan persyaratan yang ditetapkan dalam perjanjian. Karena perusahaan reasuransi adalah “*backup*” dari perusahaan asuransi maka dibutuhkan penyeleksian data fakultatif yang ketat dalam membuat keputusan apakah menerima atau menolak tawaran tersebut. Tugas dari penyeleksian data fakultatif ini ada pada *underwriter*. Untuk membantu *underwriter* dalam menganalisis data fakultatif ini dibutuhkan suatu teknik yang dapat mengklasifikasikan apakah data tersebut layak diterima atau sebaiknya ditolak karena memiliki resiko tinggi sehingga akan berpotensi menjadi resiko klaim yang besar.

Berdasarkan dari hasil laporan keuangan untuk bisnis fakultatif reasuransi jiwa pada PT. ReINDO diketahui bahwa resiko klaim yang terjadi pada tahun 2007 sebesar Rp. 7,237,954,991.24, pada tahun 2008 sebesar Rp. 6,845,758,625.24, pada tahun 2009 sebesar Rp. 6,690,765,255.29 dan pada tahun 2010 sebesar Rp. 5,854,740,225.7.



**Gambar 1.1**  
Grafik resiko klaim fakultatif reasuransi jiwa  
PT. ReINDO (2007 – 2010)

Pada gambar 1.1 diatas dapat kita lihat klaim yang terjadi dengan jumlah yang cukup besar. Untuk dapat mengurangi jumlah resiko klaim yang terjadi, akan dilakukan evaluasi pada akseptasi data-data fakultatif reasuransi jiwa. Dalam melakukan evaluasi ini akan menggunakan salah satu teknik dalam *data mining*



yaitu teknik klasifikasi. Teknik ini dipilih karena menyesuaikan dengan tujuan dari kasus tersebut yaitu memutuskan apakah menerima atau menolak suatu data polis fakultatif yang ditawarkan perusahaan asuransi..

Ada beberapa penelitian teknik klasifikasi *data mining* yang berhubungan dengan asuransi atau jasa keuangan diantaranya adalah yang dilakukan oleh Stijn Viaene, Richard A Derrig, Bart Baesens dan Guido Dedene pada tahun 2002 yang melakukan evaluasi dan komparasi algoritma klasifikasi *data mining* untuk mendeteksi penipuan klaim asuransi *expert automobile*. Algoritma klasifikasi yang dikomparasi adalah *Logistic Regression*, *C4.5 Decision Tree*, *K-Nearest Neighbors*, *Bayesian learning multilayer perceptron neural network*, *least-squares support vector machine*, *Naïve Bayes*, dan *TAN Technique*. Algoritma yang paling akurat dari komparasi tersebut adalah algoritma C4.5.

Pada kesempatan ini penelitian yang akan dilakukan adalah menganalisis komparasi algoritma klasifikasi *data mining*, diantaranya adalah algoritma *Logistic Regression*, algoritma *Decision Tree*, algoritma *K-Nearest Neighbors*, algoritma *Naïve Bayes* dan algoritma *Support Vector Machine* untuk menangani kasus akseptasi data fakultatif reasuransi jiwa dengan tujuan agar algoritma yang didapat merupakan algoritma yang paling akurat sehingga dapat melakukan seleksi data dengan tepat.

## **1.2 Permasalahan Penelitian**

### **1.2.1 Identifikasi Masalah**

Identifikasi permasalahan pada penelitian ini adalah sebagai berikut:

- a. Belum diketahuinya metode menganalisis komparasi algoritma-algoritma klasifikasi *data mining* pada akseptasi data fakultatif reasuransi jiwa
- b. Belum diketahuinya algoritma klasifikasi *data mining* yang akurat pada akseptasi data fakultatif reasuransi jiwa.

### **1.2.2 Ruang Lingkup Masalah**

Ruang lingkup permasalahan pada penelitian ini dibatasi pada bagaimana menganalisis komparasi model-model teknik klasifikasi *data mining* untuk algoritma *Logistic Regression*, *Decision Trees*, *K-Nearest Neighbors*, *Naïve Bayes*

dan *Support Vector Machines* yang akan diterapkan pada kasus akseptasi data fakultatif reasuransi jiwa.

