

**KAJIAN PENERAPAN ALGORITMA C4.5 DALAM REKRUITMEN NASABAH
BANK**



TESIS

**SAIFUDIN
14000418**

**PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER NUSA
MANDIRI
JAKARTA
2012**

**KAJIAN PENERAPAN ALGORITMA C4.5 DALAM REKRUITMEN
NASABAH BANK**



TESIS

Diajukan sebagai salah satu syarat untuk memperoleh gelar
Magister Ilmu Komputer (M.Kom)

SAIFUDIN
14000418

PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
NUSA MANDIRI
JAKARTA
2012

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini :

Nama : Saifudin
NIM : 14000418
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *e-Business*

Dengan ini menyatakan bahwa tesis yang telah saya buat dengan judul: “Kajian Penerapan Algoritma C4.5 Dalam Rekrutmen Nasabah Bank” adalah hasil karya sendiri, dan semua sumber baik yang kutip maupun yang dirujuk telah saya nyatakan dengan benar dan tesis belum pernah diterbitkan atau dipublikasikan dimanapun dan dalam bentuk apapun.

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

Jakarta, April 2012

Yang menyatakan,

Materai Rp. 6.000,-

Saifudin

HALAMAN PENGESAHAN

Tesis ini diajukan oleh:

Nama : Saifudin
NIM : 14000418
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : e-Business
Judul Tesis : “Kajian Penerapan Algoritma C4.5 Dalam Rekrutmen Nasabah Bank”

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

Jakarta, September 2012
Pascasarjana Magister Ilmu Komputer
STMIK Nusa Mandiri
Direktur



Prof. Dr. Ir. Kaman Nainggolan, MS

DEWAN PENGUJI

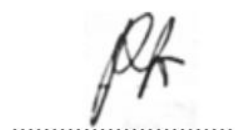
Penguji I : Dr. Sularso Budilaksono



Penguji II : Mochammad Wahyudi, MM, M.Kom, M.Pd



Penguji III/
Pembimbing : Dr. Ir. Prabowo Pudjo Widodo, MS



	Lembar Konsultasi Bimbingan Tesis
	Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri

- NIM : 14000418
- Nama Lengkap : Saifudin
- Dosen Pembimbing : Dr. Ir. Prabowo Pudjo Widodo, MS
- Judul Tesis : Kajian Penerapan Algoritma C4.5 Dalam Rekrutmen Nasabah Bank

No	Tanggal Bimbingan	Materi Bimbingan	Pembimbing Paraf Dosen
1	12 Mei 2012	Pengajuan Judul	<i>PK</i>
2	12 Mei 2012	Bab I	<i>PK</i>
3	26 Mei 2012	ACC Bab I, Pengajuan Bab II	<i>PK</i>
4	26 Mei 2012	Kuesioner	<i>PK</i>
5	9 Juni 2012	Bab III	<i>PK</i>
6	9 Juni 2012	Penataan Data di Excel	<i>PK</i>
7	7 Juli 2012	Olah Data	<i>PK</i>
8	7 Juli 2012	Pengajuan Bab IV	<i>PK</i>
9	4 Agustus 2012	ACC Bab IV	<i>PK</i>
10	4 Agustus 2012	Pengajuan Bab V, Daftar Pustaka	<i>PK</i>
11	13 Agustus 2012	Lampiran-lampiran	<i>PK</i>
12	13 Agustus 2012	PPT	<i>PK</i>

Catatan :

Total bimbingan yang harus dilakukan adalah 6 (enam) kali pertemuan

- Bimbingan dimulai pada tanggal : 12 Mei 2012
- Bimbingan diakhiri pada tanggal : 13 Agustus 2012
- Jumlah pertemuan : 12 kali pertemuan

Yogyakarta, 13 Agustus 2012

Dosen Pembimbing



Dr. Ir. Prabowo Pudjo Widodo, MS

KATA PENGANTAR

Puji syukur alhamdulillah, penulis haturkan kehadiran Allah, SWT, yang telah melimpahkan rahmat dan karunia-Nya, sehingga pada akhirnya penulis dapat menyelesaikan tesis ini tepat pada waktunya. Dimana tesis ini penulis sajikan dalam bentuk buku yang sederhana. Adapun judul tesis, yang penulis ambil sebagai berikut “Kajian Penerapan Algoritma C4.5 Dalam Rekrutmen Nasabah Bank”.

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

Tesis ini diambil berdasarkan hasil penelitian atau riset mengenai Kajian Penerapan Algoritma C4.5 Dalam Rekrutmen Nasabah Bank. Penulis juga lakukan mencari dan menganalisa berbagai macam sumber referensi, baik dalam bentuk jurnal ilmiah, buku-buku literatur, *internet*, dan lain-lain yang terkait dengan pembahasan pada tesis ini.

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

1. Bapak Dr. Ir. Prabowo Pudjo Widodo, MS selaku pembimbing tesis yang telah menyediakan waktu, pikiran dan tenaga dalam membimbing penulis dalam menyelesaikan tesis ini.
2. Orang tua tercinta yang telah memberikan dukungan material dan moral kepada penulis.
3. Istri dan anak saya tercinta yang juga telah memberikan dukungan dalam pembuatan tesis.
4. Seluruh staf pengajar (dosen) PPs MIK STMIK Nusa Mandiri yang telah memberikan pelajaran yang berarti bagi penulis selama menempuh studi.
5. Seluruh staf dan karyawan PPs MIK STMIK Nusa Mandiri yang telah melayani penulis dengan baik selama kuliah.

6. Teman-teman seperjuangan yang memberi inspirasi dalam pembuatan tesis ini.

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

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

Jakarta, 27 Februari 2012

Saifudin
Penulis

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini, saya :

Nama : Saifudin
NIM : 14000418
Program Studi : Magsiter Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *e-Business*
Jenis Karya : Tesis

Demi pengembangan ilmu pengetahuan, dengan ini menyetujui untuk memberikan ijin kepada pihak Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri) **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalti-Free Right*)** atas karya ilmiah kami yang berjudul : “Kajian Penerapan Algoritma C4.5 Dalam Rekrutmen Nasabah Bank” beserta perangkat yang diperlukan (apabila ada).

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

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

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, April 2012

Yang menyatakan,

Materai Rp. 6.000,-

Saifudin

ABSTRAK

Nama : Saifudin
NIM : 14000418
Program Studi : Magsiter Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *e-Business*
Judul Tesis : “Kajian Penerapan Algoritma C4.5 Dalam Rekrutmen Nasabah Bank”

Dunia perbankan sekarang ini berusaha mencari nasabah deposito sebanyak-banyaknya dengan melakukan promosi pemasaran langsung untuk rekrutmen nasabah bank, dengan adanya promosi pemasaran langsung terdapat permasalahan yaitu banyak nasabah yang dilakukan promosi pemasaran langsung dengan cara dihubungi menggunakan kontak telepon tidak menjadi nasabah deposito. Tujuan dari penelitian ini adalah menguji keakurasian dari algoritma C4.5 berdasarkan metodologi CRISP-DM. Dengan menggunakan algoritma C4.5 terbukti akurasi dalam mengklasifikasikan nasabah deposito yang berkaitan dengan promosi pemasaran langsung rekrutmen nasabah bank.

Kata kunci: bank, data mining, algoritma C4.5, klasifikasi, rekrutmen, CRISP-DM

ABSTRACT

Name : Saifudin
NIM : 14000418
Study of Program : Magsiter Ilmu Komputer
Levels : Strata Dua (S2)
Concentration : *e-Business*
Title : “Kajian Penerapan Algoritma C4.5 Dalam Rekrutmen Nasabah Bank”

Nowdays, bank is trying to attract as many customers deposits by doing direct marketing to recruitment of bank customers, with the direct marketing there is a problem that many customers who do direct marketing contact by using the contact phone no interest in being customer deposits. The purpose of this study was to test the accuracy of the C4.5 algorithm based methodology CRISP-DM. By using C4.5 algorithm is proved the accuracy in classifying customer deposits which related with direct marketing to recruitment bank customers.

Keyword: bank, data mining, C4.5 algorithm, classifying, recruitment, CRISP-DM

DAFTAR ISI

	Halaman
HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
HALAMAN PERNYATAAN ORISINALITAS	iii
HALAMAN PENGESAHAN	iv
KATA PENGANTAR	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	vii
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Masalah Penelitian	2
1.3. Tujuan dan Manfaat	2
1.4. Sistematika Penulisan	3
BAB II LANDASAN DAN KERANGKA PEMIKIRAN	5
2.1. Tinjauan Pustaka	5
2.2. Tinjauan Studi	13
2.3. Tinjauan Obyek Penelitian	16
2.4. Kerangka Pemikiran	17
2.5. Hipotesis	19
BAB III METODE PENELITIAN	20
3.1. Jenis Penelitian	20
3.2. Kerangka Pendekatan	21
3.3. Metode Penelitian	22
3.4. Populasi, Sampel dan Metode Pemilihan Sampel	30
3.5. Metode Pengumpulan Data	31
3.6. Instrumen Penelitian	31
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	33
4.1. Hasil Penelitian	33
4.2. Pembahasan	33
4.3. Implikasi Penelitian	40
BAB V PENUTUP	43
5.1. Kesimpulan	43
5.2. Saran	43
DAFTAR REFERENSI	44

DAFTAR TABEL

	Halaman
Tabel 2.1. Contoh masukan untuk algoritma C4.5.....	11
.....	
Tabel 2.2. Tinjauan studi dari penelitian yang sudah ada	15
Tabel 3.1. Atribut dan Nilai Kategori	23
Tabel 3.2. Sebagian data murni nasabah yang dihubungi pada saat promosi langsung.....	25
Tabel 3.3. Data training diperoleh dari proses <i>data cleaning</i> , <i>data integration</i> dan <i>data reduction</i>	28
Tabel 4.1. Hasil perhitungan seluruh atribut	34
Tabel 4.2. Nilai <i>entropy</i> dan <i>gain</i> untuk menentukan simpul 1.1	35

DAFTAR GAMBAR

	Halaman
Gambar 2.1. Data Mining sebagai Proses CRISP-DM	1
Gambar 2.2. Tahap pertama proses klasifikasi yaitu proses <i>learning</i>	13
Gambar 2.3. Tahap kedua proses klasifikasi yaitu proses <i>training</i>	13
Gambar 2.4. Kerangka Pemikiran.....	18
Gambar 3.1. Kerangka Pendekatan.....	21
Gambar 4.1. <i>Tree</i> hasil perhitungan dengan algoritma C4.5	38
Gambar 4.2. Hasil evaluasi <i>confusion matrix</i> menggunakan Rapid Miner.....	39
Gambar 4.3. Nilai <i>Acurracy</i> , <i>Recall</i> dan <i>Precision</i> menggunakan Rapid Miner	40
Gambar 4.4. Kurva ROC.....	40
Gambar 4.5. <i>Graphical User Interface</i> (GUI) untuk promosi langsung terhadap nasabah bank	41

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perbankan semakin mendominasi perkembangan ekonomi dan bisnis suatu negara, tidak hanya di negara maju tapi juga di negara berkembang. Perkembangan teknologi sekarang ini cepat sekali, sekarang ini memasuki era komunikasi tanpa batas. Perkembangan dunia perbankan mengikuti dari perkembangan teknologi yang modern baik dari segi macam produk yang ditawarkan (misalnya kredit, tabungan, deposito), kualitas pelayanan yang dimiliki.

Bank merupakan badan penyedia dana pembiayaan pembangunan, antara lain melalui kegiatan penyaluran kredit dan investasi. Kegiatan perkreditan dan berbagai jasa yang diberikan oleh bank ini membantu debitur mengatasi kekurangan modal dalam mengelola, membiayai operasi, dan mengembangkan usaha sehingga mampu meningkatkan efisiensi, produktivitas, dan daya saing. Bank menjadi bisnis yang unik karena pihak bank melakukan bisnisnya dengan menggunakan dana orang lain. Bank berfungsi menjadi “perantara keuangan” dengan usaha utama menghimpun dana dari masyarakat dan menyalurkannya kembali kepada masyarakat serta memberikan jasa-jasa lainnya dalam lalu lintas pembayaran.

Salah satu produk yang ditawarkan oleh perbankan adalah deposito. Deposito atau yang sering juga disebut sebagai deposito berjangka, merupakan produk bank sejenis jasa tabungan. Dana dalam deposito dijamin oleh pemerintah melalui Lembaga Penjamin Simpanan (LPS) dengan persyaratan tertentu. Deposito biasanya memiliki jangka waktu tertentu di mana uang didalamnya tidak boleh ditarik nasabah. Deposito baru bisa dicairkan sesuai dengan tanggal jatuh temponya, biasanya deposito mempunyai jatuh tempo 1, 3, 6, atau 12 bulan. Bila deposito dicairkan sebelum tanggal jatuh tempo, maka akan kena penalti. Untuk menarik minat nasabah menjadi nasabah deposito kemudian bank melakukan kegiatan promosi pemasaran langsung terhadap nasabah, tetapi dengan banyak nasabah yang

dilakukan promosi pemasaran langsung atau dihubungi tidak banyak nasabah yang menjadi nasabah deposito.

1.2. Masalah Penelitian

1.2.1. Identifikasi Masalah

Permasalahan yang sering ada berkaitan dengan nasabah bank merupakan masalah penelitian dapat diidentifikasi sebagai berikut :

1. Aspek Teknis

Aspek ini menyangkut tingkat kesehatan bank yang memenuhi standar kesehatan Bank Indonesia yang memberikan fasilitas pelayanan yang memuaskan dan memberikan tingkat suku bunga deposito yang kompetitif.

2. Aspek Perbankan

Aspek ini menyangkut banyaknya lembaga perbankan yang menawarkan produk deposito yang sama ke masyarakat.

3. Pengelolaan deposito

Aspek ini menyangkut meningkatnya pengelolaan dari lembaga perbankan kearah yang semakin profesional membuat masyarakat lebih percaya dengan lembaga perbankan penyedia produk deposito.

4. Perilaku nasabah

Aspek ini menyangkut ketidaktahuan masyarakat tentang produk deposito yang ditawarkan oleh perbankan.

1.2.2. Batasan Masalah

Mengingat luasnya masalah dan keterbatasan waktu, maka dalam tesis ini hanya akan membahas pelanggan nasabah deposito sebagai salah satu solusi dari beberapa masalah yang ada. Untuk mengetahui keberhasilan sistem tersebut, melakukan Kajian Penerapan Algoritma C4.5 Dalam Rekrutmen Nasabah Bank.

1.2.3. Rumusan Masalah

Berdasarkan latar belakang masalah yang ada adalah berapa tingkat akurasi dari diterapkannya algoritma C4.5 dalam rekrutmen nasabah bank?

1.3. Tujuan dan Manfaat

1.3.1. Tujuan Penelitian

Tujuan dari penelitian ini untuk menggali atau mencari data dan informasi yang berhubungan dengan nasabah bank. Sesuai dengan permasalahan yang telah diuraikan, maka tujuan dari penelitian ini adalah: Mendeteksi tingkat akurasi dari algoritma C4.5 yang digunakan.

1.3.2. Manfaat Penelitian

Hasil dari penelitian yang disajikan ini diharapkan dapat memberikan manfaat bagi:

1. Manfaat teoritis dari hasil penelitian ini diharapkan dapat mengembangkan ilmu pengetahuan, terutama bagi para akademisi yang menganalisis tentang pengkajian tentang nasabah bank. Kemudian temuan penelitian ini dapat dijadikan sebagai referensi pada penelitian selanjutnya.
2. Manfaat Praktis dari hasil penelitian ini diharapkan dapat digunakan oleh tim *marketing* dalam penentuan strategi *marketing* yang tepat sasaran.
3. Manfaat Kebijakan dari hasil penelitian ini sebagai bahan masukan dan evaluasi bagi lembaga-lembaga perbankan yang ada untuk digunakan sebagai penunjang alat bantu untuk meningkatkan akurasi dalam strategi *marketing*.

1.4. Sistematika Penulisan

Bab I Pendahuluan

Bab ini berisi uraian mengenai; latar belakang, masalah penelitian, tujuan dan manfaat penelitian, serta sistematika penulisan

Bab II Landasan dan Kerangka Pemikiran

Bab ini membahas tentang; tinjauan pustaka, tinjauan studi, tinjauan obyek penelitian, kerangka pendekatan dan hipotesis

Bab III Metodologi Penelitian

Bab ini membahas tentang jenis penelitian, kerangka pendekatan, model dan variabel, populasi dan sample, metode pengumpulan data, metode analisa atau analisa data.

Bab IV Hasil dan Pembahasan

Bab ini membahas tentang hasil, pembahasan serta implikasi penelitiannya.

Bab V Penutup

Bab ini membahas tentang kesimpulan dari penelitian serta saran yang diberikan untuk penelitian ini.

BAB II

LANDASAN DAN KERANGKA PEMIKIRAN

2.1. Tinjauan Pustaka

2.1.1. Bank

Bank merupakan sebuah lembaga intermediasi keuangan umumnya didirikan dengan kewenangan untuk menerima simpanan uang, meminjamkan uang, dan menerbitkan promes atau yang dikenal sebagai *banknote* (Hoggson, 1926). Kata bank berasal dari bahasa Italia yaitu *banca* berarti tempat penukaran uang. Bank adalah badan usaha yang menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkannya kepada masyarakat dalam bentuk kredit dan atau bentuk-bentuk lainnya dalam rangka meningkatkan taraf hidup rakyat banyak (UU no 10 tahun 1998).

Bank juga dikenal sebagai tempat untuk menukar uang, memindahkan uang atau menerima segala macam bentuk pembayaran dan setoran seperti pembayaran listrik, telepon, air, pajak, uang kuliah, dan pembayaran lainnya. (Kasmir, 2000) dalam (Triwahyuniati, 2008).

Industri perbankan telah mengalami perubahan besar dalam beberapa tahun terakhir. Industri ini menjadi lebih kompetitif karena deregulasi peraturan. Saat ini, bank memiliki fleksibilitas pada layanan yang mereka tawarkan, lokasi tempat mereka beroperasi, dan tarif yang mereka bayar untuk simpanan deposito.

2.1.2. Deposito

Deposito merupakan produk simpanan di bank yang penyetoran maupun penarikannya hanya dapat dilakukan pada waktu tertentu saja atau sesuai dengan jatuh temponya sehingga deposito dikenal juga sebagai tabungan berjangka. Sebagaimana layaknya tabungan yang sudah memasyarakat, deposito juga banyak dipilih orang sebagai alternatif lain dalam menyimpan uangnya.

Fungsi deposito adalah sebagai simpanan, mengapa tidak di tabung saja semuanya di tabungan? Walaupun deposito adalah tabungan juga tetapi mempunyai

karateristik berbeda dari tabungan biasa, yang menyebabkan deposito mempunyai kelebihan dan kekurangan, antara lain:

1. Setoran minimal.

Tidak seperti tabungan yang dapat dibuka dengan setoran awal yang kecil. Minimal penempatan deposito lebih besar, sehingga Anda harus punya uang lebih banyak untuk membuka deposito. Besarnya minimal pembukaan deposito pada tiap bank bervariasi, tetapi rata-rata saat ini yang paling minimal Rp 1.000.000.

2. Jangka waktu.

Penempatan deposito mengharuskan adanya pengendapan dana selama jangka waktu tertentu yang dapat dipilih oleh nasabahnya yaitu 1, 3, 6 atau 12 bulan. Karena itu jika ingin menambah jumlah saldo deposito hal itu tidak dapat dilakukan setiap saat. Begitu juga jika membutuhkan uang kemudian ingin mencairkan dana pada deposito. Karena adanya jangka waktu tadi maka deposito juga tidak bisa dicairkan setiap saat, tetapi pada saat jatuh tempo saja. Jika Anda terpaksa harus mencairkan deposito, biasanya bank akan mengenakan denda penalti pada tiap penarikan dana deposito yang belum jatuh tempo. Besarnya denda penalti juga bervariasi di berbagai bank. Ada yang berupa prosentase dari nilai deposito pada saat dicairkan (pokok + bunga), atau berupa prosentase dari nilai pokok depositonya saja.

3. Bunga deposito.

Bunga deposito selalu lebih besar dari bunga tabungan sehingga otomatis dana akan berkembang lebih cepat. Inilah biasanya yang menjadi daya tarik utama deposito, sehingga deposito lebih cocok dijadikan sarana investasi dibandingkan tabungan

4. Risiko rendah.

Walaupun tingkat suku bunga deposito lebih tinggi dari tabungan maupun giro, namun karena masih sama-sama produk simpanan di bank maka deposito bisa dogolongkan produk simpanan berisiko rendah.

5. Biaya administrasi dan pajak.

Keuntungan lainnya dari deposito adalah tidak dikenakannya biaya administrasi bulanan. Tidak seperti tabungan dan giro yang dikenakan biaya

administrasi bulanan. Walaupun demikian pemotongan tetap ada yaitu sebesar pajak deposito yang diperhitungkan dari hasil bunga deposito saja tidak termasuk pokok.

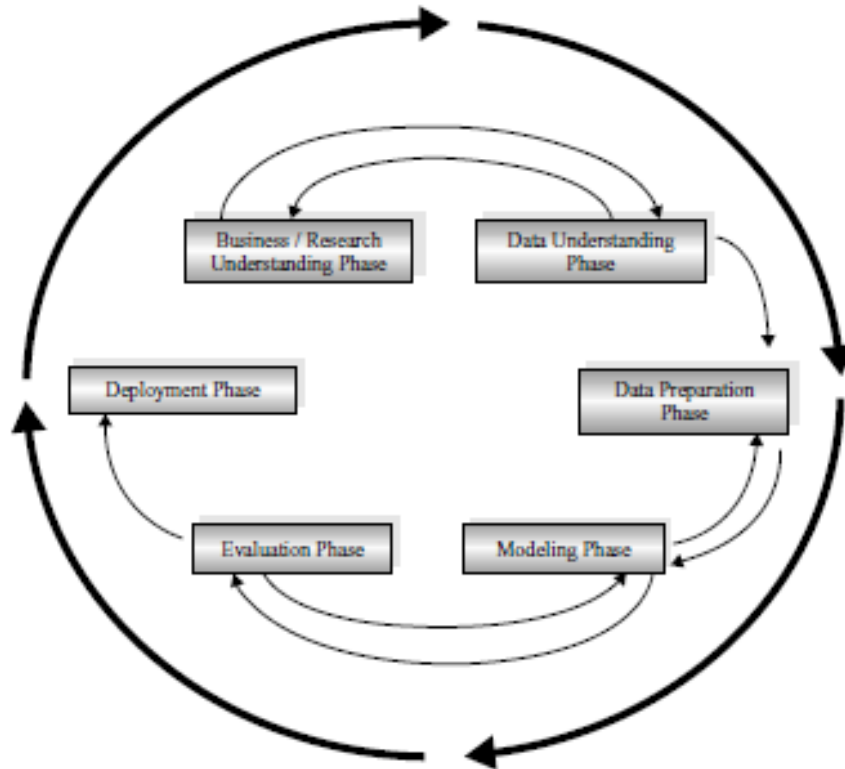
Kalau memilih bank yang bagus untuk penempatan deposito, maka hal utama yang harus dipertimbangkan tentu saja mengenai reputasi banknya. Pilihlah bank yang bonafide, artinya bank tersebut selain sehat, minimal memenuhi standar kesehatan dari Bank Indonesia juga memberikan fasilitas pelayanan yang memuaskan dan bisa memberikan tingkat suku bunga deposito yang kompetitif. Sebaiknya jangan mudah tergiur dengan banyak tawaran undian berhadiah, namun selalu dahulukan faktor kesehatan banknya terlebih dahulu. Berhati-hatilah jika ada bank yang menawarkan suku bunga deposito di atas tingkat suku bunga penjaminan bank Indonesia. Peraturan Bank Indonesia hanya membolehkan sebuah bank memberikan suku bunga deposito maksimal atau dibawah suku bunga penjaminan bank Indonesia. Jika suatu bank berani melanggar ketentuan tersebut bisa jadi hal ini mengindikasikan banknya sedang tidak sehat dan sedang sangat membutuhkan dana.

2.1.3. Data Mining

Data mining merupakan analisis (data yang besar) observasional data set untuk menemukan hubungan tak terduga dan untuk meringkas data dengan cara baru yang baik dimengerti dan berguna bagi pemilik data (Hand, Manilla dan Smith, 2001) dalam (Larose, 2007). *Data mining* merupakan proses ekstraksi pengetahuan dari data yang besar. Sesuai dengan fungsinya *data mining* merupakan proses pengambilan pengetahuan dari jumlah data yang besar yang disimpan dalam basis data, data *warehouse* atau informasi yang disimpan dalam repositori (Han dan Kamber, 2007).

Data Mining adalah sebuah proses, yang mana dalam melakukan prosesnya harus sesuai dengan prosedur dari proses tersebut, yaitu *Cross-Industry Standard Process for Data Mining* (CRISP-DM), yang terdiri dari keseluruhan proses, preprosesing data, pembentukan model, model evaluasi, dan tahap akhir penyebaran model

(Larose, 2005). Gambar 2.1 merupakan proses *Data Mining* CRISP-DM sebagai berikut :



Sumber: *Discovering Knowledge In Data*, (Larose, 2005)

Gambar 2.1. Data Mining sebagai Proses CRISP-DM

Cross-Industry Standard Process for Data Mining (CRISP-DM) terdiri dari enam fase (Larose, 2005), yaitu:

1. Fase Pemahaman Bisnis (*Business/Research Understanding Phase*)
 - a. Penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan.
 - b. Menerjemahkan tujuan dan batasan menjadi formula dari permasalahan *data mining*.
 - c. Menyiapkan strategi awal untuk mencapai tujuan.
2. Fase Pemahaman Data (*Data Understanding Phase*)
 - a. Mengumpulkan data

- b. Menggunakan analisis penyelidikan data untuk mengenali lebih lanjut data dan pencarian pengetahuan awal.
 - c. Mengevaluasi kualitas data.
 - d. Jika diinginkan, pilih sebagian kecil grup data yang mungkin mengandung pola dari permasalahan.
3. Fase Persiapan Data (*Data Preparation Phase*)
- a. Siapkan dari data awal, kumpulan data yang akan digunakan untuk keseluruhan fase berikutnya.
 - b. Pilih kasus dan variabel yang ingin dianalisis dan yang sesuai analisis yang akan dilakukan.
 - c. Lakukan perubahan pada beberapa variabel jika dibutuhkan.
 - d. Siapkan data awal sehingga siap untuk perangkat pemodelan.
4. Fase Pemodelan (*Modeling Phase*)
- a. Pilih dan aplikasikan teknik pemodelan yang sesuai.
 - b. Kalibrasi atur model untuk mengoptimalkan hasil.
 - c. Perlu diperhatikan bahwa beberapa teknik mungkin untuk digunakan pada permasalahan *data mining* yang sama.
 - d. Jika diperlukan, proses dapat kembali ke fase pengolahan data untuk menjadikan data ke dalam bentuk yang sesuai dengan spesifikasi kebutuhan *data mining* tertentu.
5. Fase Evaluasi (*Evaluation Phase*)
- a. Mengevaluasi satu atau lebih model yang digunakan dalam fase pemodelan untuk mendapatkan kualitas dan efektivitas sebelum disebarkan untuk digunakan.
 - b. Menetapkan apakah terdapat model yang memenuhi tujuan pada fase awal.
 - c. Menentukan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik.
 - d. Mengambil keputusan berkaitan dengan penggunaan hasil dari *data mining*.
6. Fase Penyebaran (*Deployment Phase*)
- a. Menggunakan model yang dihasilkan. Terbentuknya model tidak menandakan telah terselesaikannya proyek.

- b. Contoh sederhana penyebaran : Pembuatan laporan
- c. Contoh kompleks penyebaran : Penerapan proses *data mining* secara paralel pada departemen lain.

2.1.4. Algoritma C4.5

Algoritma C4.5 adalah rangkaian algoritma untuk menyelesaikan masalah klasifikasi yang sering digunakan dalam *machine learning* dan *data mining*. Hal ini ditargetkan pada suatu metode pembelajaran: Mengingat nilai atribut *dataset* yang dijelaskan oleh koleksi atribut dan termasuk salah satu set kelas yang saling eksklusif, Algoritma C4.5 mempelajari pemetaan dari nilai atribut ke kelas yang dapat diterapkan untuk mengklasifikasikan baru dari data yang belum diketahui. Algoritma C4.5 dan pohon keputusan merupakan dua model yang tak terpisahkan, karena untuk membangun sebuah pohon keputusan, dibutuhkan algoritma C4.5. Di akhir tahun 1970 hingga di awal tahun 1980-an, J. Ross Quinlan seorang peneliti di bidang *machine learning* mengembangkan sebuah model pohon keputusan yang dinamakan *Iterative Dichotomiser* (ID3), walaupun sebenarnya proyek ini telah dibuat sebelumnya oleh E.B. Hunt, J. Marin, dan P.T. Stone. Kemudian Quinlan membuat algoritma dari pengembangan ID3 yang dinamakan C4.5 yang berbasis *supervised learning*.

Sebagai contoh, lihat Gambar 2.3 di mana baris menunjukkan hari-hari tertentu, atribut menunjukkan kondisi cuaca pada hari tertentu, dan kelas menunjukkan apakah kondisi yang kondusif untuk bermain golf. Dengan demikian, setiap baris menunjukkan sebuah contoh, dijelaskan oleh nilai-nilai untuk atribut seperti *Outlook* (variabel acak), Suhu (bernilai kontinu), Kelembaban.

Tabel 2.1. Contoh masukan untuk algoritma C4.5

Day	Outlook	Temperature	Humidity	Windy	Play Golf?
1	Sunny	85	85	False	No
2	Sunny	80	90	True	No
3	Overcast	83	78	False	Yes
4	Rainy	70	96	False	Yes
5	Rainy	68	80	False	Yes
6	Rainy	65	70	True	No
7	Overcast	64	65	True	Yes
8	Sunny	72	95	False	No
9	Sunny	69	70	False	Yes
10	Rainy	75	80	False	Yes
11	Sunny	75	70	True	Yes
12	Overcast	72	90	True	Yes
13	Overcast	81	75	False	Yes
14	Rainy	71	80	True	No

Ada beberapa tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5 (Larose, 2005) yaitu :

1. Mempersiapkan *data training*. *Data training* biasanya diambil dari data histori yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menghitung akar dari pohon. Akar akan diambil dari atribut yang akan terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan rumus:

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

Keterangan :

S= Himpunan kasus

n = jumlah partisi S

P_i = proporsi S_i terhadap S

3. Kemudian hitung nilai *gain* menggunakan rumus :

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

Keterangan :

S = Himpunan Kasus

A = Fitur

n = jumlah partisi atribut A

|S_i| = Proporsi S_i terhadap S

|S| = jumlah kasus dalam S

4. Ulangi langkah kedua dan langkah ketiga hingga semua record terpartisi
5. Proses partisi pohon keputusan akan berhenti saat :
 - a. semua record dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut didalam record yang dipartisi lagi
 - c. Tidak ada record didalam cabang yang kosong

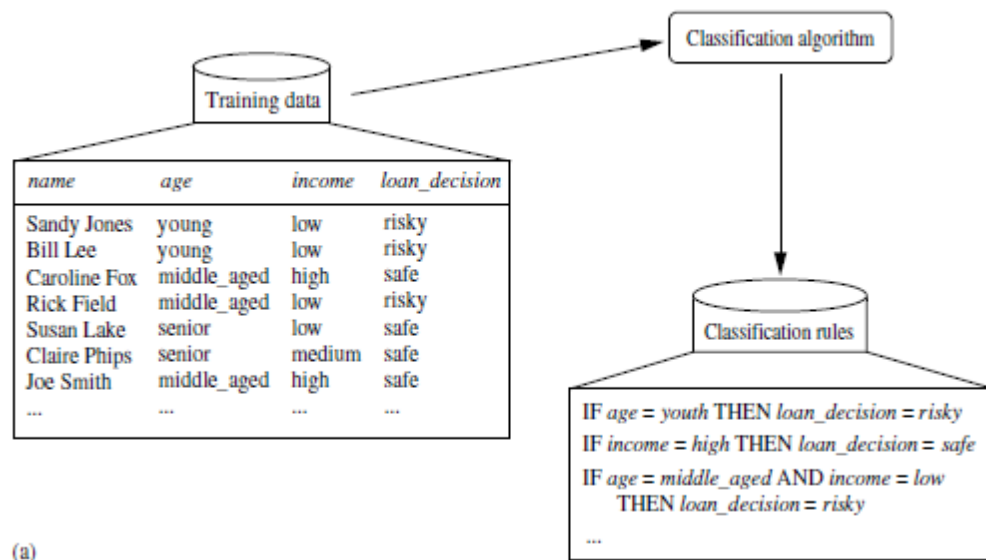
2.1.5. Klasifikasi

Klasifikasi merupakan model atau *classifier* dibangun untuk memprediksi kategori “aman” atau “berisiko” untuk permohonan pinjaman, “ya” atau “tidak” untuk pemasaran, sedangkan bidang Medis adalah “perlakuan A”, Perlakuan B” atau “Perlakuan C”. Kategori ini dapat direpresentasikan oleh nilai-nilai diskrit, dimana pemesanan di antara nilai-nilai tidak memiliki makna. Sebagai contoh, nilai 1, 2, dan 3 dapat digunakan untuk mewakili perlakuan A, B, dan C, di mana tidak ada pemesanan bahan baku tersirat antara kelompok rezim pengobatan.

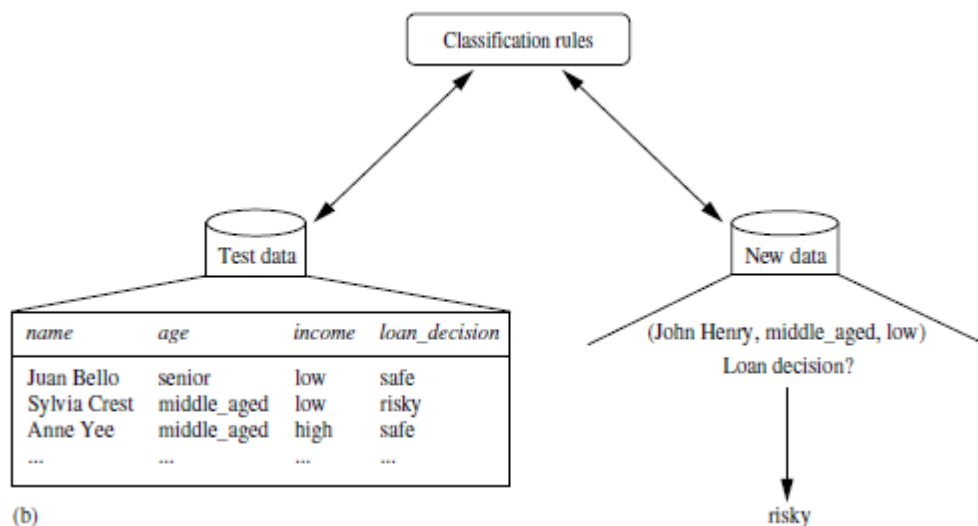
Tahap-tahap proses klasifikasi adalah sebagai berikut:

1. Proses *learning* yaitu data *training* dianalisa oleh algoritma klasifikasi dan menghasilkan sejumlah aturan, yaitu proses menganalisa data *training* oleh algoritma klasifikasi.
2. Proses *training* adalah menguji akurasi dari algoritma klasifikasi dengan memasukkan sejumlah data.

Untuk lebih jelasnya tahap-tahap proses klasifikasi dapat digambarkan sebagai berikut:



Gambar 2.2. Tahap pertama proses klasifikasi yaitu proses *learning*.



Gambar 2.3. Tahap kedua proses klasifikasi yaitu proses *training*.

2.2. Tinjauan Studi

1. *Using Data Mining For Bank Direct Marketing: An Application Of The CRISP-DM Methodology* (Moro S., 2012)

Semakin banyak promosi pemasaran dari waktu ke waktu telah mengurangi efeknya pada masyarakat umum. Tekanan ekonomi dan persaingan telah menyebabkan manajer pemasaran berinvestasi pada promosi yang diarahkan

dengan seleksi ketat dan teliti dari kontak. Promosi langsung dapat ditingkatkan melalui penggunaan teknik *Business Intelligence* (BI) dan *Data Mining* (DM).

Pelaksanaan proyek DM berdasarkan metodologi CRISP-DM. Data diambil dari promosi pemasaran di Portugal terkait dengan berlangganan deposito. Tujuan bisnis adalah untuk menemukan model yang dapat menjelaskan keberhasilan kontak, yaitu jika nasabah berlangganan deposito. Model tersebut dapat meningkatkan efisiensi promosi dengan mengidentifikasi karakteristik utama yang mempengaruhi keberhasilan, membantu dalam manajemen yang lebih baik dari sumber daya yang tersedia (misalnya usaha manusia, panggilan telepon, waktu) dan pemilihan kualitas yang tinggi dan menetapkan terjangkau pelanggan potensial membeli.

2. *Identifying Bank Frauds Using CRISP-DM And Decision Trees* (Rocha B. C., 2010)

Bertujuan untuk mengevaluasi penggunaan teknik *decision trees*, bersama dengan model pengelolaan CRISP-DM, untuk membantu dalam pencegahan penipuan bank. Dalam studi ini menawarkan tentang *decision trees*, konsep penting di bidang kecerdasan buatan. Studi ini difokuskan pada membahas bagaimana pohon-pohon ini dapat membantu dalam proses pengambilan keputusan untuk mengidentifikasi penipuan dengan analisis informasi mengenai transaksi bank. Informasi ini diambil dengan penggunaan teknik dan CRISP-DM model pengelolaan *data mining* dalam database operasional besar dari *log* transaksi bank internet.

3. *The use of CRISP-DM methodology for data mining in banking* (Konda P., 2009)

Data mining telah diakui sebagai medan independen dari penelitian selama lebih dari satu dekade. Diperkenalkan pada tahun 2000 CRISP-DM dianggap sebagai metodologi formal pertama yang sepenuhnya mencakup proses *data mining*. Perusahaan-perusahaan besar sekarang berusaha untuk memasukkan teknologi ini ke dalam sistem yang ada. Penggunaan data mining di bank NLB, d. d. Enterprizes di Slovenia membentuk sistem data *warehousing*. Menggunakan *online analytical processing* (OLAP) karyawan dapat

melakukan analisa bisnis dengan mudah, tetapi memiliki masalah menemukan pola yang kompleks dalam data. Oleh karena itu *data mining* merupakan *upgrade* dari sistem yang ada. Sejak penawaran *data mining* dengan data termasuk sejarah singkat pembangunan penyimpanan data. Bagian penelitian mencakup proses *data mining* dalam praktek. Tujuannya adalah untuk menghitung skor kecenderungan untuk setiap pelanggan. Hal ini dilakukan secara iteratif menggunakan SQL Server 2008 *platform database* dengan penekanan kuat pada *loading* data dan analisis. Saya membandingkan keakuratan model klasifikasi yang berbeda dengan menggunakan representasi grafis dan *cross validation*.

4. Penerapan Algoritma Klasifikasi C4.5 Untuk Penentuan Kelayakan Pemberian Kredit Koperasi (Firmansyah, 2011)

Analisa kredit yang dilakukan oleh analis terkadang tidak akurat, sehingga beberapa debitur yang diberikan kredit tidak mempunyai kemampuan untuk membayar yang menyebabkan kredit macet. Dari permasalahan ini diperlukan suatu model yang mampu mengklasifikasikan sekaligus memprediksi debitur mana saja yang bermasalah dan tidak bermasalah. Algoritma C4.5 ternyata mampu meningkatkan akurasi dalam menganalisa kelayakan kredit yang diajukan calon debitur. Semakin kaya informasi atau pengetahuan yang dikandung oleh *data training*, maka akurasi Algoritma C4.5 akan semakin meningkat.

Untuk lebih jelasnya tinjauan studi dari penelitian yang sudah ada bisa dilihat pada tabel 2.1 berikut ini:

Tabel 2.2. Tinjauan studi dari penelitian yang sudah ada

No	Judul	Latar Belakang	Tujuan	Metodologi	Hasil	Komentar
1.	<i>Using Data Mining For Bank Direct Marketing: An Application Of The CRISP-DM Methodology</i>	Banyak promosi langsung yang tidak berhasil mendapatkan nasabah	menemukan model yang dapat menjelaskan keberhasilan kontak, yaitu jika nasabah berlangganan deposito	CRISP-DM	Model tersebut dapat meningkatkan efisiensi promosi dengan mengidentifikasi karakteristik utama yang mempengaruhi	Dalam jurnal ini digunakan untuk menemukan model yang tepat untuk promosi langsung, tanpa ada GUI nya

					hi keberhasilan	
2.	<i>Identifying Bank Frauds Using CRISP-DM And Decision Trees</i>	Banyak penipuan dari pihak Bank	membahas bagaimana pohon-pohon ini dapat membantu dalam proses pengambilan keputusan untuk mengidentifikasi penipuan dengan analisis informasi mengenai transaksi bank	Algoritma <i>decision tree</i> dan CRISP-DM	dapat membantu dalam proses pengambilan keputusan untuk mengidentifikasi penipuan oleh pihak bank	Dalam jurnal ini membahas tentang penipuan yang dilakukan oleh pihak bank, sedangkan dalam tesis yang dibuat mencari promosi yang tepat untuk rekrutmen nasabah
3.	<i>The use of CRISP-DM methodology for data mining in banking</i>	menemukan pola yang kompleks dalam data	karyawan dapat melakukan analisa bisnis dengan mudah	CRISP-DM dan OLAP	skor kecenderungan untuk setiap pelanggan	Dalam tesis ini untuk mengukur kecenderungan pelanggan, yang membedakan pada tujuannya yaitu untuk menemukan <i>rule</i> yang tepat disertai GUInya
4.	Penerapan Algoritma Klasifikasi C4.5 Untuk Penentuan Kelayakan Pemberian Kredit Koperasi (Firmansyah, 2011)	Analisa kredit yang dilakukan oleh analis terkadang tidak akurat	mengklasifikasikan sekaligus memprediksi debitur mana saja yang bermasalah dan tidak bermasalah	Algoritma C4.5	Algoritma C4.5 ternyata mampu meningkatkan akurasi dalam menganalisa kelayakan kredit yang diajukan calon debitur	menggunakan algoritma C4.5 untuk mengklasifikasi nasabah pelanggan kredit koperasi

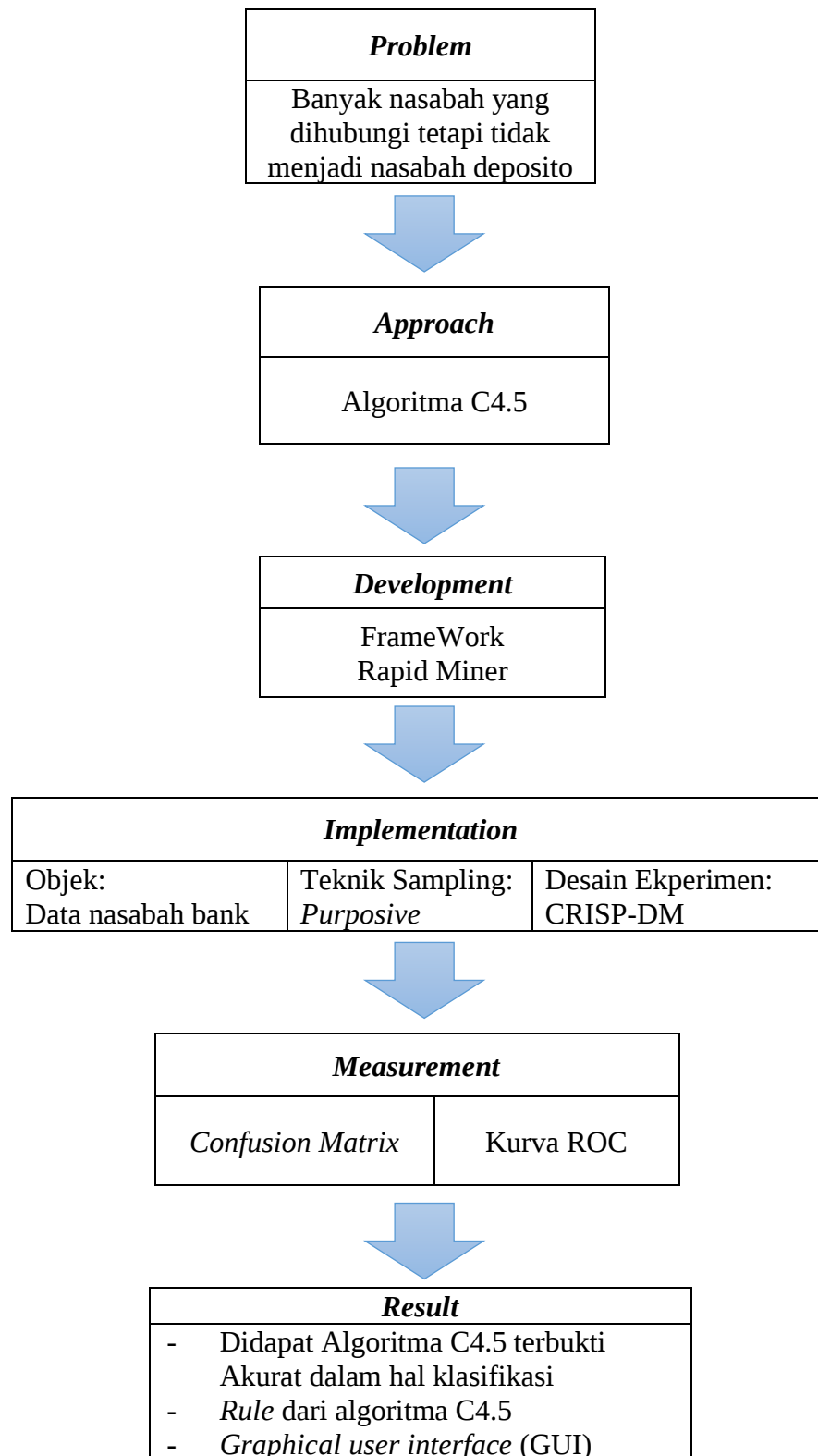
2.3. Tinjauan Obyek Penelitian

Data yang digunakan untuk Kajian Penerapan Algoritma C4.5 Dalam Rekrutmen Nasabah Bank diambil dari *UCI machine learning repository*. Data

tersebut terkait dengan promosi pemasaran langsung dari lembaga perbankan Portugal. Promosi pemasaran didasarkan pada panggilan telepon. Seringkali, lebih dari satu kontak untuk klien yang sama diperlukan, untuk mengakses jika produk (bank deposito berjangka) akan (atau tidak) berlangganan.

2.4. Kerangka Pemikiran

Kerangka pemikiran dari penelitian dimulai dari permasalahan banyak nasabah yang tidak menjadi nasabah deposito kemudian dibuat model dengan algoritma C4.5 untuk mengklasifikasikan permasalahan. Untuk mengembangkan aplikasi berdasarkan model yang dibuat, digunakan *Rapid Miner*. Tahap berikutnya yaitu implementasi, pada tahap ini objek implementasi dilakukan pengujian dari data yang didapatkan, teknik *sampling* menggunakan *purposive sampling*, dan desain ekperimennya digunakan proses *Cross Industry Standard Process for Data Mining* (CRISP-DM). Untuk mengukur tingkat keakurasian menggunakan *Confusion Matrix* dan kurva *Receiver Operating Characteristic* (ROC), setelah dilakukan pengukuran akan dihasilkan algoritma yang terbukti akurat dalam hal klasifikasi, rule dari algoritma yang terbukti akurat dalam hal klasifikasi dan *graphical user interface* (GUI). Kerangka pemikiran dalam penulisan tesis ini yaitu:



Gambar 2.4 Kerangka Pemikiran

2.5. Hipotesis

2.5.1. Hipotesa Umum

Secara umum hipotesis yang diajukan dalam penelitian ini, diduga algoritma C4.5 dapat diacu dalam promosi langsung untuk rekrutmen nasabah deposito.

2.5.2. Hipotesa Khusus

Secara khusus hipotesis berdasarkan pada kerangka pemikiran dapat disusun sebagai berikut:

1. Membuktikan algoritma C4.5 akurat digunakan dalam penelitian promosi langsung nasabah deposito.
2. Membantu lembaga untuk memecahkan masalah promosi langsung rekrutmen nasabah deposito lebih cepat dan akurat.

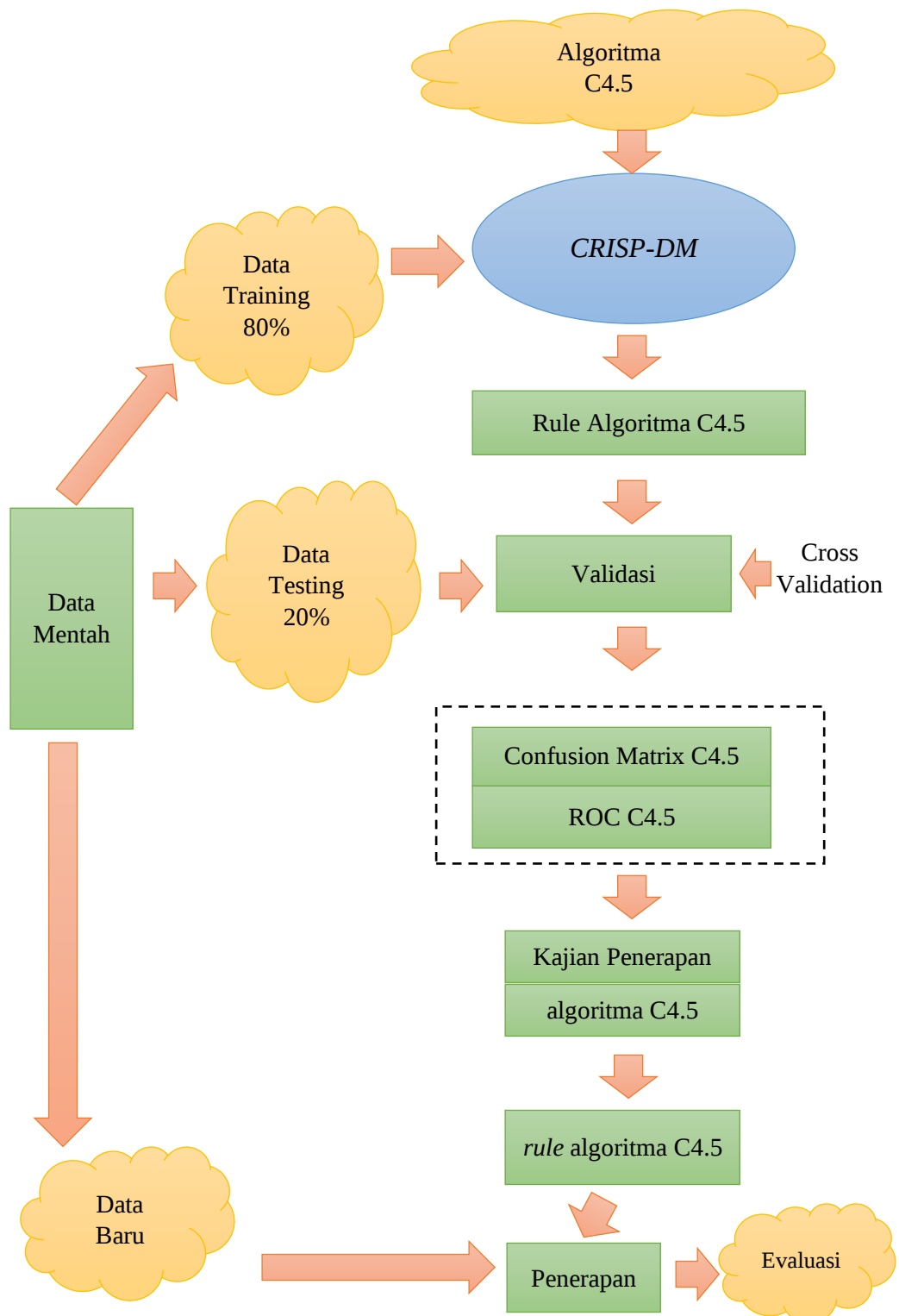
BAB III

METODE PENELITIAN

3.1. Jenis Penelitian

Jenis penelitian yang dilakukan merupakan penelitian eksperimen, metode eksperimen memiliki ciri khas tersendiri terutama dengan adanya kelompok kontrol. Dalam bidang ilmu pengetahuan, penelitian-penelitian dapat menggunakan desain eksperimen karena variabel-variabel dapat dipilih dan variabel-variabel lain yang dapat mempengaruhi proses eksperimen itu dapat dikontrol secara ketat. Sehingga dalam metode ini, peneliti memanipulasi paling sedikit satu variabel, mengontrol variabel lain yang relevan, dan mengobservasi pengaruhnya terhadap variabel terikat. Manipulasi variabel bebas inilah yang merupakan salah satu karakteristik yang membedakan penelitian eksperimental dari penelitian-penelitian lain.

3.2. Kerangka Pendekatan



Gambar 3.1. Kerangka Pendekatan

Gambar 3.1. merupakan gambaran kerangka pemikiran dalam tesis, prosesnya dimulai dari adanya data mentah yang dibagi menjadi dua yaitu untuk *data training* sebanyak 80% dan *data testing* sebanyak 20%, selanjutnya *data training* digunakan untuk proses *Cross Standard Industry Process for Data Mining* (CRISP-DM) dengan menggunakan algoritma C4.5, dari proses tersebut akan didapatkan *rule* dari algoritma C4.5. Kemudian data testing divalidasi dengan *rule* yang dihasilkan dalam proses *Cross Standard Industry Process for Data Mining* (CRISP-DM), sedangkan untuk memvalidasinya menggunakan *Cross Validation*, hasil dari validasi tersebut adalah *confusion matrix* algoritma C4.5 dan *receiver operating characteristic* (ROC) algoritma C4.5. Langkah selanjutnya dilakukan kajian tingkat akurasi algoritma C4.5 serta didapat *rule* dari algoritma C4.5. Untuk menguji *rule* dari algoritma C4.5 dilakukan penerapan dengan menggunakan data mentah baru, dan akhirnya dilakukan evaluasi dari hasil penerapan algoritma C4.5.

3.3. Metode Penelitian

Dalam metode penelitian ini, penulis menggunakan model proses *Cross-Standard Industry Process for Data Mining* (CRISP-DM) yang terdiri dari 6 tahap (Larose, 2005):

1. Fase Pemahaman Bisnis (*Business/Research Understanding Phase*)

Berdasarkan data yang ada ditemukan nasabah yang menjadi nasabah deposito dari 3200 nasabah yang dihubungi hanya sebanyak 364 orang, sehingga ini menjadi permasalahan di perbankan di Portugal dan merupakan imbas dari metode promosi pemasaran langsung untuk rekrutmen nasabah bank yang kurang akurat. Maka dibutuhkan model klasifikasi algoritma C4.5 untuk menentukan metode promosi pemasaran langsung untuk rekrutmen nasabah bank dengan tujuan analisa yang dilakukan lebih akurat.

2. Fase Pemahaman Data (*Data Understanding Phase*)

Berdasarkan data promosi pemasaran langsung untuk rekrutmen nasabah bank yang mulai diambil pada bulan mei 2008 sampai dengan bulan November 2010, dimana dari 3200 nasabah yang dihubungi, 2836 nasabah yang dihubungi tidak menjadi nasabah deposito. Atribut atau variabel yang ada sebanyak 17 atribut. Setelah dilakukan proses *data preparation* atribut atau

variabel yang digunakan terdiri dari 13 atribut yang didapat dalam klasifikasi rekrutmen nasabah bank yang menjadi nasabah deposito. Dengan 12 atribut *predictor* dan 1 (satu) atribut tujuan. Atribut yang menjadi parameter nasabah deposito dapat dilihat sebagai berikut :

Tabel 3.1. Atribut dan Nilai Kategori

No	Atribut	Nilai	Kategori
1	<i>age</i>		
		<20	
		21-60	
		>61	
2	<i>job</i>		
		<i>admin</i>	
		<i>blue-collar</i>	
		<i>entrepreneur</i>	
		<i>housemaid</i>	
		<i>management</i>	
		<i>retired</i>	
		<i>self-employed</i>	
		<i>services</i>	
		<i>student</i>	
		<i>technician</i>	
		<i>unemployed</i>	
		<i>unknown</i>	
3	<i>matrial</i>		
		<i>single</i>	
		<i>married</i>	
		<i>divorced</i>	
4	<i>education</i>		
		<i>primary</i>	
		<i>scondary</i>	
		<i>teritary</i>	
		<i>unknown</i>	
5	<i>balance</i>		
		minus	Nilainya dibawah nol
		kecil	0-500
		sedang	501-5000
		tinggi	>5000
6	<i>housing</i>		
		<i>yes</i>	
		<i>no</i>	

7	<i>loan</i>		
		<i>yes</i>	
		<i>no</i>	
8	<i>contact</i>		
		<i>cellular</i>	
		<i>telephone</i>	
		<i>unknown</i>	
9	<i>day</i>		
		awal	1-10
		tengah	11-20
		Akhir	21-31
10	<i>month</i>		
		Q1	Januari-April
		Q2	Mei-Agustus
		Q3	September-Desember
11	<i>duration</i>		
		singkat	<100
		sedang	101-500
		lama	>500
12	<i>poutcome</i>		
		<i>success</i>	
		<i>unknown</i>	
		<i>failure</i>	
13	<i>class</i>		
		<i>yes</i>	
		<i>no</i>	

3. Fase Pengolahan Data (*Data Preparation Phase*)

Tabel 3.2 berikut ini merupakan sebagian data murni nasabah sebagai data yang digunakan dalam klasifikasi rekrutmen nasabah deposito.

Tabel 3.2. Sampel data murni nasabah yang dihubungi pada saat promosi langsung

no	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	class
1	30	unemployed	married	primary	no	1787	no	no	cellular	19	oct	79	1	-1	0	unknown	no
2	33	services	married	secondary	no	4789	yes	yes	cellular	11	may	220	1	339	4	failure	no
3	35	management	single	tertiary	no	1350	yes	no	cellular	16	apr	185	1	330	1	failure	no
4	30	management	married	tertiary	no	1476	yes	yes	unknown	3	jun	199	4	-1	0	unknown	no
5	59	blue-collar	married	secondary	no	0	yes	no	unknown	5	may	226	1	-1	0	unknown	no
6	35	management	single	tertiary	no	747	no	no	cellular	23	feb	141	2	176	3	failure	no
7	36	self-employed	married	tertiary	no	307	yes	no	cellular	14	may	341	1	330	2	other	no
8	39	technician	married	secondary	no	147	yes	no	cellular	6	may	151	2	-1	0	unknown	no
9	41	entrepreneur	married	tertiary	no	221	yes	no	unknown	14	may	57	2	-1	0	unknown	no
10	43	services	married	primary	no	-88	yes	yes	cellular	17	apr	313	1	147	2	failure	no
11	39	services	married	secondary	no	9374	yes	no	unknown	20	may	273	1	-1	0	unknown	no
12	43	admin.	married	secondary	no	264	yes	no	cellular	17	apr	113	2	-1	0	unknown	no
13	36	technician	married	tertiary	no	1109	no	no	cellular	13	aug	328	2	-1	0	unknown	no
14	20	student	single	secondary	no	502	no	no	cellular	30	apr	261	1	-1	0	unknown	yes
15	31	blue-collar	married	secondary	no	360	yes	yes	cellular	29	jan	89	1	241	1	failure	no
16	40	management	married	tertiary	no	194	no	yes	cellular	29	aug	189	2	-1	0	unknown	no
17	56	technician	married	secondary	no	4073	no	no	cellular	27	aug	239	5	-1	0	unknown	no
18	37	admin.	single	tertiary	no	2317	yes	no	cellular	20	apr	114	1	152	2	failure	no
19	25	blue-collar	single	primary	no	-221	yes	no	unknown	23	may	250	1	-1	0	unknown	no
20	31	services	married	secondary	no	132	no	no	cellular	7	jul	148	1	152	1	other	no
21	38	management	divorced	unknown	no	0	yes	no	cellular	18	nov	96	2	-1	0	unknown	no
22	42	management	divorced	tertiary	no	16	no	no	cellular	19	nov	140	3	-1	0	unknown	no
23	44	services	single	secondary	no	106	no	no	unknown	12	jun	109	2	-1	0	unknown	no

24	44	entrepreneur	married	secondary	no	93	no	no	cellular	7	jul	125	2	-1	0	unknown	no
25	26	housemaid	married	tertiary	no	543	no	no	cellular	30	jan	169	3	-1	0	unknown	no
26	41	management	married	tertiary	no	5883	no	no	cellular	20	nov	182	2	-1	0	unknown	no
27	55	blue-collar	married	primary	no	627	yes	no	unknown	5	may	247	1	-1	0	unknown	no
28	67	retired	married	unknown	no	696	no	no	telephone	17	aug	119	1	105	2	failure	no
29	56	self-employed	married	secondary	no	784	no	yes	cellular	30	jul	149	2	-1	0	unknown	no
30	53	admin.	married	secondary	no	105	no	yes	cellular	21	aug	74	2	-1	0	unknown	no
31	68	retired	divorced	secondary	no	4189	no	no	telephone	14	jul	897	2	-1	0	unknown	yes
32	31	technician	married	secondary	no	171	no	no	cellular	27	aug	81	3	-1	0	unknown	no
33	59	management	married	secondary	no	42	no	no	cellular	21	nov	40	1	-1	0	unknown	no
34	32	management	single	tertiary	no	2536	yes	no	cellular	26	aug	958	6	-1	0	unknown	yes
35	49	technician	married	tertiary	no	1235	no	no	cellular	13	aug	354	3	-1	0	unknown	yes
36	42	admin.	divorced	secondary	no	1811	yes	no	unknown	14	may	150	1	-1	0	unknown	no
37	78	retired	divorced	primary	no	229	no	no	telephone	22	oct	97	1	-1	0	unknown	yes
38	32	blue-collar	married	secondary	no	2089	yes	no	cellular	14	nov	132	1	-1	0	unknown	yes
39	33	management	married	secondary	no	3935	yes	no	cellular	6	may	765	1	342	2	failure	yes
40	23	services	single	tertiary	no	363	yes	no	unknown	30	may	16	18	-1	0	unknown	no
41	38	management	single	tertiary	no	11971	yes	no	unknown	17	nov	609	2	101	3	failure	no
42	36	management	single	tertiary	no	553	no	no	cellular	11	aug	106	2	-1	0	unknown	no
43	52	blue-collar	married	secondary	no	1117	yes	no	cellular	13	may	365	1	-1	0	unknown	no
.																	
.																	
4521	44	entrepreneur	single	tertiary	no	1136	yes	yes	cellular	3	apr	345	2	249	7	other	no

Tabel 3.2 merupakan data murni yang belum diolah menjadi *data training*. Jika kita lihat secara seksama pada tabel 3.2, maka akan didapati beberapa atribut yang tidak digunakan dan *record* yang duplikasi. Untuk itu maka diperlukan tehnik dalam *preprocessing* (Han & Kamber, 2006) yaitu :

- a. *Data Cleaning*, untuk membersihkan nilai yang kosong, tidak konsisten atau mungkin tupel yang kosong (*missing values* dan *noisy*).
- b. *Data Integration*, untuk meningkatkan akurasi dan efisiensi algoritma, data yang digunakan dikategorikalkan.
- c. *Data Reduction*. Jumlah atribut dan tupel yang digunakan untuk *data training* mungkin terlalu besar, hanya beberapa atribut yang diperlukan sehingga atribut yang tidak diperlukan akan dihapus. Tupel dalam *data set* mungkin terjadi duplikasi atau terdapat tupel yang sama, sehingga untuk memperkecil jumlah tupel, tupel yang sama akan dijadikan dalam satu tupel untuk mewakili tupel tersebut sehingga akan terlihat pada tabel 3.3 di bawah :

Tabel 3.3. Sampel *Data Training* diperoleh dari proses *data cleaning*, *data integration* dan *data reduction*.

no	age	job	marital	education	balance	housing	loan	contact	day	month	duration	poutcome	class
1	21-60	services	married	secondary	sedang	yes	yes	cellular	tengah	Q2	sedang	failure	no
2	21-60	management	single	tertiary	sedang	yes	no	cellular	tengah	Q1	sedang	failure	no
3	21-60	management	married	tertiary	sedang	yes	yes	unknown	awal	Q2	sedang	unknown	no
4	21-60	blue-collar	married	secondary	kecil	yes	no	unknown	awal	Q2	sedang	unknown	no
5	21-60	self-employed	married	tertiary	kecil	yes	no	cellular	tengah	Q2	sedang	other	no
6	21-60	technician	married	secondary	kecil	yes	no	cellular	awal	Q2	sedang	unknown	no
7	21-60	entrepreneur	married	tertiary	kecil	yes	no	unknown	tengah	Q2	singkat	unknown	no
8	21-60	services	married	primary	minus	yes	yes	cellular	tengah	Q1	sedang	failure	no
9	21-60	admin	married	secondary	kecil	yes	no	cellular	tengah	Q1	sedang	unknown	no
10	21-60	technician	married	tertiary	sedang	no	no	cellular	tengah	Q2	sedang	unknown	no
11	<20	student	single	secondary	sedang	no	no	cellular	akhir	Q1	sedang	unknown	yes
12	21-60	blue-collar	married	secondary	kecil	yes	yes	cellular	akhir	Q1	singkat	failure	no
13	21-60	technician	married	secondary	sedang	no	no	cellular	akhir	Q2	sedang	unknown	no
14	21-60	admin	single	tertiary	sedang	yes	no	cellular	tengah	Q1	sedang	failure	no
15	21-60	blue-collar	single	primary	minus	yes	no	unknown	akhir	Q2	sedang	unknown	no
16	21-60	services	married	secondary	kecil	no	no	cellular	awal	Q2	sedang	other	no
17	21-60	management	divorced	tertiary	kecil	no	no	cellular	tengah	Q3	sedang	unknown	no
18	21-60	services	single	secondary	kecil	no	no	unknown	tengah	Q2	sedang	unknown	no
19	21-60	entrepreneur	married	secondary	kecil	no	no	cellular	awal	Q2	sedang	unknown	no
20	21-60	housemaid	married	tertiary	sedang	no	no	cellular	akhir	Q1	sedang	unknown	no
21	21-60	blue-collar	married	primary	sedang	yes	no	unknown	awal	Q2	sedang	unknown	no
22	>61	retired	married	unknown	sedang	no	no	telephone	tengah	Q2	sedang	failure	no
23	21-60	self-employed	married	secondary	sedang	no	yes	cellular	akhir	Q2	sedang	unknown	no

24	21-60	admin	married	secondary	kecil	no	yes	cellular	akhir	Q2	singkat	unknown	no
25	21-60	technician	married	secondary	kecil	no	no	cellular	akhir	Q2	singkat	unknown	no
26	21-60	management	married	secondary	kecil	no	no	cellular	akhir	Q3	singkat	unknown	no
27	21-60	management	single	tertiary	sedang	yes	no	cellular	akhir	Q2	lama	unknown	yes
28	21-60	technician	married	tertiary	sedang	no	no	cellular	tengah	Q2	sedang	unknown	yes
29	>61	retired	divorced	primary	kecil	no	no	telephone	akhir	Q3	singkat	unknown	yes
30	21-60	blue-collar	married	secondary	sedang	yes	no	cellular	tengah	Q3	sedang	unknown	yes
31	21-60	management	married	secondary	sedang	yes	no	cellular	awal	Q2	lama	failure	yes
32	21-60	services	single	tertiary	kecil	yes	no	unknown	akhir	Q2	singkat	unknown	no
33	21-60	management	single	tertiary	sedang	no	no	cellular	tengah	Q2	sedang	unknown	no
34	21-60	blue-collar	married	secondary	sedang	yes	no	cellular	tengah	Q2	sedang	unknown	no
35	21-60	technician	married	tertiary	kecil	yes	no	cellular	tengah	Q2	sedang	unknown	no
36	21-60	technician	single	tertiary	sedang	yes	no	cellular	akhir	Q3	singkat	unknown	no
37	21-60	blue-collar	married	primary	kecil	no	no	telephone	awal	Q1	singkat	other	no
38	21-60	blue-collar	married	primary	kecil	yes	no	unknown	akhir	Q2	sedang	unknown	no
39	21-60	entrepreneur	single	primary	minus	yes	yes	cellular	awal	Q1	sedang	unknown	no
40	>61	admin	married	unknown	sedang	yes	no	cellular	akhir	Q1	sedang	success	yes
41	21-60	technician	single	secondary	kecil	yes	no	cellular	tengah	Q2	lama	unknown	no
42	21-60	technician	single	secondary	kecil	no	no	cellular	akhir	Q2	singkat	unknown	no
43	21-60	technician	married	tertiary	sedang	yes	no	cellular	tengah	Q2	sedang	other	yes
44	21-60	blue-collar	married	secondary	sedang	yes	no	cellular	tengah	Q2	sedang	unknown	no
45	21-60	management	married	secondary	kecil	no	yes	telephone	awal	Q1	sedang	unknown	no
.													
.													
3200	21-60	services	married	secondary	kecil	no	no	cellular	akhir	Q2	sedang	unknown	no

4. Fase Pemodelan (*Modeling Phase*)

Tahap ini juga dapat disebut tahap *learning* karena pada tahap ini *data training* diklasifikasikan dengan model dan selanjutnya menghasilkan sejumlah *rule*. Model yang digunakan adalah algoritma C4.5. Seperti yang telah dijelaskan sebelumnya, ada beberapa tahap yang harus dilalui dalam membentuk pohon keputusan, algoritma C4.5 digunakan untuk membuat pohon keputusan. Untuk membuat pohon keputusan, dibutuhkan *data training* yang diambil dari tabel 3.3.

Menghitung jumlah kasus *class yes* dan jumlah *class no* dan *Entropy* dari semua kasus dan kasus yang dibagi berdasarkan atribut *age, job, marital, education, balance, housing, loan, contact, day, month, duration, poutcome*. Kemudian, dilakukan penyeleksian atribut dengan menghitung *gain* tertinggi. Dengan menggunakan persamaan *gain*, hitung jumlah kasus dari masing-masing atribut dan hitung *gain*.

5. Fase Evaluasi (*Evaluation Phase*)

Pada tahap ini adalah tahap dilakukan pengujian menggunakan data *training* berjumlah 3200 dan data *testing* berjumlah 800 data untuk melihat hasil akurasi pada proses klasifikasi menggunakan metode algoritma C4.5. Serta evaluasi dengan menggunakan metode *confusion matrix* dan kurva ROC.

6. Fase Penyebaran (*Deployment Phase*)

Setelah pembentukan model, dilakukan analisa, dan pengukuran, selanjutnya pada tahap ini diterapkan model yang akurat untuk kajian penerapan cara promosi pemasaran langsung untuk rekrutmen nasabah bank yang tepat.

3.4. Populasi, Sampel dan Metode Pemilihan Sampel

3.4.1. Populasi

Yang dimaksud populasi dalam penelitian kali ini diperoleh dari data yang terkait dengan promosi pemasaran langsung dari perbankan di Portugal. Mulai dari bulan Mei 2008 sampai dengan bulan November 2010 sejumlah 4521 data pelanggan nasabah.

3.4.2. Sampel

Sampel yang diambil dari populasi lembaga perbankan di Portugal diambil individu yang dapat dijadikan responden sejumlah 3200 nasabah.

3.4.3. Metode Pemilihan Sampel

Mengingat jumlah populasi besar, hingga tidak memungkinkan dilakukan pemilihan sampel dengan cara pengundian. Dalam keadaan populasi yang demikian, digunakanlah teknik *sampling* random sistematis (*Systematic Random Sampling*). Persyaratan yang harus dipenuhi agar teknik *sampling* ini dapat digunakan, yakni tersedianya kerangka *sampling* (ukuran populasinya diketahui dengan pasti), dan populasinya mempunyai pola beraturan yang memungkinkan untuk diberikan nomor urut serta bersifat homogen.

Cara penggunaan teknik *sampling* random sistematis atau pengundian hanya dilakukan satu kali, yakni ketika menentukan unsur pertama dari *sampling* yang akan diambil. Penentuan unsur *sampling* selanjutnya ditempuh dengan cara memanfaatkan interval sampel. Interval sampel merupakan angka yang menunjukkan jarak antara nomor-nomor urut yang terdapat dalam kerangka *sampling* yang akan dijadikan patokan dalam menentukan atau memilih unsur-unsur *sampling* kedua dan seterusnya hingga unsur ke- n . Interval sampel biasanya dilambangkan dengan huruf k .

Interval sampel atau juga disebut *sampling ratio* diperoleh dengan cara membagi ukuran populasi dengan ukuran sampel yang dikehendaki (N/n). Misalnya, dari populasi (N) berukuran 4000, akan diambil sampel (n) berukuran 800, maka interval *sampling*nya adalah $4000/800 = 5$ atau $k = 5$. Maka unsur *sampling* pertama adalah satuan elementer yang bernomor s , maka penentuan unsur-unsur sampel berikutnya adalah:

Unsur pertama = s

Unsur kedua = $s + k$

Unsur ketiga = $s + 2k$

Unsur keempat = $s + 3k$, dan seterusnya hingga unsur ke- n

3.5. Metode Pengumpulan Data

Dalam penelitian ini data diperoleh dengan cara mendokumentasikan data yang berada di *UCI Machine Learning Repository*, dalam penelitian ini data dibedakan menjadi dua yaitu:

1. Data sekunder merupakan data yang diambil dari data promosi langsung yang dilakukan oleh bank di Portugal.
2. Data primer didapatkan dari hasil komputasi dengan menggunakan algoritma C4.5

3.6. Instrumen Penelitian

Berikut ini adalah instrumen yang digunakan dalam proses penelitian:

1. Perangkat Lunak

Perangkat lunak yang digunakan dalam penelitian ini antara lain:

- a. Sistem operasi : Windows 7 Ultimate
- b. Perangkat pemrograman : Rapid Miner 5
- c. Pembuatan GUI : Microsoft Visual Basic 6.0
- d. Pengolah data : Microsoft Excel 2010

2. Perangkat Keras

Perangkat keras yang digunakan dalam penelitian ini adalah komputer dengan spesifikasi sebagai berikut:

- a. Prosesor : Intel Atom N280 1.66 GHz
- b. RAM : 2 Gb
- c. Harddisk : 250Gb
- d. VGA : Mobile Intel® 945 256Mb

BAB IV

HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

Tujuan dari penelitian ini menguji keakuratan promosi pemasaran langsung yang dilakukan kepada nasabah bank dengan menggunakan algoritma C4.5. Data yang dianalisa adalah data dari bank di Portugal yang diambil dari *UCI machine learning repository*.

4.2. Pembahasan

Data training pada tabel 3.3 digunakan untuk menentukan apakah seorang nasabah menjadi nasabah deposito “yes” atau “no”. Berikut akan dibahas prediksi apakah nasabah deposito “yes” atau “no”, dengan menggunakan metode klasifikasi. Langkah untuk membuat pohon keputusan, yaitu :

1. Tabel 3.3 adalah *data training* beserta kelasnya
2. Hitung nilai *entropy*. Dari *data training* diketahui jumlah kasus ada 3200, konsumen yang termasuk kelas yes 364 *record* dan no 2836 *record* sehingga didapat *entropy*:

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

$$= (-364/3200 \cdot \log_2 (364/3200)) + (-2836/3200 \cdot \log_2 (2836/3200))$$

$$= 0,1543$$

3. Hitung nilai *gain* untuk tiap atribut, lalu tentukan nilai *gain* tertinggi. Yang mempunyai nilai *gain* tertinggi itulah yang akan dijadikan akar dari pohon. Misalkan untuk atribut status perkawinan = menikah, didapat nilai *gain*:

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

$$= 0,1543 - (87/3200) * 0,5297 + (2633/3200) * 0,1218 + (337/3200) * 0,1901$$

$$= 0,0197$$

Perhitungan *entropy* dan *gain* untuk semua atribut dilakukan, untuk mendapatkan nilai gain tertinggi. Hasil perhitungan seluruh atribut terlihat pada Tabel 4.1.

Tabel 4.1. Hasil perhitungan seluruh atribut

Node			Jumlah Kasus (S)	Yes (Si)	No (Si)	Entropy	Gain
1	Total		3200	364	2836	0,1543	
	<i>age</i>						0,00167
		<20	4	2	2	0,5	
		21-60	3105	330	2775	0,1448	
		>61	91	32	59	0,4053	
	<i>job</i>						0,00115
		<i>admin</i>	352	44	308	0,1685	
		<i>blue-collar</i>	665	42	623	0,0881	
		<i>enterpreneur</i>	117	10	107	0,1178	
		<i>housemaid</i>	80	7	73	0,1205	
		<i>management</i>	706	101	605	0,1908	
		<i>retired</i>	160	32	128	0,2575	
		<i>self-employed</i>	133	13	120	0,1338	
		<i>services</i>	282	33	249	0,1585	
		<i>student</i>	55	12	43	0,2776	
		<i>technician</i>	536	57	479	0,1449	
		<i>unemployed</i>	86	8	78	0,1277	
		<i>unknown</i>	28	5	23	0,2331	
	<i>matrial</i>						0,00023
		<i>single</i>	843	119	724	0,1885	
		<i>married</i>	2006	204	1802	0,1389	
		<i>divorced</i>	351	41	310	0,1582	
	<i>education</i>						0,00053
		<i>primary</i>	486	43	443	0,1218	
		<i>scondary</i>	1622	164	1458	0,1382	
		<i>tertiary</i>	960	146	814	0,2018	
		<i>unknown</i>	132	11	121	0,1150	
	<i>balance</i>						0,00069
		<i>minus</i>	262	20	242	0,1058	
		<i>kecil</i>	1419	125	1294	0,1213	
		<i>sedang</i>	1298	188	1110	0,1930	
		<i>tinggi</i>	221	31	190	0,1874	
	<i>housing</i>						0,00088
		<i>yes</i>	1821	155	1666	0,1174	

		<i>no</i>	1379	209	1170	0,2011	
	<i>loan</i>						0,00038
		<i>yes</i>	496	31	465	0,0872	
		<i>no</i>	2704	333	2371	0,1662	
	<i>contact</i>						0,00166
		<i>cellular</i>	2063	294	1769	0,1901	
		<i>telephone</i>	213	31	182	0,1938	
		<i>unknown</i>	924	39	885	0,0595	
	<i>day</i>						0,21811
		<i>awal</i>	926	107	819	0,1566	
		<i>tengah</i>	1334	154	1180	0,1565	
		<i>akhir</i>	940	103	837	0,1490	
	<i>month</i>						0,00183
		Q1	501	98	403	0,2526	
		Q2	2314	196	2118	0,1168	
		Q3	385	70	315	0,2368	
	<i>duration</i>						0,28784
		<i>singkat</i>	782	12	770	0,0219	
		<i>sedang</i>	2006	196	1810	0,1338	
		<i>lama</i>	412	156	256	0,4265	
	<i>poutcome</i>						0,27227
		<i>Success</i>	87	53	34	0,5297	
		<i>Unknown</i>	2633	233	2400	0,1218	
		<i>failure</i>	337	48	289	0,1901	
		<i>other</i>	143	30	113	0,2684	

Dari hasil perhitungan *entropy* dan *gain* yang didapat pada Tabel 4.1, terlihat bahwa atribut *duration* mempunyai nilai gain tertinggi yaitu 0,28784. Oleh karena itu *duration* merupakan simpul akar pada pohon keputusan.

Untuk menentukan simpul berikutnya, yaitu simpul 1.1 dilakukan lagi perhitungan *entropy* dan *gain* berdasarkan atribut *duration*. Jumlah kasus yang dihitung adalah sejumlah kasus dengan nilai dari simpul akar yaitu *duration*.

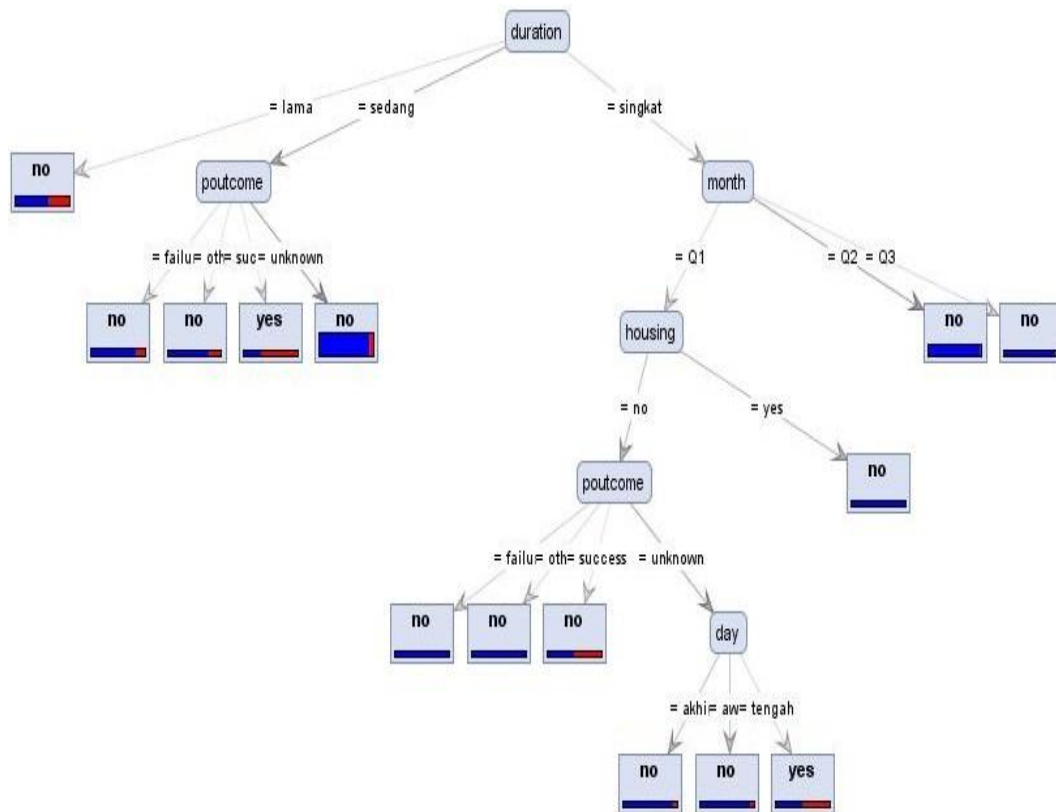
Tabel 4.2. Nilai *entropy* dan *gain* untuk menentukan simpul 1.1

Node			Jumlah Kasus (S)	Yes (Si)	No (Si)	Entropy	Gain
1.1	<i>duration</i>		412	156	256	0,4266	
	<i>lama</i>						
	<i>age</i>						0,00009

		<20	0	0	0	0	
		21-60	397	151	246	0,4279	
		>61	15	5	10	0,3900	
	<i>job</i>						0,00570
		<i>admin</i>	40	15	25	0,4238	
		<i>blue-collar</i>	82	25	57	0,3647	
		<i>entrepreneur</i>	18	6	12	0,3900	
		<i>housemaid</i>	11	3	8	0,3341	
		<i>management</i>	102	42	60	0,4503	
		<i>retired</i>	19	8	11	0,4565	
		<i>self-employed</i>	13	6	7	0,4809	
		<i>services</i>	41	18	23	0,4678	
		<i>student</i>	8	2	6	0,3113	
		<i>technician</i>	57	25	32	0,4676	
		<i>unemployed</i>	18	4	14	0,2820	
		<i>unknown</i>	3	2	1	0,5283	
	<i>matrial</i>						0,00352
		<i>single</i>	132	59	73	0,4726	
		<i>married</i>	230	76	154	0,3875	
		<i>divorced</i>	50	21	29	0,4558	
	<i>education</i>						0,00657
		<i>primary</i>	59	21	38	0,4088	
		<i>secondary</i>	210	74	136	0,4059	
		<i>tertiary</i>	129	60	69	0,4828	
		<i>unknown</i>	14	1	13	0,0993	
	<i>balance</i>						0,00100
		<i>minus</i>	38	12	26	0,3746	
		<i>kecil</i>	166	60	106	0,4132	
		<i>sedang</i>	189	77	112	0,4473	
		<i>tinggi</i>	19	7	12	0,4187	
	<i>housing</i>						0,00176
		<i>yes</i>	240	83	157	0,4005	
		<i>no</i>	172	73	99	0,4587	
	<i>loan</i>						0,00158
		<i>yes</i>	68	20	48	0,3547	
		<i>no</i>	344	136	208	0,4389	
	<i>contact</i>						0,00549
		<i>cellular</i>	276	117	159	0,4584	
		<i>telephone</i>	20	8	12	0,4422	
		<i>unknown</i>	116	31	85	0,3287	

	<i>day</i>						0,00079
		awal	107	40	67	0,4229	
		tengah	180	73	107	0,4461	
		akhir	125	43	82	0,3990	
	<i>month</i>						0,00039
		Q1	67	28	39	0,4544	
		Q2	289	108	181	0,4228	
		Q3	56	20	36	0,4098	
	<i>poutcome</i>						0,05468
		<i>Success</i>	13	8	5	0,5302	
		<i>Unknown</i>	335	123	212	0,4177	
		<i>failure</i>	41	15	26	0,4167	
		<i>other</i>	23	10	13	0,4652	

Gambar 4.1 adalah pohon keputusan akhir yang dihasilkan dari perhitungan *entropy* dan *gain* untuk seluruh atribut. Terlihat bahwa atribut *duration* menjadi simpul akar karena dari perhitungan pada Tabel 4.1, *duration* mempunyai nilai *gain* yang paling besar. Dari simpul akar bercabang menjadi tiga simpul sesuai dengan nilai yang dimilikinya, yaitu *duration* lama, sedang dan singkat. Kemudian pada gambar 4.1 juga terlihat simpul 1.1 adalah *poutcome*, karena atribut tersebut mempunyai nilai *gain* tertinggi, terlihat dari hasil perhitungan pada Tabel 4.2.



Gambar 4.1. Tree hasil perhitungan dengan algoritma C4.5

Pohon keputusan pada gambar 4.1 untuk menginterpretasikan pohon keputusan di atas menjadi aturan-aturan yaitu :

1. R1: IF *duration*=lama then *class*= no
2. R2: IF *duration*=sedang AND *poutcome*=failure AND *poutcome*=unknown AND *poutcome*=other AND then *class*= no
3. R3: IF *duration*=sedang AND *poutcome*=success THEN *class*= yes
4. R4: IF *duration*=singkat AND *month*=Q2 AND *month*=Q3 THEN *class*= no
5. R5: IF *duration*=singkat AND *month*=Q1 AND *housing*=yes THEN *class*= no
6. R6: IF *duration*=singkat AND *month*=Q1 AND *housing*=no AND *poutcome*=failure AND *poutcome*=success AND *poutcome*=other THEN *class*= no
7. R7: IF *duration*=singkat AND *month*=Q1 AND *housing*=no AND *poutcome*=unknown AND *day*=tengah THEN *class*= yes

8. R7: IF *duration*=singkat AND *month*=Q1 AND *housing*=no AND *poutcome*=unknown AND *day*=awal AND *day*=akhir THEN *class*= no

Dari hasil pengujian di atas akan dievaluasi tingkat akurasinya menggunakan 2 model yaitu menggunakan *confusion matrix* dan *Receiver Operating Characteristic* (ROC). Setelah diuji coba dengan metode *cross validation*, didapatkan hasil pengukuran terhadap data training yaitu *accuracy* = 89.62%, *precision* = 78.95% dan *recall* = 15.96%.

1. *Confusion Matrix*

Setelah dievaluasi dengan *confusion matrik* menggunakan Rapid Miner menghasilkan seperti pada gambar 4.2 berikut ini:

PerformanceVector

```

PerformanceVector:
accuracy: 89.62%
ConfusionMatrix:
True:  no    yes
no:    702   79
yes:    4    15
precision: 78.95% (positive class: yes)
ConfusionMatrix:
True:  no    yes
no:    702   79
yes:    4    15
recall: 15.96% (positive class: yes)
ConfusionMatrix:
True:  no    yes
no:    702   79
yes:    4    15
AUC (optimistic): 0.930 (positive class: yes)
AUC: 0.831 (positive class: yes)
AUC (pessimistic): 0.733 (positive class: yes)

```

Gambar 4.2. Hasil evaluasi *confusion matrix* menggunakan Rapid Miner

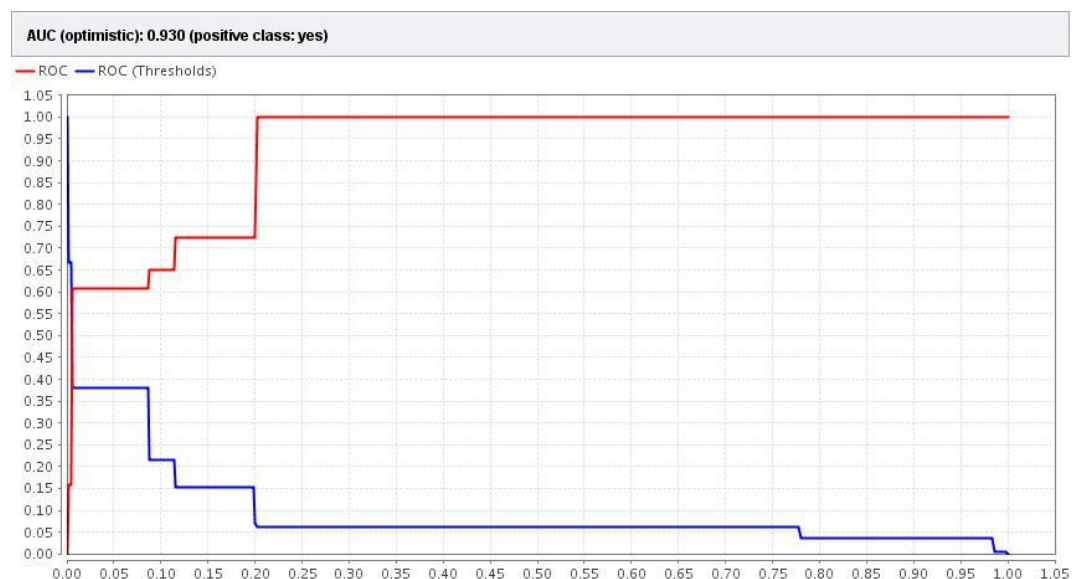
Terlihat pada gambar 4.3, nilai *accuracy*, *recall* dan *precision* yang dihasilkan oleh Rapid Miner menggunakan model *confusion matrix*.

accuracy: 89.62%			
	true no	true yes	class precision
pred. no	702	79	89.88%
pred. yes	4	15	78.95%
class recall	99.43%	15.96%	

Gambar 4.3. Nilai *Accuracy*, *Recall* dan *Precision* menggunakan Rapid Miner

2. Kurva *Receiver Operating Characteristic* (ROC)

Gambar 4.4 menunjukkan grafik ROC dengan nilai *Area Under Curve* (AUC), Akurasi AUC dikatakan sempurna apabila nilai AUC mencapai 1.000 dan akurasinya buruk jika nilai AUC dibawah 0.500.



Gambar 4.4. Kurva ROC

4.3. Implikasi Penelitian

Dari hasil evaluasi baik secara *confusion matrix* maupun kurva ROC ternyata terbukti akurat. Dengan hasil ini, maka kemungkinan terjadinya salah sasaran dalam promosi langsung ke nasabah bank dapat dikurangi dan menekan jumlah nasabah bank yang dihubungi untuk promosi tidak menjadi nasabah deposito, dengan demikian algoritma C4.5 dapat memberikan solusi untuk permasalahan penentuan strategi promosi langsung untuk nasabah deposito.

4.3.1. Aspek Sistem

Untuk memudahkan dalam pengambilan keputusan yang berkaitan dengan promosi langsung terhadap nasabah bank dibutuhkan piranti yang dapat membantu pengambilan keputusan yang cepat dan efektif. Pada gambar 4.5 merupakan sebuah *Graphical User Interface* (GUI) untuk memudahkan dalam hal pengambilan keputusan mengenai promosi langsung terhadap nasabah bank.

The screenshot shows a window titled "Aplikasi Perbankan" with the following fields and values:

Age	30	Loan	no
job	unemployed	Contact	telephon
Matrial	single	Day	11-20
Education	primary	Month	maret-agustus
Balance	sedang	Duration	200 second
Housing	yes	Phone outcome	success

Below the input fields, the output section displays:

Hasil Klasifikasi C4.5

Menjadi Nasabah

At the bottom, there are three buttons: **Process**, **Reset**, and **Exit**.

Gambar 4.5. *Graphical User Interface* (GUI) untuk promosi langsung terhadap nasabah bank

4.3.2. Aspek Manajerial

Dalam aspek manajerial dibutuhkan langkah-langkah sebagai berikut:

1. Adanya *Standard Operating Procedure* (SOP)

Dengan ada SOP setiap karyawan akan menjalankan tugas sesuai dengan ketentuan yang diberikan perusahaan dan akan menghasilkan jumlah nasabah yang sesuai dengan target yang diinginkan.

2. Sosialisasi

Setelah adanya SOP yang dikeluarkan oleh bank maka pihak manajemen bank wajib melakukan sosialisasi supaya seluruh karyawan dapat memahami isi dari SOP tersebut.

3. Pelatihan.

Untuk meningkatkan kemampuan karyawan memahami SOP perlu dilakukan pelatihan-pelatihan yang bertujuan melatih karyawan dapat menjalankan SOP sesuai dengan yang diinginkan dari pihak manajemen bank.

4. Penerapan.

Untuk melihat hasil dari pelatihan yang telah dilakukan maka perlu diterapkan dilapangan, berhadapan langsung dengan calon nasabah deposito yang akan diberi promosi tentang deposito.

4.3.3. Aspek Penelitian Lanjutan

Ada tiga hal yang harus dikembangkan:

1. Ekstrakulasi

Dari hasil penelitian ini diharapkan bisa diterapkan ditempat lain yang hampir sama permasalahannya yaitu masalah promosi langsung terhadap nasabah.

2. Ekstension

Dalam penelitian ini bisa dikembangkan lagi dengan menambah variabel-variabel baru sehingga dapat meningkatkan nilai akurasi.

3. Periodisasi

Penelitian ini bisa dikembangkan kembali setiap dua tahun sekali, mengingat perkembangan teknologi sekarang ini.

BAB V

PENUTUP

5.1. Kesimpulan

Dari hasil penelitian terbukti bahwa algoritma C4.5 akurat untuk menganalisa promosi pemasaran langsung yang dilakukan oleh bank. Hal ini dibuktikan dengan tingkat akurasi dari hasil evaluasi penelitian bahwa algoritma C4.5 mampu menganalisa nasabah yang menjadi nasabah deposito atau “yes” dan yang nasabah yang tidak menjadi nasabah atau “no” dengan tingkat keakuratan 89,62%.

5.2. Saran

Agar penelitian ini bisa ditingkatkan nilai akurasinya dengan penerapan algoritma C4.5, berikut adalah saran-saran yang diusulkan:

1. Hasil penelitian ini diharapkan dapat diterapkan dilembaga perbankan, khususnya pada bank riset lakukan, untuk meningkatkan akurasi dari promosi pemasaran langsung yang dilakukan terhadap calon nasabah deposito.
2. Untuk mendukung pengambilan keputusan dan pengembangan sistem informasi manajemen, model ini dapat menerapkan sistem yang menggunakan perangkat lunak.
3. Disertai dengan pembuatan *Standard Operational Procedure* (SOP), dilakukan sosialisasi bagi seluruh karyawan, pelatihan bagi *end-user*, dan berikutnya penerapan dilapangan supaya hasil yang didapatkan maksimal.
4. Penelitian ini diharapkan dapat dikembangkan lagi setiap dua tahun, dikarenakan perkembangan teknologi sekarang ini cepat sekali.

DAFTAR REFERENSI

<http://id.wikipedia.org/wiki/Deposito>

<http://id.wikipedia.org/wiki/Bank>

<http://archive.ics.uci.edu/ml/machine-learning-databases/00222/bank.zip>

Han dan Kamber (2007), *Data Mining Concepts and Techniques*, San Fransisco: Morgan Kauffman.

Hoggson, N. F. (1926) *Banking Through the Ages*, New York: Dodd, Mead & Company.

Konda, P. (2009) *The use of CRISP-DM methodology for data mining in banking*, Slovenia: Universitas Ljubljana.

Larose, D. T.(2005), *Discovering Knowledge In Data*, New Jersey: John Wiley & Sons, Inc.

Moro, S. and Laureano, RMS (2011) *Using Data Mining For Bank Direct Marketing: An Application Of The CRISP-DM Methodology*, Portugal: ESM

Rocha, B. C. and Junior, R. T. S (2010) *Identifying Bank Frauds Using CRISP-DM And Decision Trees*, Brasil: IJCSIT

Triwahyuniati, N (2008) *Pelaksanaan Analisis Pemberian Kredit Di PT Bank Haga Cabang Semarang*, Semarang: Universitas Diponegoro.

Lampiran

```

Private Sub Command1_Click()
If ((Val(Text1.Text) <= 500) And (Combo10.Text = "failure") And
(Combo10.Text = "unknown") And (Combo10.Text = "other")) Then
Label14.Caption = "No"
ElseIf ((Val(Text2.Text) <= 500) And (Combo10.Text = "success")) Then
Label14.Caption = "Yes"
ElseIf ((Val(Text2.Text) <= 100) And (Combo9.Text = "mei-agustus") And
(Combo9.Text = "september-desember")) Then
Label14.Caption = "No"
ElseIf ((Val(Text2.Text) <= 100) And (Combo9.Text = "januari-april") And
(Combo5.Text = "yes")) Then
Label14.Caption = "No"
ElseIf ((Val(Text2.Text) <= 100) And (Combo9.Text = "januari-april") And
(Combo5.Text = "no") And (Combo10.Text = "failure") And (Combo10.Text =
"success") And (Combo10.Text = "other")) Then
Label14.Caption = "No"
ElseIf ((Val(Text2.Text) <= 100) And (Combo9.Text = "januari-april") And
(Combo5.Text = "no") And (Combo10.Text = "unknown") And (Combo8.Text =
"11-20")) Then
Label14.Caption = "Yes"
ElseIf ((Val(Text2.Text) <= 100) And (Combo9.Text = "januari-april") And
(Combo5.Text = "no") And (Combo10.Text = "unknown") And (Combo8.Text =
"1-10") And (Combo8.Text = "21-31")) Then
Label14.Caption = "No"
Else
Label14.Caption = "No"
End If

```

```
End Sub
```

```

Private Sub Command2_Click()
Combo1.Text = ""
Combo2.Text = ""
Combo3.Text = ""
Combo4.Text = ""
Combo5.Text = ""
Combo6.Text = ""
Combo7.Text = ""
Combo8.Text = ""
Combo9.Text = ""
Combo10.Text = ""
Text1.Text = ""
Text2.Text = ""
Label14.Caption = ""
Text1.SetFocus

```

End Sub

```
Private Sub Command3_Click()
Unload Me
End Sub
```

```
Private Sub Form_Load()
'----- job -----
Combo1.AddItem "admin"
Combo1.AddItem "unknown"
Combo1.AddItem "unemployed"
Combo1.AddItem "management"
Combo1.AddItem "housemaid"
Combo1.AddItem "entrepreneur"
Combo1.AddItem "student"
Combo1.AddItem "blue-collar"
Combo1.AddItem "self-employed"
Combo1.AddItem "retired"
Combo1.AddItem "technician"
Combo1.AddItem "services"
'----- matrial -----
Combo2.AddItem "single"
Combo2.AddItem "married"
Combo2.AddItem "divorced"
'----- education -----
Combo3.AddItem "unknown"
Combo3.AddItem "secondary"
Combo3.AddItem "primary"
Combo3.AddItem "tertiary"
'----- education -----
Combo4.AddItem "minus"
Combo4.AddItem "kecil"
Combo4.AddItem "sedang"
Combo4.AddItem "tinggi"
'----- Housing -----
Combo5.AddItem "yes"
Combo5.AddItem "no"
'----- loan -----
Combo6.AddItem "yes"
Combo6.AddItem "no"
'----- contact -----
Combo7.AddItem "unknown"
Combo7.AddItem "telephone"
Combo7.AddItem "cellular"
'----- day -----
```

```
Combo8.AddItem "1-10"  
Combo8.AddItem "11-20"  
Combo8.AddItem "21-31"  
'----- day -----  
Combo9.AddItem "januari-april"  
Combo9.AddItem "maret-agustus"  
Combo9.AddItem "september-desember"  
'----- poutcome -----  
Combo10.AddItem "unknown"  
Combo10.AddItem "other"  
Combo10.AddItem "failure"  
Combo10.AddItem "success"  
  
End Sub
```