### **1.2.3 Rumusan Masalah**

Rumusan masalah yang berdasarkan kepada identifikasi masalah dan ruang lingkup masalah yaitu algoritma klasifikasi *data mining* apa yang paling akurat pada akseptasi data fakultatif reasuransi jiwa.

## **1.3 Tujuan dan Manfaat Penelitian**

### **1.3.1 Tujuan Penelitian**

Tujuan dari penelitian ini adalah untuk mendapatkan algoritma yang paling akurat dalam menangani kasus akseptasi data fakultatif reasuransi jiwa. Algoritma-algoritma yang digunakan adalah *Logistic Regression*, *Decision Trees*, *K-Nearest Neighbors*, *Naïve Bayes* dan *Support Vector Machines*.

### **1.3.2 Manfaat Penelitian**

Manfaat yang dapat diambil dari penelitian ini adalah:

- a. Manfaat praktis, dapat dimanfaatkan oleh *underwriter* sebagai alat bantu dalam membuat keputusan untuk menentukan akseptasi data fakultatif reasuransi jiwa.
- b. Manfaat teoritis, dapat memberikan kontribusi pada metode *data mining* dalam membandingkan atau mengkomparasikan beberapa algoritma klasifikasi *data mining* untuk kasus-kasus yang belum pernah dilakukan komparasi atau pada kasus-kasus yang sudah pernah dilakukan komparasi tetapi dengan algoritma yang berbeda.
- c. Manfaat kebijakan, dapat membantu Departemen Underwriter Reasuransi Jiwa dalam menetapkan suatu kebijakan kepada perusahaan asuransi jiwa tertentu saat memberikan penawaran bisnis fakultatif.

#### **1.4 Sistematika Penulisan**

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

##### **BAB I: PENDAHULUAN**

Bab ini mencakup beberapa sub bab antara lain: latar belakang, masalah penelitian, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

##### **BAB II: LANDASAN TEORI DAN KERANGKA PEMIKIRAN**

Bab ini mencakup beberapa subbab yaitu: tinjauan studi, tinjauan pustaka, teori-teori yang mendukung topik dan kerangka pemikiran.

##### **BAB III: METODE PENELITIAN**

Bab ini mencakup beberapa sub bab antara lain: analisa kebutuhan, perancangan penelitian dan teknik analisis.

##### **BAB IV: HASIL DAN PEMBAHASAN**

Bab ini mencakup beberapa sub bab antara lain: analisis data dan metode, evaluasi data dan metode, komparasi metode, pengujian metode.

##### **BAB V: KESIMPULAN DAN SARAN**

Bab ini berisi kesimpulan dari penelitian ini dan memberikan saran untuk pengembangan penelitian lebih lanjut.

## **BAB II**

### **LANDASAN TEORI DAN KERANGKA PEMIKIRAN**

#### **2.1 Tinjauan Studi**

Tinjauan studi ini diambil dari beberapa jurnal dan artikel penelitian yang berhubungan dengan komparasi algoritma *data mining* teknik klasifikasi yang memiliki keterkaitan dalam topik penelitian.

##### **a. Application and Comparison of Classification Techniques in Controlling Credit Risk**

Penelitian yang dilakukan oleh Lan Yu, Guoqing Chen, Andy Koronios, Shiwu Zhu, Xunhua Guo (Yu, Chen, Koronios, Zhu, & Guo, 2007: p1) adalah tentang:

1. Kasus penelitiannya adalah penentuan algoritma terbaik untuk menentukan prediksi rating kredit.
2. Algoritma yang dikomparasi ada sepuluh algoritma yaitu LDA, QDA, *Logistic Regression*, *K-Nearest Neighbors*, *Naïve Bayes*, *The TAN Technique*, *Decision Tree*, *Associative Classification*, *Artificial Neural Network*, dan *Support Vector Machines*.
3. Metode analisis dan komparasi yang digunakan adalah *Confusion Matrix*, *ROC Curve*, dan *Delong-Pearson statistical method*.
4. Hasil komparasi dengan algoritma terbaik adalah algoritma *Associative Classification* (CBA).

##### **b. A Comparison Of State-Of-The-Art Classification Techniques For Expert Automobile Insurance Claim Fraud Detectio**

Penelitian yang dilakukan oleh Stijn Viaene, Richard A Derrig, Bart Baesens, Guido Dedene (Viaene, Derrig, Baesens, & Dedene, 2002: p1) adalah tentang: