ANALISIS ALGORITMA *NAIVE BAYES CLASSIFIER* DENGAN SELEKSI ATRIBUT BERBASIS *PARTICLE SWARM OPTIMIZATION* UNTUK KLASIFIKASI PELANGGAN *CROSS SELLING* PRODUK INDIHOME PT. TELKOM JAKARTA

TESIS

Oleh

INDAH PURNAMASARI

NIM : 14001240

PROGRAM PASCA SARJANA MAGISTER ILMU KOMPUTER

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER

NUSAMANDIRI

JAKARTA

2016

ANALISIS ALGORITMA *NAIVE BAYES CLASSIFIER* DENGAN SELEKSI ATRIBUT BERBASIS *PARTICLE SWARM OPTIMIZATION* UNTUK KLASIFIKASI PELANGGAN *CROSS SELLING* PRODUK INDIHOME PT. TELKOM JAKARTA

TESIS

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

Oleh

INDAH PURNAMASARI

NIM : 14001240

PROGRAM PASCA SARJANA MAGISTER ILMU KOMPUTER

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER

NUSAMANDIRI

JAKARTA

2016

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini :

Nama : Indah Purnamasari
NIM : 14001240
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Management Information System* (MIS)

Dengan ini menyatakan bahwa tesis yang telah saya buat dengan judul : "Analisis Algoritma *Naive Bayes Classifier* dengan seleksi atribut berbasis *Particle Swarm Optimization* untuk Klasifikasi Pelanggan *Cross selling* Produk Indihome PT. Telkom Jakarta" 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, 20 Febuari 2016
Yang menyatakan,

Indah Purnamasari

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

Tesis ini diajukan oleh :

Nama	: Indah Purnamasari
NIM	: 14001240
Program Studi	: Magsiter Ilmu Komputer
Jenjang	: Strata Dua (S2)
Konsentrasi	: *Management Information System* (MIS)
Judul Tesis	: "Analisis Algoritma *Naive Bayes Classifier* dengan seleksi atribut berbasis *Particle Swarm Optimization* untuk Klasifikasi Pelanggan *Cross selling* Produk Indihome PT. Telkom Jakarta"

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, 20 Febuari 2016
Pascasarjana Magister Ilmu Komputer
STMIK Nusa Mandiri
Direktur

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

D E W A N P E N G U J I

Penguji I	: Dr. Said Mirza Pahlevi, M.Eng.	.....
Penguji II	: Dr. Windu Gata, M.Kom.	.....
Penguji III / Pembimbing	: Dr. Sfenrianto, M. Kom.	.....

iv

KATA PENGANTAR

Puji syukur alhamdullillah, penulis panjatkan kehadirat Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya, sehingga pada akhirnya penulis dapat menyelesaikan tesis ini tepat pada waktunya. Dimana tesis ini penulis sajikan dalam bentuk buku yang sederhana. Adapun judul tesis, yang penulis ambil sebagai berikut "Analisis Algoritma *Naive Bayes Classifier* dengan seleksi atribut berbasis *Particle Swarm Optimization* untuk Klasifikasi Pelanggan *Cross selling* Produk Indihome PT. Telkom Jakarta".

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

Tesis ini diambil berdasarkan hasil penelitian atau riset mengenai *data mining Naive Bayes Classifier*. Penulis juga lakukan mencari dan menganalisa berbagai macam sumber referensi, baik dalam bentuk jurnal ilmiah, buku-buku literatur, *internet*, dll yang terkait dengan pembahasan pada tesis ini.

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

1. Bapak Dr. Sfenrianto, M.Kom. selaku pembimbing tesis yang telah menyediakan waktu, pikiran dan tenaga dalam membimbing penulis menyelesaikan tesis ini.
2. Bapak Eddy Djunaedy (alm) dan Ibu Suharti atas dukungan moril dan doa-doanya yang selalu menyertai penulis.
3. Suami dan anak-anak tercinta Hylmi dan Latifa sebagai penyemangat dalam menyelesaikan tesis ini..
4. Seluruh staf pengajar (dosen) Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri yang telah memberikan ilmu yang bermanfaat bagi penulis selama menempuh studi.

5. Seluruh staf dan karyawan Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri yang telah melayani penulis dengan baik selama kuliah.
6. 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, 20 Febuari 2016

Penulis

Indah Purnamasari

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini, saya :

Nama : Indah Purnamasari
NIM : 14001240
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Management Information System* (MIS)
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 : "Analisis Algoritma *Naive Bayes Classifier* dengan seleksi atribut berbasis *Particle Swarm Optimization* untuk Klasifikasi Pelanggan *Cross selling* Produk Indihome PT. Telkom Jakarta" beserta perangkat yang diperlukan (apabila ada).

Dengan **Hak Bebas Royalti Non-Eksklusif** ini pihak STMIK Nusa Mandiri berhak menyimpan, mengalih-media atau *bentuk*-kan, mengelolaannya 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, 20 Febuari 2016

Yang menyatakan,

Indah Purnamasari

ABSTRAKSI

Nama	: Indah Purnamasari
NIM	: 14001240
Program Studi	: Magister Ilmu Komputer
Jenjang	: Strata Dua (S2)
Konsentrasi	: *Management Information System* (MIS)
Judul	: "Analisis Algoritma *Naive Bayes Classifier* dengan seleksi atribut berbasis *Particle Swarm Optimization* untuk Klasifikasi Pelanggan *Cross selling* Produk Indihome PT. Telkom Jakarta "

Salah satu strategi pemasaran PT. TELKOM yaitu dengan *cross selling.* Selain untuk memperoleh keuntungan dari produk tambahan yang dijual, juga untuk meningkatkan ketergantungan pelanggan pada vendor sehingga mengurangi *churn*. Pemasaran *cross selling* dapat dilakukan dengan menawarkan produk baru. Tetapi terlalu banyak penawaran terhadap pelanggan yang tidak tepat, menjadikan pemasaran tidak efisien dan efektif. Untuk itu perlu dilakukan klasifikasi pelanggan telepon dan speedy eksisting PT. TELKOM Jakarta untuk *cross selling* produk Indihome. Penelitian ini menggunakan *data mining* klasifikasi dengan algoritma *Naive Bayes Classifier* dan seleksi atribut *Particle Swarm Optimization* untuk meningkatkan akurasi klasifikasi dari 85,08 % menjadi 89,31%.

Kata kunci : *Naive Bayes Classification (NBC), Particle Swarm Optimization (PSO),* Indihome, *Data Mining*

ABSTRACT

Name	: Indah Purnamasari
NIM	: 14001240
Study of Program	: Magister Ilmu Komputer
Levels	: Strata Dua (S2)
Concentration	: *Management Information System (MIS)*
Title	: *"Analysis Naive Bayes Classifier Algorithm with Selection Attributes based Particle Swarm Optimization Classification for Cross Selling Customers Indihome Products PT. Telkom Jakarta"*

One of PT. TELKOM marketing strategy is by cross selling. In addition to the benefit of additional products sold, as well as to enhance the customer's dependence on vendors, thereby reducing churn. Marketing cross selling can be done by offering new products. But too many deals to customers who are not appropriate, make marketing inefficient and ineffective. It is necessary for the classification of existing telephone customers and speedy PT. TELKOM Jakarta *for cross selling Indihome. This study uses data mining classification with Naive Bayes classifier algorithm and selection of attributes Particle Swarm Optimization to improve the classification accuracy of 85,08% to 89,31%.*

Keyword : Naive Bayes Classification (NBC), *Particle Swarm Optimization* (PSO), *Indihome, Data Mining*

DAFTAR ISI

Halaman
HALAMAN SAMPUL i
HALAMAN JUDUL ii
SURAT PERNYATAAN ORISINALITAS iii
HALAMAN PENGESAHAN iv
KATA PENGANTAR v
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH
UNTUK KEPENTINGAN AKADEMIS vii
ABSTRAKSI viii
ABSTRACT ix
DAFTAR ISI x
DAFTAR GAMBAR xii
DAFTAR TABEL xiii
DAFTAR LAMPIRAN xiv
BAB I PENDAHULUAN 1
1.1 Latar Belakang Masalah 1
1.2 Identifikasi dan Rumusan Masalah Penelitian 3
1.3 Tujuan Penelitian 3
1.4 Ruang Lingkup Penelitian 4
1.5 Sistematika Penulisan 4
BAB II LANDASAN TEORI 5
2.1 Tinjauan Pustaka 5
2.1.1 *Data Mining* 5
2.1.2 *Naive Bayes Classifier* (NBC) 8
2.1.3 *Particle Swarm Optimization* (PSO) 10
2.1.4 Evaluasi dan Validasi 11
2.1.5 Indihome 15
2.2 Tinjauan Studi 17
2.3 Tinjauan Objek Penelitian 21
BAB III METODOLOGI PENELITIAN 22
3.1 Desain Penelitian 22
3.2 Pengumpulan Data 23
3.3 *Preprocessing Data* 23
3.4 Pendekatan *Data Mining* 25
3.5 Eksperimen 25
3.6 Pengukuran 25
3.7 Hasil 26
3.8 Penerapan 27
BAB IV HASIL ANALISA DAN PEMBAHASAN 28
4.1 Pengumpulan Data 28
4.2 *Preprocessing Data* 30
4.3 Pendekatan *Data Mining* 33
4.3.1 Algoritma *Naive Bayes Classifier* 33
4.3.2 Algoritma *Particle Swarm Optimization - Naive Bayes* 35
4.4 Eksperimen 37
4.4.1 Algoritma *Naive Bayes Classifier* 38

4.4.2 Algoritma *Particle Swarm Optimization-Naive Bayes* 39
4.5 Pengukuran 41
4.5.1 Algoritma *Naive Bayes Classifier* 41
4.5.2 Algoritma *Particle Swarm Optimization - Naive Bayes* 42
4.6 Hasil Analisa 44
4.6.1 Algoritma *Naive Bayes Classifier* 44
4.6.2 Algoritma *Particle Swarm Optimization - Naive Bayes* 44
4.7 Pembahasan 46
4.7.1 Peningkatan akurasi 46
4.7.2 Uji *T-test* 49
4.8 Penerapan 49
4.9 Implikasi Penelitian 51
BAB V PENUTUP 53
5.1 Kesimpulan 53
5.2 Saran 53
DAFTAR REFERENSI 54
DAFTAR RIWAYAT HIDUP 56
SURAT KETERANGAN RISET 57
KARTU BIMBINGAN TESIS 58
LAMPIRAN 59

DAFTAR GAMBAR

Halaman

Gambar 2.1 Proses KDD 7
Gambar 2.2 *Ten cross validation* 12
Gambar 2.3 ROC *Curve* 14
Gambar 2.4 Layanan Triple Play Indihome 15
Gambar 3.1 Tahapan Penelitian 22
Gambar 4.1 Cara melihat missing value 38
Gambar 4.2 Replace missing value 38
Gambar 4.3 Model Naive Bayes Classifier 39
Gambar 4.4 Model Particle Swarm Optimization - Naive Bayes Classifier 40
Gambar 4.5 Validasi model Naive Bayes Classifier 41
Gambar 4.6 Pengukuran model Naive Bayes Classifier 42
Gambar 4.7 Validasi model Particle Swarm Optimization - Naive Bayes Classifier 43
Gambar 4.8 Pengukuran model Particle Swarm Optimization - Naive Bayes Classifier 44
Gambar 4.9 Performansi Naive Bayes Classifier 46
Gambar 4.10 *ROC Curve Naive Bayes Classifier* 47
Gambar 4.11 Performansi Particle Swarm Optimization - Naive Bayes Classifier 48
Gambar 4.12 *ROC Curve Particle Swarm Optimization - Naive Bayes Classifier* 48
Gambar 4.13 Uji T-test 49
Gambar 4.14 GUI Sistem Klasifikasi Cross Selling Produk Indihome 50
Gambar 4.15 Hasil klasifikasi GUI 50

DAFTAR TABEL

Halaman

Tabel 2.1 Confusion Matrix 13
Tabel 2.2 Penelitian yang Relevan 19
Tabel 4.1 Data pelanggan telepon dan speedy yang digunakan 31
Tabel 4.2 Missing Value data 32
Tabel 4.3 Hasil Replace Missing Value 33
Tabel 4.4 Probabilitas Posterior NBC 34
Tabel 4.5 Probabilitas Posterior PSO-NBC 36
Tabel 4.6 Spesifikasi hardware dan software 37
Tabel 4.7 Hasil performansi pendekatan Naive Bayes Classifier 44
Tabel 4.8 Hasil seleksi atribut Particle Swarm Optimization 45
Tabel 4.9 Hasil performansi pendekatan Particle Swarm Optimization-Naive Bayes Clasifier 45
Tabel 4.10 Nilai perbandingan performansi 46

DAFTAR LAMPIRAN

Halaman

Lampiran 1 Tarif berdasarkan Paket eksisting dan Zona 59
Lampiran 2 Perhitungan Probabilitas Prior NBC..... 61
Lampiran 3 Probabilitas Prior PSO-NBC 63
Lampiran 4 Data pelanggan eksisting dan Hasil penelitian 64

BAB 1 PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi telekomunikasi dan informasi semakin pesat dari waktu ke waktu sehingga memudahkan kita untuk mendapatkan berbagai informasi yang kita butuhkan dari belahan dunia manapun, terutama saat teknologi internet mulai diperkenalkan. Saat ini internet dapat diakses oleh semua kalangan dengan mudah. Kemajuan teknologi internet telah membuka peluang bisnis tersendiri bagi perusahaan yang bergerak di bidang telekomunikasi dan informasi. Berbagai perusahaan telekomunikasi dan informasi menyediakan akses kecepatan tinggi (*broadband*) untuk layanan-layanan berbasis internet dengan biaya terjangkau.

PT. Telkom sebagai perusahaan yang bergerak di bidang telekomunikasi dan informasi di Indonesia memperkenalkan layanan *Internet Protocol Television* (IPTV) yaitu metode penyaluran gambar dan suara televisi melalui IP pada tahun 2011 dengan nama produk Groovia TV. Mulai tahun 2015 layanan Groovia TV mengalami evolusi menjadi produk Indihome yaitu produk *bundling* yang memberikan layanan *Triple Play* dari PT. Telkom yang terdiri dari 3 layanan yaitu Internet Fiber atau *High Speed Internet* (Internet cepat), *Interactive* TV (Usee TV) dan Telepon Rumah dengan menggunakan 100% fiber optic artinya fiber optic digelar sampai ke rumah pelanggan (*Fiber To The Home /* FTTH).

Strategi pemasaran terhadap pelanggan yang telah ada (eksisting) disebut dengan *cross selling. Cross-selling* adalah strategi menjual produk baru untuk pelanggan yang telah melakukan pembelian lainnya sebelumnya. Selain untuk memperoleh keuntungan dari produk tambahan yang dijual, juga meningkatkan ketergantungan pelanggan pada vendor sehingga mengurangi churn (Jaroszewicz, 2008). Strategi pemasaran *cross selling* dilakukan karena pesatnya pertumbuhan telekomunikasi yang semakin jenuh dimana layanan komunikasi suara biasa sudah sangat umum sehingga perusahaan telekomunikasi mengalami kesulitan dalam meningkatkan kuantitas pelanggan mereka (Ahn et al., 2011). Pemasaran untuk pelanggan baru akan lebih mahal daripada mempertahankan pelanggan setia yang sudah ada (Sharma & Panigrahi, 2011). Oleh karena itu, perusahaan menjual

produk baru terhadap pelanggan yang telah ada untuk meningkatkan keuntungan perusahaan.

Pemasaran *cross selling* dapat dilakukan dengan menawarkan produk baru melalui SMS, panggilan telepon ke pelanggan dan surat ke pelanggan bersamaan dengan tagihan bulanan. Akan tetapi terlalu banyak penawaran terhadap pelanggan yang tidak tepat, hanya akan membuat kejenuhan di sisi pelanggan dan akan berdampak negatif yaitu pelanggan menolak produk yang ditawarkan (Jaroszewicz, 2008). Untuk itu agar pemasaran produk Indihome dapat lebih efisien dan efektif maka dilakukan klasifikasi kepada pelanggan telepon dan pelanggan speedy eksisting PT. Telkom Jakarta yang berpotensi untuk meningkatkkan layanannya menjadi pelanggan Indihome.

Banyak penelitian yang telah dilakukan dalam klasifikasi dengan algoritma *Naive Bayes Classifier* (NBC) diantaranya yaitu penelitian yang menyatakan bahwa NBC membutuhkan waktu lebih sedikit untuk membangun model (Karim & Rahman, 2013). Penelitian lain menggunakan metodologi CRISP-DM dengan algoritma NBC dapat digunakan pada database besar dengan banyak atribut (Moro & Laureano, 2011). NBC merupakan salah satu teknik data mining yang paling baik (Ting et al., 2011) dan telah berhasil diterapkan pada domain masalah yang berbeda dari tugas klasifikasi seperti prakiraan cuaca, intrusi deteksi, gambar dan pengenalan pola, diagnosa medis, persetujuan kredit dan bioinformatika dll (Kabir et al., 2011) (Bramer, 2007, p. 3).

Akan tetapi NBC memiliki kelemahan utama pada asumsi fitur *independent* sehingga dengan menggunakan seleksi atribut dapat meningkatkan kinerja *classifier* (Taha et al., 2013). Asumsi atribut *independent* membatasi penerapan data aktual sehingga kinerja Naive Bayes tidak sempurna (Li et al., 2014). *Naive Bayes Classifier* (NBC) merupakan *classifier* yang sederhana namun asumsi atribut NBC sering dilanggar dan sebagai hasilnya perkiraan probabilitas sering kurang optimal (Zaidi & Cerquides, 2013).

Klasifikasi dilakukan dengan metode *Naive Bayes Classifier* karena kesederhanaannya serta kecepatan komputasinya dan seleksi atribut PSO dapat meningkatkan akurasinya (Correa et al., 2008). Data optimum diseleksi oleh algoritma PSO (Ghanad & Ahmadi, 2015). Penggabungan seleksi atribut dan

algoritma klasifikasi akan menghilangkan fitur yang tidak relevan dan bising (Vihar et al., 2009). Penurunan fitur memiliki efek besar pada mempercepat algoritma data mining, dan meningkatkan kinerja pertambangan seperti akurasi prediksi dan hasil dengan komperhensif (Harb, 2014).

Penelitian ini bertujuan untuk mengklasifikasikan pelanggan telepon dan speedy untuk *cross selling* produk Indihome menggunakan algoritma *Naive Bayes Classifier* (NBC) dan *Particle Swarm Optimization* (PSO) sebagai seleksi atribut sehingga tingkat akurasi lebih tinggi.

1.2 Identifikasi dan Rumusan Masalah Penelitian

Untuk mencapai strategi pemasaran kepada pelanggan yang tepat maka perusahaan dapat melakukan klasifikasi pelanggan yang telah ada yang berpotensi untuk produk lain (*cross selling*). Klasifikasi pelanggan dilakukan berdasarkan data-data pelanggan eksisting di PT. TELKOM.

Naive Bayes Classifier (NBC) adalah algoritma sederhana dan tidak membutuhkan waktu yang lama dalam klasifikasi dengan jumlah data yang besar. Akan tetapi NBC memiliki kelemahan dalam asumsi atribut *independent* sehingga menyebabkan tingkat akurasi menjadi rendah.

Dari masalah tersebut diatas, maka rumusan masalah penelitian ini adalah sebagai berikut :

1. Bagaimana akurasi *Naive Bayes Classifier* untuk klasifikasi pelanggan eksisting *cross selling* produk Indihome ?
2. Bagaimana peningkatan akurasi *Naive Bayes Classifier* dengan seleksi atribut *Particle Swarm Optimization* untuk klasifikasi pelanggan eksisting *cross selling* produk Indihome ?

1.3 Tujuan Penelitian

Sesuai rumusan masalah diatas, maka penelitian ini memiliki tujuan antara lain sebagi berikut :

1. Memperoleh akurasi *Naive Bayes Classifier* untuk klasifikasi pelanggan eksisting *cross selling* produk Indihome.

2. Mengetahui peningkatan akurasi *Naive Bayes Classifier* dengan seleksi atribut *Particle Swarm Optimization* untuk klasifikasi pelanggan eksisting *cross selling* produk Indihome.

1.4 Ruang Lingkup Penelitian

Ruang lingkup penelitian ini dibatasi pada data mining klasifikasi metode *Naive Bayes Classifier* (NBC) dan *Naive Bayes Classifier* (NBC) berbasis *Particle Swarm Optimization* (PSO) untuk seleksi atribut. Penggunaan dataset diambil dari data primer pelanggan telepon dan speedy eksisting divisi *Bussiness Planning & Performance* PT. TELKOM Jakarta.

1.5 Sistematika Penulisan

Sistematika penulisan dalam penelitian ini disusun dalam 5 (lima) bab yaitu:

BAB 1. PENDAHULUAN

Bab ini terdiri dari latar belakang, identifikasi dan rumusan masalah, tujuan penelitian, ruang lingkup dan sistematika penulisan

BAB 2. LANDASAN TEORI

Bab ini berisi tentang tinjauan pustaka, tinjauan studi, tinjauan obyek penelitian.

BAB 3. METODE PENELITIAN

Bab ini membahas tentang desain, pengumpulan data, *preprocessing data*, pembagian data, pendekatan, eksperimen, pengukuran dan hasil penelitian.

BAB 4. HASIL PENELITIAN

Bab ini berisi pengumpulan data, *preprocessing data*, pembagian data, pendekatan, eksperimen, pengukuran, hasil, pembahasan dan implikasi penelitian.

BAB 5. PENUTUP

Bab ini mencakup kesimpulan dari penelitian ini dan saran untuk penelitian selanjutnya.

BAB 2 LANDASAN TEORI

2.1 Tinjauan Pustaka

2.1.1 Data Mining

Data mining adalah proses menemukan korelasi bermakna baru, pola dan trend oleh sejumlah data besar yang disimpan dalam repositori menggunakan teknologi pola serta teknik statistik dan matematika (Larose, 2005, p. 2). *Data mining* adalah proses analitik dirancang untuk mengeksplorasi sejumlah besar data mencari ilmu tersembunyi yang konsisten dan berharga, langkah pertama terdiri dalam eksplorasi dan data persiapan awal (Gorunescu, 2011, p. 57). *Data mining* didefinisikan sebagai proses menemukan pola dalam data(Witten et al., 2011, p. 5).

Proses pencarian informasi ini menggunakan teknik-teknik pembelajaran mesin (*machine learning*) untuk memeriksa *database* berukuran besar sebagai cara untuk menemukan pola yang baru dan berguna. Proses pencarian untuk menemukan pola atau model yang sahih, baru, bermanfaat, dan dimengerti dilakukan secara iteratif dengan menggunakan perangkat lunak dalam menganalisis data untuk menemukan pola dan relasi data agar dapat menghasilkan prediksi yang lebih akurat. Meskipun tidak semua pekerjaan pencarian informasi dinyatakan sebagai *data mining*, seperti halnya pencarian *situs* tertentu menggunakan *query search engine* merupakan pekerjaan pencarian informasi yang berhubungan dengan *information retrieval*. Akan tetapi kemampuan sistem-sistem *information retrieval* dapat ditingkatkan dengan menggunakan teknik-teknik *data mining*. Algoritma *data mining* berdasarkan tugasnya (Larose, 2005, p. 11), yaitu :

1. *Description*

   Deskripsi merupakan cara untuk menggambarkan pola dan tren dalam data yang memberikan saran dengan penjelasan untuk pola dan trend tersebut. Beberapa algoritma yang dapat digunakan yaitu *association* dan *decision trees.*

2. *Estimation*

   Estimasi seperti klasifikasi hanya variabel target numerik lebih dari kategori. Model yang dibangun menggunakan catatan lengkap, yang memberikan nilai

dari variabel target serta prediktor. Kemudian, untuk pengamatan baru, memperkirakan dari nilai variabel target yang dibuat, didasarkan pada nilai-nilai prediktor. Algoritma yang dapat digunakan yaitu *linier regression.*

3. *Prediction*

   Prediksi mirip dengan klasifikasi dan estimasi akan tetapi menghasilkan prediksi masa depan. Algoritma yang dapat digunakan yaitu *k-nearest neighbor, neural network, decision trees.*

4. *Classification*

   Dalam klasifikasi, ada kategori variabel sasaran. Model data mining memeriksa dataset, masing-masing dataset berisi informasi tentang variabel sasaran serta satu variabel prediktor. Langkah pertama, memeriksa kumpulan data yang mengandung variabel prediktor dan (sudah diklasifikasikan) variabel sasaran. Dengan cara ini, algoritma "belajar" tentang kombinasi variabel yang berhubungan. Kumpulan data ini disebut *training set*. Berdasarkan klasifikasi pada training set, algoritma akan menugaskan klasifikasi ke rekor baru. Algoritma yang digunakan dalam *classification* diantaranya yaitu *decision tree, bayesian classifier, neural networks.*

5. *Clustering*

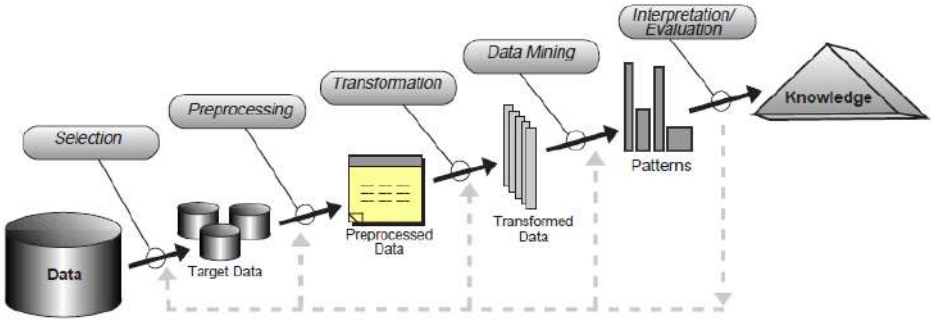
   Pada dasamya clustering terhadap data adalah suatu proses untuk mengelompokkan sekumpulan data tanpa suatu atribut kelas yang telah didefinisikarn sebelumnya *(unsupervised learning).* Melakukan analisa dengan *clustering*, akan sangat membantu untuk membentuk partisi-partisi yang berguna terhadap sejumlah besar himpunan obyek dengan didasarkan pada prinsip "*divide and conquer*" yang mendekomposisikan suatu sistem skala besar, menjadi komponen-komponen yang lebih kecil, untuk menyederhanakan proses desain dan implementasi. Algoritma yang digunakan dalam *clustering* diantaranya yaitu *K-Means, Neural Network.*

6. *Association Rules*

   Association Rules disebut juga *market bases analysis* yang digunakan untuk menemukan relasi atau korelasi antara himpunan. aturan asosiasi adalah dari bentuk "Jika mendahului, maka konsekuen" dengan ukuran dari *support* dan terkait dengan aturan. Metode ini termasuk metode *Unsupervised.* Algoritma

yang digunakan dalam *clustering* diantaranya yaitu *A Priori* dan Generalized Rule Induction (GRI).

Data mining merupakan salah satu tahap dari proses *Knowledge Discovery in Data mining* merupakan langkah dalam proses KDD yang terdiri dari analisis data menerapkan dan algoritma penemuan yang menghasilkan enumerasi tertentu pola (atau model) atas data (Fayyad et al., 1996). Proses KDD adalah sebagai berikut :



Gambar 2.1 Proses KDD

1. *Data Selection*

   Mengembangkan pemahaman tentang aplikasi utama dan pengetahuan yang relevan serta mengidentifikasi tujuan dari proses KDD dari sudut pandang pelanggan.

2. *Preprocessing*

   Pembersihan data dan *preprocessing* merupakan operasi dasar seperti penghapusan *noise*, pemungutan informasi yang diperlukan untuk model, memutuskan strategi untuk menangani data hilang (*missing value)*, menghitung urutan waktu untuk informasi dan mengetahui perubahan.

3. *Transformation*

   Transformasi merupakan pencocokan tujuan dari proses KDD ke metode data mining tertentu misalnya, klasifikasi, regresi, clustering, dll.

4. *Data Mining*

   Data Mining adalah proses mencari pola atau informasi menarik dari data terpilih dengan menggunakan algoritma atau metode tertentu bergantung pada fungsinya seperti yang telah disebutkan diatas.

5. *Interpretation / Evaluation*

   Tahap ini mencakup validasi dan evaluasi terhadap model yang diusulkan sehingga pengetahuan dapat digunakan untuk mendukung tujuan penelitian.

2.1.2 Naive Bayes Classifier *(NBC)*

Naive Bayes Classifier adalah salah satu algoritma klasifikasi data mining untuk menemukan model atau fungsi yang menjelaskan atau membedakan kelas data. Teorema Bayes dikemukakan oleh Reverand Thomas Bayes pada abad ke 18 (Bramer, 2007, p. 24-28). Berdasarkan teorema Bayes dikembangkan *Naive Bayes* yaitu pengklasifikasi dengan metode probabilistik dan statistik istilah "naif" menunjukkan keindependenan antara atribut kelas (Madyatmadja & Aryuni, 2014). Klasifikasi *Naive Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya. Asumsi keindependenan atribut akan menghilangkan kebutuhan banyaknya jumlah data latih dari perkalian kartesius seluruh atribut yang dibutuhkan untuk mengklasifikasikan suatu data.

Algoritma *Naïve Bayes* (NB) merupakan algoritma yang cepat, mudah diimplementasikan dengan struktur yang sederhana dan efektif dalam pengklasifikasian (Taheri & Mammadov, 2013). Selain itu NB terbukti memiliki akurasi yang tinggi saat diaplikasikan ke dalam database yang besar.

Pada dasarnya metode NBC menempuh dua tahap klasifikasi, yaitu tahap pelatihan dan tahap klasifikasi. Pada tahap pelatihan dilakukan proses analisis terhadap *sampel*. Selanjutnya adalah penentuan probabilitas *prior* bagi tiap kategori berdasarkan sampel yang ada. Pada algoritma *Naïve Bayes* (NB) didasari oleh Persamaan Teorema Bayes, yang membedakan adalah asumsi keindependenan antara atribut. Persamaan Bayes adalah (Gorunescu, 2011, p. 186-190):

$$\boldsymbol{P}(\boldsymbol{H}|\boldsymbol{X}) = \frac{\boldsymbol{P}(\boldsymbol{X}|\boldsymbol{H}).\ \boldsymbol{P}(\boldsymbol{H})}{\boldsymbol{P}(\boldsymbol{X})} \qquad (2.1)$$

Keterangan :

X : Data dengan class yang belum diketahui

H : Hipotesis data X merupakan suatu class spesifik

$P(H|X)$: Probabilitas hipotesis H berdasar kondisi X (*posterior probability*)

$P(H)$: Probabilitas hipotesis H (*prior probability*)

$P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$: Probabilitas X

Untuk menjelaskan teorema *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, teorema Bayes di atas disesuaikan sebagai berikut (Gorunescu, 2011, p. 186-190):

$$\boldsymbol{P}(\boldsymbol{H}|\boldsymbol{X}) = \boldsymbol{P}(\boldsymbol{X}|\boldsymbol{H}).\ \boldsymbol{P}(\mathbf{H}) \qquad (2.2)$$

Persamaan diatas merupakan model dari *Naive Bayes* yang selanjutnya akan digunakan dalam proses klasifikasi. Sedangkan untuk klasifikasi dengan data kontinyu digunakan rumus *Densitas Gauss* (Gorunescu, 2011, p. 190):

$$\boldsymbol{P}(\boldsymbol{X}, \boldsymbol{\mu}, \boldsymbol{\sigma}) = \frac{1}{\sqrt{2\pi\sigma}}\, e^{-\frac{(x-\mu)2}{2\sigma 2}} \qquad (2.3)$$

Keterangan :

P : Peluang

X : Atribut

μ : Mean, menyatakan rata – rata dari seluruh atribut

σ : Deviasi standar, menyatakan varian dari seluruh atribut

π = 3,1416

e = 2,7183

$$\sigma^2 = \frac{1}{N-1} \sum (X_i - \mu)^2 \qquad (2.4)$$

Akan tetapi pada kasus khusus untuk menghitung probabilitas posterior dimana P (X | C) = 0 selama |X| = 0 akan menghilangkan efek semua *posterior* P (C). Untuk

mengatasi hal ini, kita asumsikan dengan menambahkan nilai satu pada setiap perhitungan data pelatihan. Dalam perbedaan probabilitas perkiraan ini, perlu dibuat kesepakatan. Teknik ini dikenal sebagai koreksi Laplacian :

$$P(X \mid C) = \frac{1}{|C| + \text{jumlah kelas}} \quad (2.5)$$

2.1.3 Particle Swarm Optimization (PSO)

Metode PSO merupakan teknik optimasi yang dikembangkan oleh Kennedy dan Eberhart pada tahun 1995. *Particle Swarm Optimization* merupakan algoritma yang sangat sederhana yang efektif untuk mengoptimalkan berbagai fungsi. (Kennedy & Eberhart, 1995). PSO menggunakan sekumpulan partikel yang saling bekerjasama. Setiap partikel merepresentasikan suatu solusi sehingga metode PSO dapat mengeksplorasi solusi-solusi yang memungkinkan bagi permasalahan optimasi. *Particle Swarm Optimization* (PSO) adalah algoritma pencarian berbasis populasi yang terinspirasi oleh perilaku komunitas biologis baik individu dan perilaku sosial, contoh komunitas-komunitas ini kawanan burung, sekolah ikan dan kawanan lebah. Anggota masyarakat seperti berbagi tujuan bersama (misalnya, menemukan makanan) yang diwujudkan dengan mengeksplorasi lingkungannya saat berinteraksi di antara mereka (Escalante, 2009).

Dalam PSO setiap solusi untuk masalah yang dihadapi disebut partikel. Variabel *i* merupakan partikel ke-*i* dalam kawanan, t merupakan jumlah iterasi, *Vi* adalah kecepatan partikel ke-*i* dan *Xi* adalah variabel partikel vektor (contohnya posisi vektor) dari partikel ke-*i* pada permasalahan *N* dimensional. Masing-masing partikel akan diinisialisasi secara acak atau heuristik. Kemudian pada setiap langkah optimasi, partikel-partikel tersebut diperbolehkan untuk mengevaluasi kemampuannya dan kemampuan partikel-partikel di sekitarnya dinyatakan sebagai *pbest*. Masing-masing partikel dapat menyimpan solusi yang menghasilkan kemampuan terbaik sebagai salah satu kandidat solusi terbaik untuk semua partikel disekitarnya dinyatakan sebagai *gbest*. Masing-masing partikel akan berpindah berdasarkan jarak dan kecepatan. Partikel memperbaharui kecepatan (*velocity*) dengan persamaan:

$$\boldsymbol{v_{id}} = \boldsymbol{\omega} * \boldsymbol{v_{id}} + \boldsymbol{\eta_X} * \boldsymbol{R}(\quad) * (\boldsymbol{p_{id}} - \boldsymbol{x_{id}}) + \boldsymbol{\eta_2} * \boldsymbol{R}(\quad) * (\boldsymbol{p_{gd}} - \boldsymbol{x_{id}}) \quad (2.6)$$

$$\boldsymbol{x_{id}} = \boldsymbol{x_{id}} + \boldsymbol{v_{id}} \quad (2.7)$$

Keterangan:

$\boldsymbol{v_{id}}$: kecepatan partikel ke-i pada iterasi ke-i

$\boldsymbol{\omega}$: faktor bobot inersia

$\boldsymbol{\eta_{X,2}}$: konstanta akeselerasi (*learning rate)*

$\boldsymbol{R}$: bilangan random (0-1)

$\boldsymbol{x_{id}}$: posisi saat ini dari partikel ke-i pada iterasi ke-i

$\boldsymbol{p_{id}}$: posisi terbaik sebelumnya dari partikel ke-i

$\boldsymbol{p_{gd}}$: partikel terbaik diantara semua partikel dalam satu kelompok/populasi

Proses dasar dari algoritma PSO adalah :

1. Menyiapkan sampel dataset untuk menghitung bobot dari record pertama
2. *Initialization* yaitu inisial partikel saat t = 0 maka diketahui lokasi awal xi(t) dari setiap partikel .
3. *Fitness* yaitu mengukur kemampuan dari masing-masing partikel dari populasi.
4. *Update* yaitu menghitung *velocity* dari masing-masing partikel.
5. Membandingkan kinerja masing-masing partikel dengan pertikel yang memiliki kinerja terbaik *Pbest.* Membandingkan kinerja masing-masing partikel dengan pertikel secara global *Gbest*.
6. *Termination* : memberhentikan algoritma jika *termination criterion* buruk dan kembali ke langkah 2.

Iterasi akan berhenti jika jumlah iterasi telah mencapai nilai maksimal dari jumlah iterasi. Dengan menerapkan PSO, diharapkan dapat meningkatkan *accuracy* pada c*lassification algorithm.*

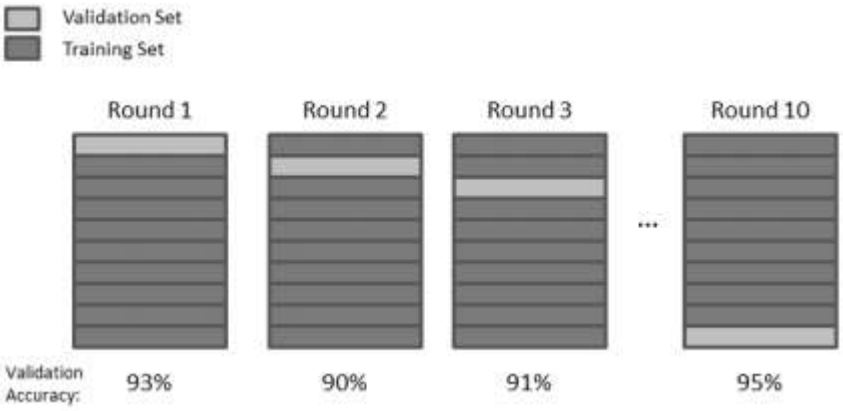
2.1.4 Evaluasi dan Validasi

Evaluasi dan validasi untuk menguji model pada penelitian ini menggunakan *cross validation, confusion matriks, ROC curve, T-test.*

1. *Cross Validation*

Validasi merupakan proses evaluasi untuk mengukur keakuratan prediksi suatu model. Tahap validasi mewakili langkah paling penting dalam proses membangun sebuah model. Sebuah pilihan yang baik dari parameter k dapat dilakukan dengan menggunakan metode cross-validasi. Jika dataset terdiri N baris data (*instance*) kemudian dibagi menjadi k bagian yang sama (jika N tidak tepat habis dibagi k, bagian akhir akan memiliki lebih sedikit contoh dari k lainnya -. 1 bagian). Masing-masing bagian k pada gilirannya digunakan sebagai data test dan k - 1 bagian lainnya digunakan sebagai training set (Bramer, 2007, p. 82). Cara kerja K-*fold cross validation* adalah sebagai berikut :

1. Total *instance* dibagi menjadi k bagian.
2. *Fold* ke-1 : bagian ke-1 menjadi data uji (*testing* data) dan sisanya menjadi data latih (*training* data). Selanjutnya, hitung akurasi berdasarkan porsi data tersebut.
3. *Fold* ke-2 : bagian ke-2 menjadi data uji (*testing* data) dan sisanya menjadi data latih (*training* data). Selanjutnya, hitung akurasi berdasarkan porsi data tersebut.
4. Demikian seterusnya hingga mencapai *fold* ke-K.



Gambar 2.2 *Ten cross validation*

2. *Confusion Matriks*

Confusion matrix memberikan keputusan yang diperoleh dalam traning dan testing. C*onfusion matrix* memberikan penilaian *performance* klasifikasi berdasarkan objek dengan benar atau salah. Confusion Matrix merupakan alat

untuk menganalisis seberapa baik *classifier* dapat mengenali tupel dari kelas yang berbeda (Jiawei & Micheline, 2006) . *Confusion matrix* berisi informasi aktual (actual) dan prediksi (predicted) pada sistem klasifikasi. Rumus yang berkaitan dengan perhitungan akurasi adalah sebagai berikut (Gorunescu, 2011, p. 320) :

Tabel 2.1 *Confusion Matrix*

CLASSIFICATION	*PREDICTED CLASS*	
	Class=Yes	*Class=No*
Class=Yes	*A (True Positive - tp)*	*B (False Negative - fn)*
Class=No	*C (False Positive - fp)*	*D (True Negative - tn)*

Berikut adalah persamaan model *confusion matrix* untuk menghitung akurasi. Nilai akurasi adalah proporsi jumlah prediksi yang benar. Dapat dihitung dengan menggunakan persamaan:

$$\text{Akurasi} = \frac{tp + tn}{tp + tn + fp + fn}. \qquad (2.8)$$

False positive disebut dengan error tipe 1, yang terjadi ketika kasus seharusnya diklasifikasikan sebagai *negative* akan tetapi diklasifikasikan sebagai *positive*. Sedangkan *False negative* dikenal sebagai *error* tipe 2, terjadi ketika kasus yang seharusnya diklasifikasikan sebagai *positive* akan tetapi diklasifikasi sebagai *negative* (Kohavi and John, 1997).

3. *ROC Curve*

Curve ROC (Receiver Operating Characteristic) adalah cara lain untuk mengevaluasi akurasi dari klasifikasi secara visual. Sebuah grafik ROC adalah plot dua dimensi dengan proporsi positif salah (fp) pada sumbu X dan proporsi positif benar (tp) pada sumbu Y. Titik (0,1) merupakan klasifikasi yang sempurna terhadap semua kasus positif dan kasus negatif. Nilai positif salah adalah tidak ada (fp = 0) dan nilai positif benar adalah tinggi (tp = 1). Titik (0,0) adalah klasifikasi yang memprediksi setiap kasus menjadi negatif {-1}, dan titik (1,1) adalah klasifikasi yang memprediksi setiap kasus menjadi positif {1}.

Grafik ROC menggambarkan *trade-off* antara manfaat (‘*true positives’*) dan biaya (‘*false positives’*). Dapat disimpulkan bahwa, satu point pada kurva ROC adalah lebih baik dari pada yang lainnya jika arah garis melintang dari kiri bawah ke kanan atas didalam grafik. Sebuah panduan untuk mengklasifikasikan keakuratan tes menggunakan AUC disajikan di bawah ini (Gorunescu, 2011, p. 323-325)

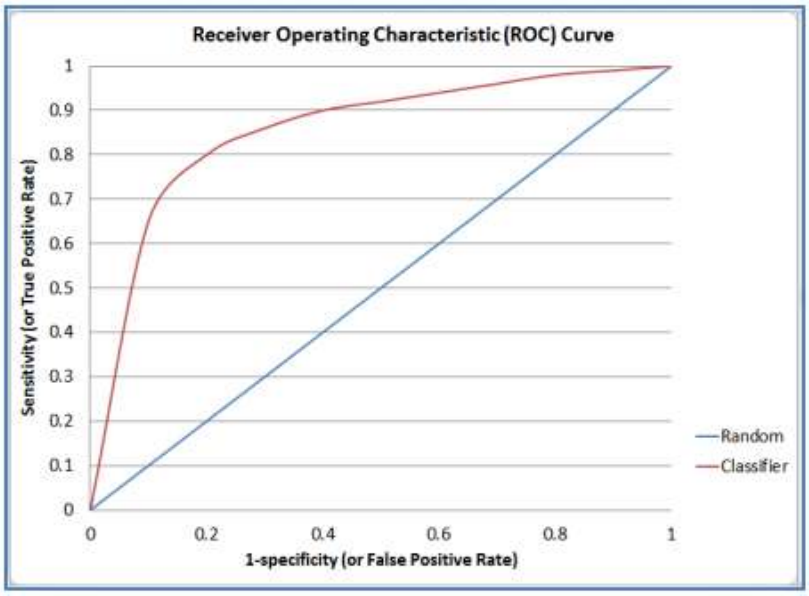
Akurasi 0.90 – 1.00 = *Excellent classification*

Akurasi 0.80 – 0.90 = *Good classification*

Akurasi 0.70 – 0.80 = *Fair classification*

Akurasi 0.60 – 0.70 = *Poor classification*

Akurasi 0.50 – 0.60 = *Failure*



Gambar 2.3 ROC *Curve*

4. *T-test*

Analisis *Paired-Samples T Test* merupakan prosedur yang digunakan untuk membandingkan rata-rata dua variabel dalam satu group. Artinya pula analisis ini berguna untuk melakukan pengujian terhadap dua sampel yang berhubungan atau dua sampel berpasangan.(Elliott & Woodward, 2007)

2.1.5 Indihome

Indihome Internet Fiber merupakan layanan *Triple Play* dari PT. TELKOM yaitu suatu produk terbaru PT. TELKOM yang terdiri dari 3 layanan sekaligus antara lain *High Speed Internet on Fiber* (Internet cepat), *Interactive* TV (Usee TV) dan Telepon Rumah dengan menggunakan 100% fiber optic artinya fiber optic digelar sampai ke rumah pelanggan (*Fiber To The Home /* FTTH). Indihome juga memiliki beberapa atribut tambahan seperti wifi.id *Seamless*, Indihome Telkomsel Mania, Indihome *Global Call*, Indihome *view*, MelOn dan *Trend Micro Security System*.

Gambar 2.4 Layanan *Triple Play* Indihome

High Speed Internet on Fiber merupakan layanan internet cepat yang menggunakan media akses jaringan fiber optik dengan bandwidth sampai dengan 100 Mbps. Berbeda dengan *High Speed Internet* biasa yang menggunakan media akses jaringan tembaga (*copper*) atau coaxial dengan bandwidth hanya 1 Mbps sampai 5 Mbps. *High Speed Internet on Fiber* mempunyai beberapa kelebihan, diantaranya adalah kecepatan yang lebih stabil pada saat melakukan sharing akses internet secara bersamaan, lebih tahan dalam kondisi cuaca apapun seperti serangan petir dan gangguan elektromagnet sehingga komputer yang digunakan menjadi lebih aman dan tidak perlu perawatan secara berkala.

Usee TV adalah layanan televisi berbasis internet protokol. Penyelenggaraan layanan televisi protokol internet (*Internet Protocol* Television) IPTV diatur oleh Peraturan Menteri Komunikasi dan Informatika Nomor 15 Tahun 2014 yang

merupakan perubahan dari Peraturan Menteri Komunikasi dan Informatika Nomor: 11/PER/M.KOMINFO/07/2010 tentang penyelenggaraan layanan IPTV. *Usee TV Cable* merupakan layanan televisi interaktif yang dapat dimiliki oleh perorangan dan berteknologi internet protocol yang dilengkapi fitur-fitur unggulan seperti *TV on Demand, Video on Demand, Pause and Rewind, Video Recorder*. *TV on Demand* merupakan Layanan yang dapat digunakan untuk menayangkan ulang acara-acara *live* TV tertentu dengan rentang waktu 7 (tujuh) hari ke belakang. Misalnya, pertandingan sepak bola yang pertandingannya telah lewat beberapa hari yang lalu, dapat disaksikan kembali. Sedangkan fitur Pause and Rewind pada Usee TV adalah layanan pada TV yang dapat dihentikan (*pause*) atau dimainkan kembali (*play*) tayangan program TV yang sedang berjalan. Misalnya, tayangan yang sedang disaksikan dapat dihentikan sementara apabila ingin melakukan aktivitas yang lain, dan melanjutkannya kembali setelah aktivitas tersebut selesai. *Video Recorder* pada Usee TV merupakan layanan penyimpanan yang disediakan oleh Usee TV dan dapat digunakan untuk menyimpan program pilihan dari berbagai channel yang tersedia. Selain itu juga dapat merekam video/program TV yang sedang tayang maupun yang akan tayang. Tayangan yang tersimpan melalui fasilitas *video recorder* ini akan tersimpan selama yang diinginkan dan dapat dihapus secara manual.

Telepon rumah pada Indihome adalah layanan komunikasi suara dengan biaya yang lebih murah serta kualitas suara yang lebih jernih dengan media akses jaringan fiber optik.

Fitur tambahan wifi.id *Seamless* merupakan layanan tambahan yang diberikan kepada pelanggan Indihome agar selalu bisa terkoneksi dengan mudah, murah dan *unlimited* melalui akses internet wifi cepat ke jaringan @wifi.id di seluruh Indonesia secara otomatis (*seamless*). Satu id (*username dan password*) pelanggan indihome dapat digunakan untuk lima perangkat yang berbeda (android, iphone, ipad, notebook dll).

Indihome Telkomsel Mania merupakan layanan tambahan yang diberikan kepada pelanggan Indihome untuk menelpon sepuasnya ke semua nomor dengan operator telkomsel dari telepon rumah.

Indihome *Global Call* merupakan layanan tambahan yang diberikan kepada pelanggan Indihome untuk menelpon hemat ke luar negeri melalui Telkom Global TG01017.

Indihome *view* merupakan layanan tambahan yang diberikan kepada pelanggan Indihome untuk dapat menikmati *live camera* dimana pengguna dapat menggunakan *live access* dan *recorded video* dengan proses pemasangan yang sangat mudah yaitu dengan menggunakan *plug & play* IP-cam melalui perangkat android atau IOS.

MelOn Indonesia adalah layanan tambahan yang diberikan kepada pelanggan Indihome yang merupakan portal musik digital yang menyediakan beragam musik dari berbagai genre lokal maupun mancanegara.

Trend Micro Security Sysem adalah layanan tambahan yang diberikan kepada pelanggan Indihome berupa layanan keamanan internet (*security)* dengan aplikasi Trend Micro sebagai platform sehingga komputer pelanggan akan terlindungi dari serangan virus, malware, spyware, spam, phising dan konten yang tidak layak dari internet sehingga data dan sistem aplikasi terbebas dari gangguan-gangguan tersebut.

2.2 Tinjauan Studi

Penerapan data mining yang digunakan untuk penjualan produk lain (*cross selling*) terhadap pelanggan eksisting telah dilakukan penelitian sebelumnya, diantaranya yaitu:

1. Penelitian Hyunchul Ahn, Jae Joon Ahn, Kyong Joo Oh, Dong Ha Kim (2011) melakukan penelitian untuk klasifikasi pelanggan akan membeli layanan baru atau tidak. Layanan baru yag ditawarkan adalah VAS (V*alue Added S*ervice) berupa bel melodi (atau dering nada di Amerika Serikat) layanan download. Model klasifikasi pelanggan untuk *cross-selling* di pasar telekomunikasi seluler ini dilakukan dengan menggabungkan hasil prediksi logistik regresi, jaringan saraf tiruan, dan pohon keputusan menggunakan Genetic Algorithm (GAOW) yang menghasilkan akurasi training data set 66,60% dan akurasi hold out dataset 66,02%

2. Penelitian Szymon Jaroszewicz (2008) melakukan penelitian untuk klasifikasi pelanggan akan membeli layanan baru atau tidak. Layanan baru yang ditawarkan ada tiga layanan opsional yang memungkinkan pelanggan untuk menurunkan biaya connection. Layanan dijelaskan sebagai berikut:

   - RL : panggilan lokal murah untuk panggilan dalam area lokal pelanggan;

   - TPG : harga panggilan yang lebih rendah untuk panggilan yang dibuat setelah 6 p.m;

   - TPWS: panggilan murah untuk panggilan ke pengguna lain dari jaringan yang telah ada.

   Penelitian ini menghasilkan akurasi dengan algoritma naive bayes 20,54%, decision tree 27,93%, AdaBoostMI 26,19, *support vector machine 28,15%*..
3. Penelitian (Ghanad & Ahmadi 2015) membandingkkan akurasi beberapa metode sebelumnya yaitu *Support Vector Machine* (SVM), *Decision Tree* (DT), *Nearest Neighbour* (NN), *Genetic Algorithm* (GA) dengan metode yang ditawarkan oleh penelitiannya yaitu menggunakan kombinasi *Particle Swarm Optimization* (PSO) dan *Naive Bayes* (NB) memperoleh nilai akurasi tertinggi dengan cara mencari jumlah data pelatihan terbaik yaitu 8 dataset dengan akurasi 97.95%.
4. Penelitian Masud Karim, Rashedur M. Rahman (2013) bertujuan untuk memprediksi apakah klien akan berlangganan deposito berjangka dengan menyelidiki dua teknik data mining yaitu *Naive Bayes* dan pohon keputusan C4.5. Evaluasi kinerja dari algoritma klasifikasi biasanya pada akurasinya. Namun, faktor-faktor lain, seperti waktu atau ruang komputasi juga dianggap memiliki gambaran penuh dari setiap algoritma Dari penelitian tersebut, akurasi C4.5 yaitu 0,94 lebih baik dari NB yaitu 0,85 namun C4.5 membutuhkan lebih banyak waktu untuk membangun model yaitu 6.36 detik sedangkan *Naive Bayes* 0.58 detik.
5. Sergio Moro, Raul M. S. Laureano dan Paulo Cortez (2011) menggunakan metodologi *Cross-Industry Standard Process Data Mining* (CRISP-DM) yang terdiri dari 6 fase untuk meningkatkan efisiensi pemasaran dengan mengurangi jumlah untuk melakukan kontak. Penelitian menggunakan database internal lembaga bank. Algoritma yang digunakan untuk membuat model adalah *Naive Bayes (NB), Decision Tree (DT) dan Support Vector Machine (SVM)*. Pada

iterasi pertama dengan dataset besar yaitu 79354 *record* dan atribut yang digunakan mencapai 59 atribut, hanya dengan algoritma NB dapat tercapai hasil sedangkan algoritma DT dan SVM tidak mencapai hasil melainkan terus memproses dalam waktu lama. Pada iterasi kedua menggunakan 55817 *record* dan 53 atribut, algoritma NB dan DT dapat mencapai hasil sementara SVM tidak mencapai hasil. Sedangkan pada iterasi ketiga dengan 45211 *record* dan 29 atribut maka dataset dapat diproses dengan ketiga algoritma dimana SVM mencapai kinerja tertinggi AUC 0.938 dan ALIFT 0.887, NB yaitu AUC 0.87 dan ALIFT 0.827 dan DT yaitu AUC 0.868 dan ALIFT 0.79. Hal ini menunjukkan bahwa *Naive Bayes* dapat digunakan pada database besar dengan akurasi yang lebih baik dari *Decision Tree*.

6. Penelitian (Elsalamony, 2014) membandingkan akurasi algoritma *multilayer perception neural network* (MLPNN*), tree augmented Naïve Bayes* (TAN) dikenal sebagai *Bayesian networks, Nominal regression or logistic regression* (LR), dan *Ross Quinlan new decision tree model* (C5.0) dengan tujuan untuk memprediksi pemasaran ke klien yang berpotensi berlangganan deposito. Akurasi terbaik dihasilkan oleh MPLNN 90.49%.

Tabel 2.2 Penelitian yang Relevan

No.	**Nama Peneliti**	**Judul Penelitian**	**Masalah Penelitian**	**Metode**	**Kesimpulan**
1.	Hyunchul Ahn, Jae Joon Ahn, Kyong Joo Ohb, Dong Ha Kim	*Facilitating cross-selling in a mobile telecom market to develop customer classification model based on hybrid data mining techniques*	Mendapatkan algoritma DM yang dapat menyelesaikan klasifikasi pelanggan untuk layanan baru	logistik regresi, jaringan saraf tiruan, dan pohon keputusan dengan Genetic Algorithm (GAOW)	GAOW menghasilkan akurasi training data set 66,60% dan akurasi hold out dataset 66,02%

No.	Nama Peneliti	Judul Penelitian	Masalah Penelitian	Metode	Kesimpulan
2.	Szymon Jaroszewicz	*Cross-selling models for telecommunication Services*	Mendapatkan algoritma DM yang dapat menyelesaikan klasifikasi pelanggan untuk layanan baru	Algoritma NB, DT, J4.8, SVM	Akurasi terbaik menggunakan SVM 28,15%
3.	Navid Khozein Ghanad,Saheb Ahmadi	*Combination of PSO Algorithm and Naive Bayesian Classification for Parkinson Disease Diagnosis*	Mencari metode klasifasi	PSO dan NB	Kombinasi metode PSO dan NB memperoleh akurasi data pelatihan tertinggi 97.95%
4.	Masud Karim, Rashedur M. Rahman	*Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing*	Mendapatkan algoritma DM yang dapat menyelesaikan secara efisien	Algoritma NB dan C4.5	Akurasi C4.5 yaitu 0,94 lebih baik dari NB namun NB lebih cepat yaitu 0.58 untuk membangun model

No.	Nama Peneliti	Judul Penelitian	Masalah Penelitian	Metode	Kesimpulan
5.	Sergio Moro, Raul M. S. Laureano dan Paulo Cortez	*Using Data Mining For Bank Direct Marketing :An Application Of The CRISP-DM Methodology*	Meningkatkan efisiensi pemasaran menggunakan DM terbaik	CRISP-DM dengan algoritma NB, DT, SVM	NB dapat digunakan pada database besar dengan 59 atribut. SVM mencapai kinerja tertinggi AUC 0.938 dan ALIFT 0.887
6.	Hany A. Elsalamony	*Bank Direct Marketing Analysis of Data Mining Techniques*	Mencari algoritma klasifikasi untuk prediksi pemasaran	MPLNN, TAN, LR, C5.0	Akurasi terbaik dihasilkan oleh MPLNN 90.49%.

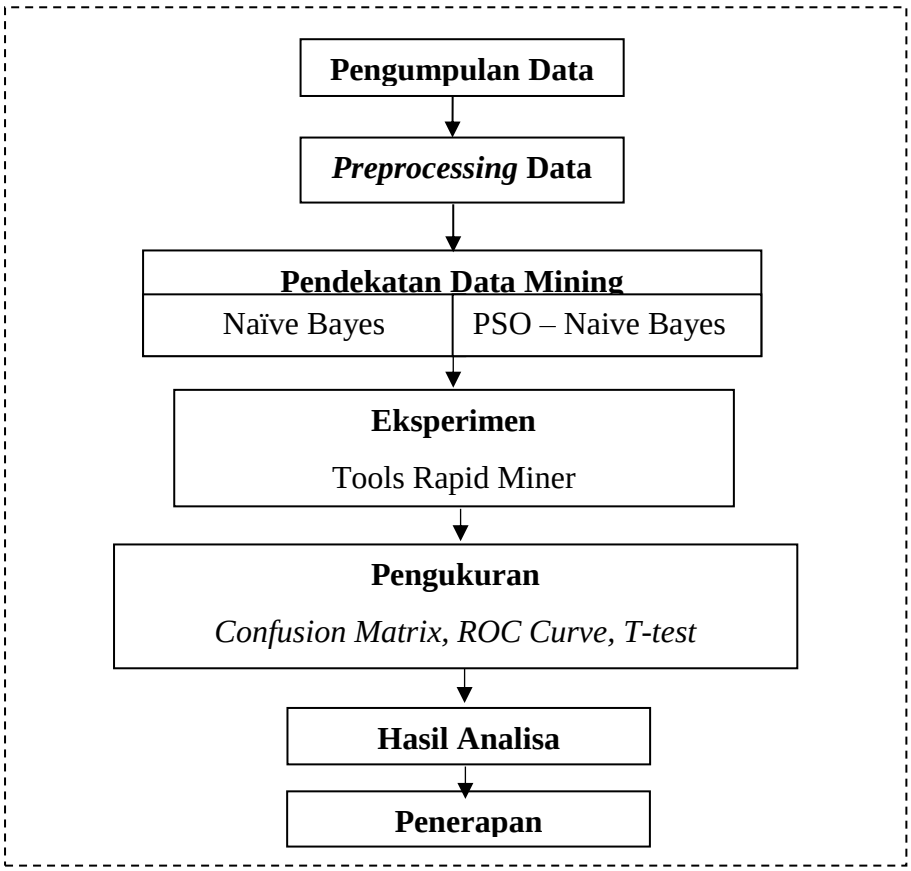
2.3 Tinjauan Objek Penelitian

Objek penelitian yang akan digunakan sebagai data dalam penelitian ini adalah data primer pelanggan telepon dan speedy *eksisting* dari divisi *Bussiness Planning and Performance* PT. Telkom Jakarta.

BAB 3 METODOLOGI PENELITIAN

3.1 Desain Penelitian

Pendekatan dalam penelitian terdiri dari pendekatan kualitatif dan pendekatan kuantitatif. Pendekatan kualitatif berkaitan dengan penilaian subjektif dari sikap, pendapat dan perilaku. Teknik pendekatan kualitatif yang digunakan secara umum adalah wawancara pada kelompok tertentu dengan mendalam (Kothari, 2004, p. 5). Sedangkan pendekatan kuantitatif digunakan untuk meneliti populasi atau *sample* tertentu, pengumpulan data menggunakan instrumen penelitian, analisis data bersifat kuantitatif atau statistik untuk menguji hipotesis yang telah ditentukan sebelumnya. Penelitian ini menggunakan pendekatan kuantitatif yang dapat disebut dengan metode *discovery* karena dengan metode ini dapat ditemukan dan dikembangkan berbagai ilmu pengetahuan dan teknologi baru. Tahapan dalam penelitian ini yaitu sebagai berikut :



Gambar 3.1 Tahapan Penelitian

Gambar diatas menunjukkan tahapan penelitian yang akan dilakukan yaitu :

1. Pengumpulan Data
2. *Preprocessing* Data
3. Pendekatan *Data Mining*
4. Eksperimen
5. Pengukuran
6. Hasil Analisa
7. Penerapan

3.2 Pengumpulan Data

Pada tahap ini dimulai dengan pengumpulan data awal, mengidentifikasi masalah kualitas data, menemukan wawasan awal tentang data dan mendeteksi subset menarik mengenai informaasi yang tersembunyi.

Sumber dalam penelitian terdiri dari sumber primer dan sumber sekunder. Sumber primer merupakan sumber obyek penelitian yang langsung memberikan data kepada pengumpul data penelitian sehingga data tersebut merupakan data primer. Data primer adalah data yang dikumpulkan pertama kali dan untuk melihat apa yang sesungguhnya terjadi (Kothari, 2004, p. 95). Sedangkan sumber sekunder merupakan sumber obyek penelitian yang tidak langsung memberikan data kepada pengumpul data penelitian melainkan melalui orang lain atau dokumen umum sehingga data tersebut merupakan data sekunder. Data sekunder adalah data yang sebelumnya pernah dibuat oleh seseorang baik data tersebut telah diterbitkan atau tidak (Kothari, 2004, p. 111).

3.3 *Preprocessing Data*

Persiapan data mencakup seluruh kegiatan yang diperlukan untuk membangun dataset yang akan dimasukkan ke dalam model. Data yang diperoleh tidak langsung diproses ke dalam model yang diusulkan melainkan melalui tahap persiapan untuk memperoleh data berkualitas tinggi sehingga menghasilkan hasil tambang dengan tingkat akurasi tinggi. Beberapa teknik yang dilakukan untuk mendapatkan data berkualitas tinggi (Harb, 2014) yaitu :

a. *Data Cleaning* .

Data Cleaning dilakukan untuk membersihkan inkonsistensi data yaitu data yang mengandung nilai-nilai yang hilang (*missing* value), berisik (*noise*) atau menghapus nilai-nilai yang salah (*outlier*). Untuk menyelesaikan data yang tidak konsisten, kita dapat memilih menggunakan metode berikut :

1. membuang *instance* pada *missing value.*
2. menggunakan nilai modus atau nilai rata-rata (misalnya, mean atau median) untuk mengisi *missing value*.
3. mengisi nilai yang hilang secara manual
4. menggunakan konstanta global untuk mengisi nilai yang hilang (mengganti semua hilang nilai atribut dengan konstan yang sama seperti label seperti "*Unknown*").

b. *Data integration*

Data integration tujuannya adalah untuk penggabungan beberapa data *warehouse*. Dalam rangka untuk membantu mengurangi dan menghindari redudansi dan inkonsistensi dalam kumpulan data yang dihasilkan, kita perlu integrasi sangat berhati-hati. Jadi, akurasi dan kecepatan proses data mining berikutnya akan ditingkatkan.

c. Data reduction

Data reduction dapat digunakan untuk menghilangkan data yang sama atau tidak lengkap, namun mempertahankan integritas data asli. Seleksi atribut dilakukan dimana akan dipilih atribut-atribut tertentu saja diantara semua atribut. Algoritma seleksi atribut akan menunjukkan atribut-atribut yang berkaitan dengan klasifikasi berdasarkan nilai pembobotan yang dihasilkan. Atribut yang berbobot 1 akan digunakan sebagai atribut prediksi untuk meningkatkan akurasi dan efisiensi model klasifikasi sedangkan atribut yang berbobot 0 akan dihilangkan sehingga meningkatkan akurasi dan efisiensi model yang diusulkan.

d. Data Transformation

Data Transformation mengubah menjadi data yang diperlukan dalam proses *data mining.*

3.4 Pendekatan *Data Mining*

Pada tahap ini pemilihan algoritma disesuaikan dengan permasalahan yang ada. Beberapa algoritma yang ada memiliki persyaratan yang spesifik pada bentuk data. Pada penelitian ini menggunakan algoritma *data mining* untuk klasifikasi yaitu *Naive Bayes Classifier* (NBC) dan *Parricle Swarm optimization – Naive Bayes Classifier* (PSO-NBC) akan diterapkan pada atribut-atribut dataset.

Fase utama dari penelitian ini adalah memperoleh algoritma yang akurat dalam klasifikasi pelanggan telepon dan speedy yang berpotensi untuk meningkatkan layanannya dengan produk indihome. Oleh karena itu pada tahap ini *data training* dilakukan klasifikasi dengan menggunakan algoritma *Naive Bayes Classifier* (NBC) dan algoritma *Naive Bayes Classifier* yang menggunakan seleksi atribut *Particle Swarm Optimization* (PSO-NBC).

Pada pendekatan *Naive Bayes Classifier* dilakukan perhitungan statistik untuk mengetahui kemungkinan (probabilitas) sebuah data masuk ke dalam klasifikasi kelas tertentu. Perhitungan probabilitas dengan menggunakan algoritma *Naive Bayes Classifier* dilakukan dengan menghitung probabilitas prior yang kemudian dilanjutkan dengan melakukan perhitungan probabilitas posterior.

Atribut - atribut pada *data training* digunakan sebagai masukan dalam pendekatan *data mining* ini untuk mengklasifikasikan pelanggan telepon dan speedy yang berpotensi untuk meningkatkan layanannya dengan indihome.

3.5 Eksperimen

Dalam eksperimen yang dilakukan menggunakan spesifikasi *hardware* dan *software* sebagai alat bantu dalam pemrosesan data dengan algoritma *Naive Bayes Classifier* dan *Particle Swarm Optimization*.

3.6 Pengukuran

Tahap ini dilakukan setelah model dibangun yang secara perspetif memiliki kualitas tinggi dari hasil analisis data. Sebelum melanjutkan model yang dibangun ke fase penerapan, fase yang penting untuk dilakukan adalah mengevaluasi model untuk memastikan bahwa model tersebut memiliki tingkat akurasi yang tinggi.

Dalam tahap ini akan dilakukan evaluasi dan validasi untuk menguji model penelitian dengan menggunakan *cross validation, confusion matriks, ROC curve.*

1. *10 - Fold Cross Validation*

   10-fold Cross Validation merupakan teknik statistik (Witten et al., 2011, p. 153). Data pelanggan eksisting akan dibagi data secara acak ke dalam 10 bagian dan masing-masing bagian tersebut akan dilakukan proses klasifikasi. Teknik pemilihan untuk data pembelajaran adalah dengan menggunakan *stratified sampling.*

2. *Confusion Matriks*

   Confusion Matrix merupakan alat untuk menganalisis seberapa baik *classifier* dapat mengenali tupel dari kelas yang berbeda. *Confusion matrix* berisi informasi aktual (actual) dan prediksi (predicted) kelas dari data pelanggan eksisting yang ada.

3. *ROC Curve*

   Curve ROC (Receiver Operating Characteristic) adalah salah satu cara untuk mengevaluasi akurasi dari model klasifikasi secara visual.

4. *T-test*

   T Test digunakan untuk membandingkan rata-rata dua variabel dalam satu group untuk pengujian terhadap dua sampel yang berhubungan atau dua sampel berpasangan (Elliott & Woodward, 2007).

3.7 Hasil

Setelah dilakukan pengukuran terhadap model dengan pendekatan *Naive Bayes Classifier* (NBC) dan *Particle Swarm optimization – Naive Bayes Classifier* (PSO-NBC) maka diperoleh hasil akurasi darikedua pendekatan untuk klasifikasi pelanggan telepon dan speedy eksisting. Pada tahap ini juga diketahui akurasi *data testing* dari model terbaik.

3.8 Penerapan

Pada tahap ini diterapkan sistem aplikasi berdasarkan model yang memiliki akurasi tertinggi pada pemasaran PT. Telkom Jakarta untuk mengklasifikasikan pelanggan telepon dan speedy eksisting *cross selling* produk Indihome. Tahap ni menggunakan perangkat lunak Microsoft Visual Basic 6.0.

BAB 4 HASIL ANALISA DAN PEMBAHASAN

4.1 Pengumpulan Data

Penelitian ini bertujuan mengetahui peningkatan akurasi algoritma *Naive Bayes Classifier* dengan menggunakan *Particle Swarm Optimization* sebagai seleksi atribut pada klasifikasi pelanggan telepon dan speedy eksisting PT. Telkom Jakarta *cross selling* produk Indihome dengan menguji keakuratan model sehingga dapat memberikan informasi yang bermanfaat khususnya dalam strategi pemasaran PT. Telkom.

Data yang diperoleh merupakan data primer dari divisi *Bussiness Planning and Performance* PT. Telkom Jakarta yang terdiri atas 878 data pelanggan telepon dan speedy eksisting dengan 14 atribut termasuk *class label atribute* (atribut output) yaitu atribut Paket. Data-data ini merupakan data fakultatif yang telah diketahui statusnya yaitu 3P (pelanggan Indihome) atau *decline* (menolak penawaran produk Indihome). Dari 878 data pelanggan telepon dan speedy eksisting terdapat 587 pelanggan Indihome (3P) dan 291 pelanggan telepon dan speedy yang menolak penawaran Indihome (*Decline*). 14 atribut yang terdapat pada data pelanggan telepon dan speedy eksisting tersebut adalah :

1. NCLI : Nomor identitas pelanggan
2. Nama : Nama pelanggan
3. Alamat : Alamat pelanggan
4. Regional : Regional 2 PT. TELKOM Jakarta meliputi Jabotabes (Jakarta, Bogor, Tangerang, Bekasi, Serang)
5. Witel : Wilayah telepon (Jakarta Utara, Jakarta Barat, Jakarta Selatan, 7 Jakarta Timur, Jakarta Pusat, Banten Barat (Serang), Banten Timur (Tangerang), Jabar Utara (Bekasi), Jabar Barat (Bogor))
6. POTS : *Plain Old Telephone Service* merupakan kode wilayah dan nomor telepon pelanggan (21xxxxxxxx, 251xxxxxxxx, 254xxxxxxxxx)

7. Speedy : Nomor identitas pelanggan speedy
8. Kwadran : Ada pemakaian maka ada tagihan (4)
9. Paket Eksisting : Jenis paket layanan yang saat ini digunakan (INETL15H, INETL50H, INETU1M, INETU2M, INETU384K, INETU384R, INETU3M, INETU512K, INETU512R, INET_1MB_H, INET_2MB_H, INET1000S, INETG10M1, INETG2M1, INETR10M1, INETR10M3, INETR1M1, INETR2M1, INETR2M3, INETR3M1, INETR3M2, INETR3M3, INETR50M1, INETR512K1, INETR512K2, INETR512K3, INETR5M1, INETR5M3)
10. R2BB : *Ready To Broadband* (GIPTV, GSP5, GTLP, IPTV, SP2, SP3, SP4, SP5, SP7)
11. Tipe Jaringan : Tipe Jaringan yang digunakan saat ini (*Copper*, FTTH, MSAN)
12. Zona : Zona waktu pemasaran (*attack_promo, super_winning, winning)*
13. Migrasi : Jenis pelanggan saat ini (2p = pelanggan telepon dan speedy)
14. Paket : Apakah pelanggan akan meningkatkan layanan menjadi pelanggan indihome (decline, 3P)

Atribut – atribut tersebut merupakan komponen yang mempengaruhi pelanggan telepon dan speedy untuk *cross selling* produk indihome. Keterangan dan informasi yang menjelaskan masing-masing atribut yang ada pada dataset, diperoleh dari sumber primer divisi *Bussiness Planning and Performance* PT. Telkom Jakarta.

4.2 Preprocessing Data

Data penelitian merupakan data yang telah diketahui statusnya yaitu 3P (pelanggan Indihome) atau *decline* (menolak penawaran produk Indihome). Dari 878 data pelanggan telepon dan speedy eksisting terdapat 587 pelanggan Indihome (3P) dan 291 pelanggan telepon dan speedy yang menolak penawaran Indihome (*Decline*). Data ini memiliki 14 atribut termasuk *class label atribute* (atribut output) yaitu atribut Paket. Akan tetapi 5 atribut NCLI, Nama, Alamat, POTS, Speedy mempunyai nilai yang tidak sama (unik) untuk setiap pelanggan sehingga pada penelitian ini hanya menggunakan 9 atribut yaitu :

1. Regional : Regional 2 PT. TELKOM Jakarta meliputi Jabotabes (Jakarta, Bogor, Tangerang, Bekasi, Serang)
2. Witel : Wilayah telepon (Jakarta Utara, Jakarta Barat, Jakarta Selatan, 7 Jakarta Timur, Jakarta Pusat, Banten Barat (Serang), Banten Timur (Tangerang), Jabar Utara (Bekasi), Jabar Barat (Bogor))
3. Kwadran : Ada pemakaian maka ada tagihan (4)
4. Paket Eksisting : Jenis paket layanan yang saat ini digunakan (INETL15H, INETL50H, INETU1M, INETU2M, INETU384K, INETU384R, INETU3M, INETU512K, INETU512R, INET_1MB_H, INET_2MB_H, INET1000S, INETG10M1, INETG2M1, INETR10M1, INETR10M3, INETR1M1, INETR2M1, INETR2M3, INETR3M1, INETR3M2, INETR3M3, INETR50M1, INETR512K1, INETR512K2, INETR512K3, INETR5M1, INETR5M3)
5. R2BB : *Ready To Broadband* (GIPTV, GSP5, GTLP, IPTV, SP2, SP3, SP4, SP5, SP7)

6. Tipe Jaringan : Tipe Jaringan yang digunakan saat ini (*Copper*, FTTH, MSAN)
7. Zona : Zona waktu pemasaran (*attack_promo, super_winning, winning)*
8. Migrasi : Jenis pelanggan saat ini (2p = pelanggan telepon dan speedy)
9. Paket : Apakah pelanggan akan meningkatkan layanan menjadi pelanggan indihome (decline, 3P)

Sehingga dataset pelanggan telepon dan speedy eksisting yang digunakan dari PT. TELKOM Jakarta dalam format excel seperti pada tabel berikut :

Tabel 4.1 Data pelanggan telepon dan speedy yang digunakan

REGIONAL	WITEL	KWADRAN	PAKET EXISTING	R2BB	TYPE JARINGAN	ZONA	MIGRASI	PAKET
2	BANTEN BARAT (SERANG)	4	INETR1M3	GTLP	FTTH	SUPER_WINNING	2P	3P
2	BANTEN BARAT (SERANG)	4	INETR1M3	GTLP	FTTH	SUPER_WINNING	2P	3P
2	JAKARTA TIMUR	4	INETR1M1	GIPTV	FTTH	SUPER_WINNING	2P	3P
2	JAKARTA TIMUR	4	INETR1M1	GIPTV	FTTH	SUPER_WINNING	2P	3P
2	JABAR BARAT UTARA (BEKASI)	4	INETR1M1	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE
2	JABAR BARAT UTARA (BEKASI)	4	INETR1M1	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE
2	JAKARTA BARAT	4	INETR3M1	GIPTV	FTTH	SUPER_WINNING	2P	3P

REGIONAL	WITEL	KWADRAN	PAKET EXISTING	R2BB	TYPE JARINGAN	ZONA	MIGRASI	PAKET
2	JAKARTA TIMUR	4	INETR1M3	SP4	COPPER	SUPER_WINNING	2P	3P
2	JAKARTA TIMUR	4	INETR1M3	SP4	COPPER	SUPER_WINNING	2P	3P
2	JABAR BARAT UTARA (BEKASI)	4	INETR1M3	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE
2	JABAR BARAT UTARA (BEKASI)	4	INETR1M3	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE
2	JABAR BARAT UTARA (BEKASI)	4	INETR1M1	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE
2	JABAR BARAT UTARA (BEKASI)	4	INETR1M1	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE

Data lengkap pelanggan eksisting telepon dan speedy terdapat pada lampiran 4. Tabel tarif berdasarkan Paket Eksisting dan Zona Waktu Pemasaran Pelanggan terdapat pada lampiran 1.

Selanjutnya data pelanggan telepon dan speedy eksisting PT. TELKOM Jakarta yang diperoleh selanjutnya ditinjau kelengkapan datanya. Dari data yang diperoleh diketahui bahwa terdapat *missing value* pada atribut kwadran dan paket eksisting pada dataset pelanggan telepon dan speedy eksisting ke 830, 838 dan 842.

Tabel 4.2 Missing Value data

NO	PAKET	REGIONAL	WITEL	KWADRAN	PAKET EXISTING	R2BB	TYPE JARINGAN	ZONA	MIGRASI
830	3P	2	JAKARTA UTARA	?	GIPTV	SP3	FTTH	SUPER_WINNING	2P
838	3P	2	JAKARTA BARAT	?	INETR2M3	SP4	COPPER	SUPER_WINNING	2P
842	3P	2	JAKARTA BARAT	?	?	SP4	COPPER	SUPER_WINNING	2P

Oleh karena itu dilakukan *replace missing value* pada data dengan *missing value* dengan nilai rata-rata. Hasil Replace Missing Value ditunjukkan pada tabel berikut ini :

Tabel 4.3 Hasil *Replace Missing Value*

NO	PAKET	REGIONAL	WITEL	KWADRAN	PAKET EXISTING	R2BB	TYPE JARINGAN	ZONA	MIGRASI
830	3P	2	JAKARTA UTARA	4	GIPTV	SP3	FTTH	SUPER_WINNING	2P
838	3P	2	JAKARTA BARAT	4	INETR2M3	SP4	COPPER	SUPER_WINNING	2P
842	3P	2	JAKARTA BARAT	4	INETR1M1	SP4	COPPER	SUPER_WINNING	2P

Tabel diatas menunjukkan hasil *Replace Missing Value* dimana data ke 830, 838 dan 842 yang pada awalnya terdapat *missing value* telah terisi dengan nilai rata-rata yang ada dimana nilai atribut untuk :

- Data 830
  kwadran : 4
- Data 838
  kwadran : 4
- Data 842
  kwadran : 4
  paket eksisting : INETR1M1

4.3 Pendekatan *Data Mining*

4.3.1 Algoritma *Naive Bayes Classifier*

Data mining dengan menggunakan algoritma *Naive Bayes Classifier* dilakukan dengan menghitung probabilitas prior yang kemudian dilanjutkan dengan melakukan perhitungan probabilitas posterior.

Algoritma Naive Bayes terlebih dahulu kita menghitung nilai probabilitas hipotesis untuk masing-masing kelas P(H). Hipotesis pelanggan telepon dan speedy yang menyetujui untuk meningkatkan layanannya dengan produk indihome adalah

3P sedangkan pelanggan yang menolak penawaran produk indihome adalah *Decline*.

Data-data yang digunakan telah diketahui statusnya yaitu 3P (pelanggan Indihome) atau *decline* (menolak penawaran produk Indihome). Dari 878 data pelanggan telepon dan speedy eksisting terdapat 587 pelanggan Indihome (3P) dan 291 pelanggan telepon dan speedy yang menolak penawaran Indihome (*Decline*) sehingga dengan menggunakan persamaan 2.1, maka :

P (3P) = 587 : 878 = 0,66856492

P (Decline) = 291 : 878 = 0,33143508

Perhitungan probabilitas prior selengkapnya terdapat pada lampiran 2 digunakan untuk menentukan kelas pada kasus baru yang terlebih dahulu dihitung probabilitas posteriornya. Jika terdapat contoh kasus seperti dibawah ini :

Tabel 4.4 Probabilitas Posterior NBC

ATRIBUT	NILAI	P(H|X)	
		3P	DECLINE
REGIONAL	2	1	1
WITEL	Jakarta Timur	0,233390119	0,068728522
KWADRAN	4	1	1
PAKET EKSISTING	INETR1M1	0,340715503	0,340206186
TIPE JARINGAN	FTTH	0,187393526	0,41580756
ZONA	*Super_Winning*	0,839863714	0,865979381
MIGRASI	2P	1	1
R2BB	GIPTV	0,182282794	0,41580756

Setelah diketahui probabilitas setiap atribut terhadap probabilitas tiap kelas atau P(H|X), maka langkah selanjutnya adalah menghitung total keseluruhan probabilitas tiap kelas yaitu :

P(H|paket = 3P) = P(Regional = 2 | Paket = 3P) x P(Witel = Jakarta Timur | Paket = 3P) x P(Kwadran = 4 | Paket = 3P) x P(Paket Eksisting = INETR1M1 | Paket = 3P) x P(Tipe Jaringan = FTTH | Paket = 3P) x P(Zona = *Super_Winning* | Paket = 3P) x P(Migrasi = 2P | Paket = 3P) x P(R2BB = GIPTV | Paket = 3P)

P(H|paket = 3P) = 1 x 0,233390119 x 1 x 0,340715503 x 0,187393526 x 0,839863714 x 1 x 0,182282794

= **0,002281305**

P(H|paket = Decline) = P(Regional = 2 | Paket = *Decline*) x P(Witel = Jakarta Timur | Paket = *Decline*) x P(Kwadran = 4 | Paket = *Decline*) x P(Paket Eksisting = INETR1M1 | Paket = *Decline*) x P(Tipe Jaringan = FTTH | Paket = *Decline*) x P(Zona = *Super_Winning* | Paket = *Decline*) x P(Migrasi = 2P | Paket = *Decline*) x P(R2BB = GIPTV | Paket = *Decline*)

= 1 x 0,068728522 x 1 x 0,340206186 x 0,41580756 x 0,865979381 x 1 x 0,41580756

= **0,003500834**

dimana berdasarkan persamaan 2.2 yaitu

P (H| paket = 3P) P (3P) = 0,002281305 x 0,66856492 = **<u>0,001525201</u>**

P(H| paket = Decline) P(Decline) = 0,003500834 x 0,33143508= **<u>0,001160299</u>**

sehingga

P (H| paket = 3P) P (3P) > P(H| paket = Decline) P(Decline)

<u>0,001525201</u> > **<u>0,001160299</u>**

Dari hasil perhitungan tesebut diketahui nilai P (H|3P) lebih besar dari pada nilai P (H|*Decline*), sehingga dapat disimpulkan bahwa untuk kasus tersebut masuk kedalam klasifikasi **<u>3P</u>**.

4.3.2 Algoritma *Particle Swarm Optimization - Naive Bayes Classifier*

Dari hasil seleksi atribut diperoleh bahwa atribut Witel dan Tipe Jaringan bernilai lebih dari 0. Hal ini menunjukkan bahwa atribut-atribut yang bernilai lebih

dari 0 memiliki keterkaitan terhadap klasifikasi suatu data sehingga digunakan sebagai atribut prediksi dalam klasifikasi pelanggan telepon dan speedy *cross seling* produk Indihome. Dengan menggunakan persamaan 2.1, maka :

P (3P) = 587 : 878 = 0,66856492
P (Decline) = 291 : 878 = 0,33143508

Perhitungan probabilitas prior selengkapnya terdapat pada lampiran 3 digunakan untuk menentukan kelas pada kasus baru yang terlebih dahulu dihitung probabilitas posteriornya. Jika terdapat contoh kasus seperti dibawah ini :

Tabel 4.5 Probabilitas Posterior PSO-NBC

ATRIBUT	NILAI	P(H|X)	
		3P	DECLINE
WITEL	Jakarta Timur	0,233390119	0,068728522
TIPE JARINGAN	FTTH	0,187393526	0,41580756

Setelah diketahui probabilitas setiap atribut terhadap probabilitas tiap kelas atau P(H|X), maka langkah selanjutnya adalah menghitung total keseluruhan probabilitas tiap kelas yaitu :

P(H|paket = 3P) = P(Witel = Jakarta Timur | Paket = 3P) x P(Tipe Jaringan = FTTH | Paket = 3P)
P(H|paket = 3P) = 0,233390119 x 0,187393526 = **0,043735797**

P (H|paket = Decline) = P(Witel = Jakarta Timur | Paket = *Decline*) x P(Tipe Jaringan = FTTH | Paket = *Decline*)
= 0,068728522 x 0,41580756 = **0,01287928**

dimana berdasarkan persamaan 2.2 yaitu
P (H| paket = 3P) P (3P) = **0,043735797** x 0,66856492 = **0,02924022**
P(H| paket = Decline) P(Decline) = **0,01287928** x 0,33143508 = **0,004268645**

sehingga

P (H| paket = 3P) P (3P) > P(H| paket = Decline) P(Decline)

0,02924022 > 0,004268645

Dari hasil perhitungan tesebut diketahui nilai P (H|3P) lebih besar dari pada nilai P (H|*Decline*), sehingga dapat disimpulkan bahwa untuk kasus tersebut masuk kedalam klasifikasi **3P**.

4.4 Eksperimen

Eksperimen yang dilakukan mengunakan alat bantu untuk pemrosesan data dengan spesifikasi *hardware* dan *software* sebagai berikut :

Tabel 4.6 Spesifikasi *hardware* dan *software*

Hardware	***Software***
CPU : Intel Core i3 (1,8GHz)	Sistem Operasi : Windows 7
Memory : 2 GB	Data Mining : Rapidminer versi 5.3
Harddisk : 500 GB	

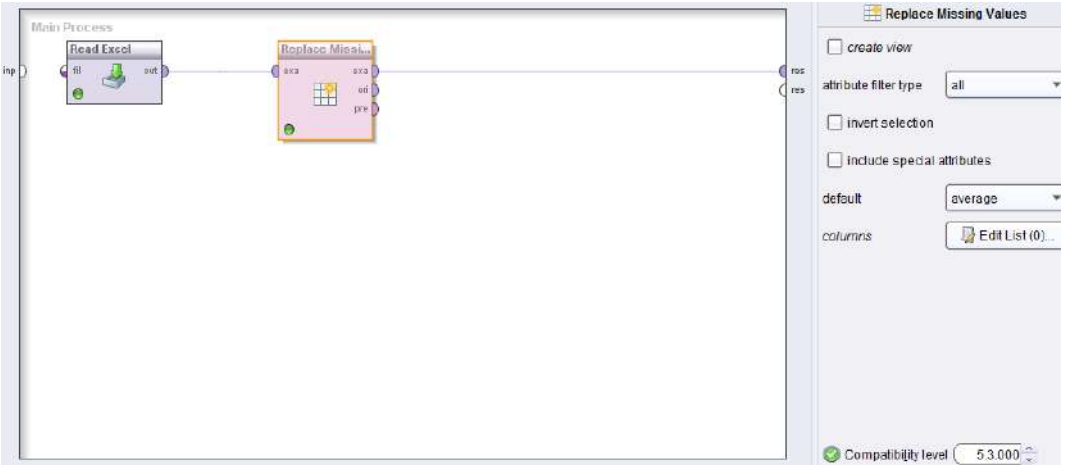
Penggunaan software Rapidminer 5.3 dilakukan dengan langkah-langkah berikut ini :

a. Perangkat lunak Rapidminer 5.3 dapat membaca data dengan format .csv, .xls, .mdb dan sebagainya. Data yang digunakan dalam penelitian ini adalah format .xls
b. Buka perangkat lunak Rapidminer 5.3.
c. Pilih *new process* untuk membuat proses baru.
d. Gunakan operator –operator yang diperlukan untuk melakukan eksperimen

Untuk melihat kelengkapan data agar mengetahui *missing value* maka dilakukan proses sebagai berikut :

Gambar 4.1 Cara melihat *missing value*

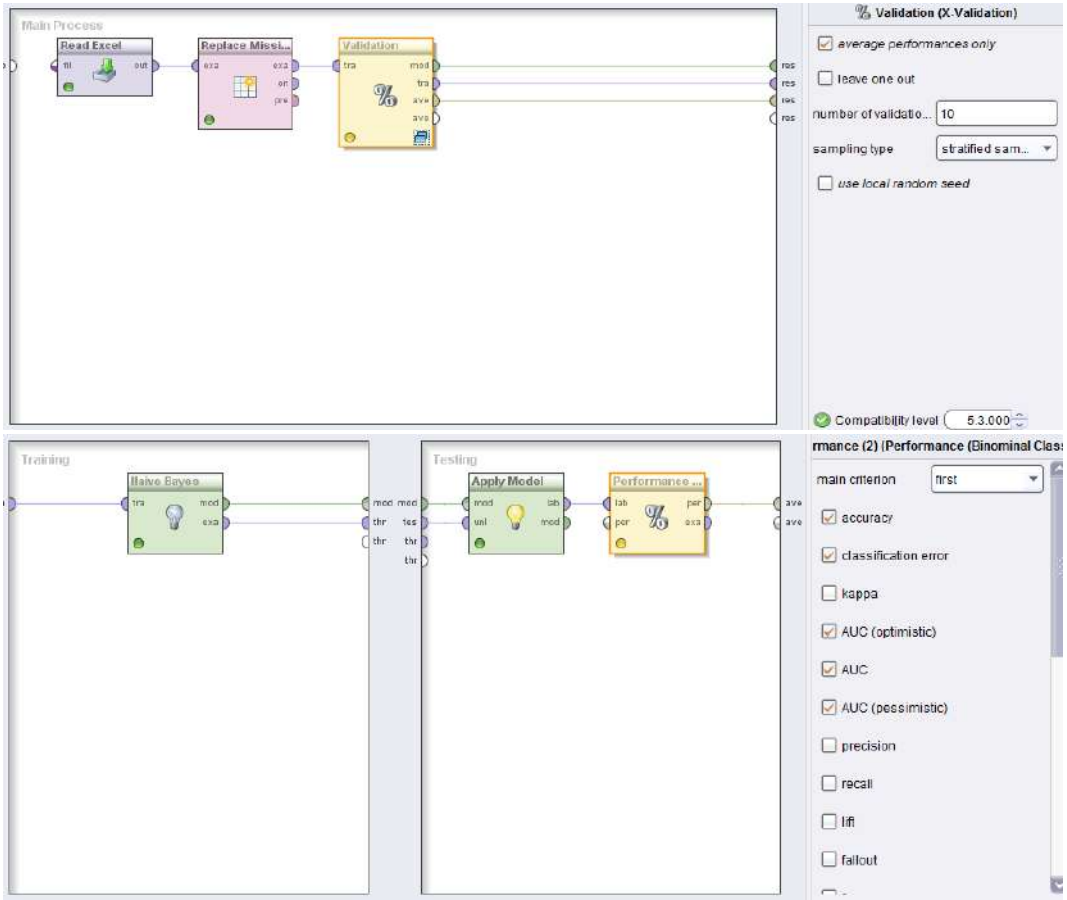
Sedangkan untuk mengisi *missing value* maka digunakan operator *Replace Missing Value* adalah *average* yang mempunyai arti data dengan *missing value* akan diisi dengan nilai rata-rata. Berikut ini merupakan desain pada Rapidminer 5.3 :



Gambar 4.2 *Replace missing value*

4.4.1 Algoritma *Naive Bayes Classifier*

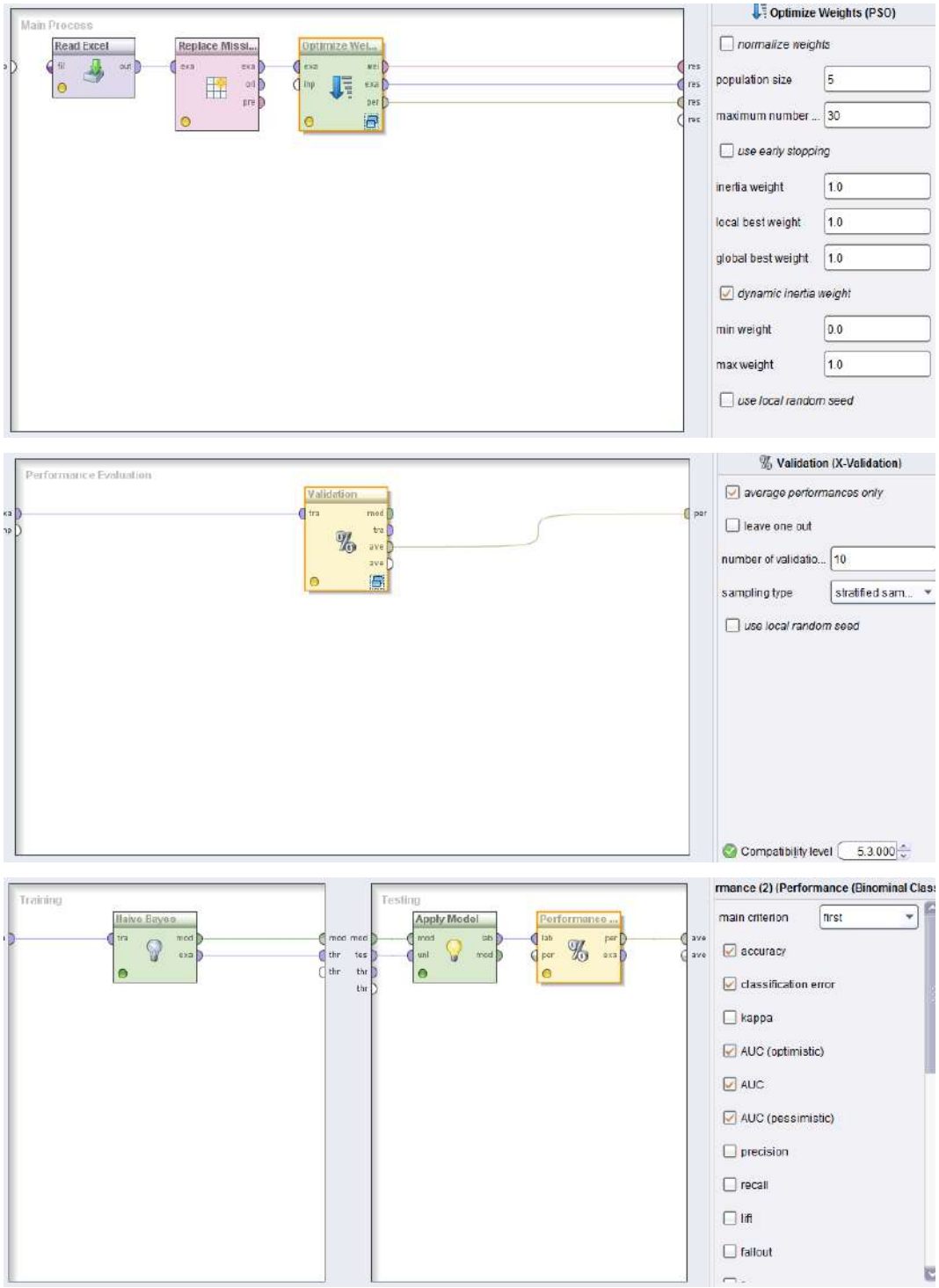
Eksperimen yang dilakukan dengan pendekatan data mining algoritma *Naive Bayes Classifier* menggunakan perangkat lunak Rapidminer 5.3 yaitu sebagai berikut :



Gambar 4.3 Model *Naive Bayes Classifier*

4.4.2 Algoritma *Particle Swarm Optimization-Naive Bayes Classifier*

Eksperimen yang dilakukan dengan pendekatan *Particle Swarm optimization* sebagai seleksi atribut pada algoritma *Naive Bayes Classifier* menggunakan perangkat lunak Rapidminer 5.3 yaitu sebagai berikut :



Gambar 4.4 Model *Particle Swarm Optimization - Naive Bayes Classifier*

Gambar diatas menunjukkan model *Particle Swarm Optimization - Naive Bayes Classifier* dimana parameter yang digunakan untuk *Particle Swarm Optimization* yaitu :

- *Population size* : 5

  Population size merupakan parameter yang menunjukkan jumlah individu tiap generasi
- *Maximum number of generation* : 30

  Maximum number of generation merupakan parameter yang menunjukkan jumlah generasi maksimum untuk menghentikan jalannya algoritma
- *Inertia weight* : 1.0

  Inertia weight merupakan parameter yang menentukan kecepatan pergerakan terhadap diri sendiri
- *Local best weight* : 1.0

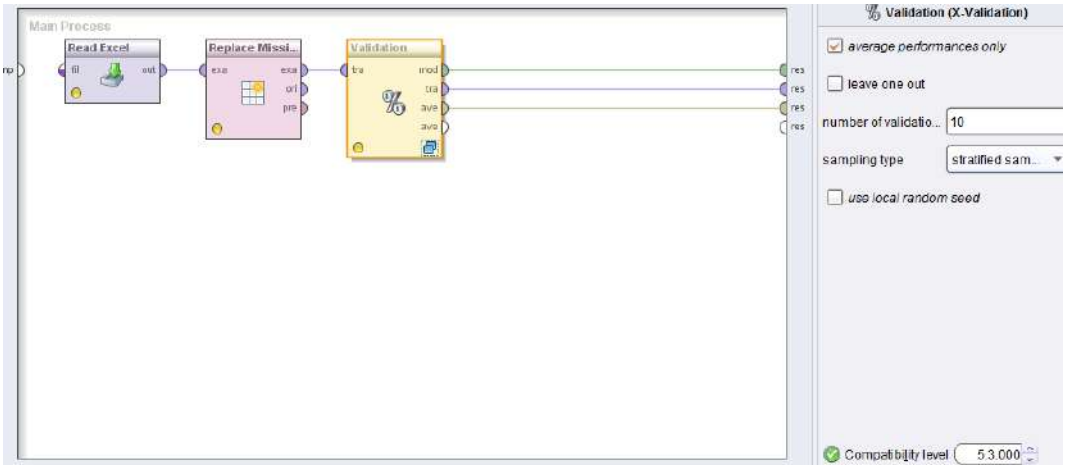
  Local best weight merupakan parameter yang menunjukkan posisi terbaik dari setiap populasi
- *Global best weight* : 1.0

  Global best weight merupakan parameter yang menunjukkan posisi terbaik dari keseluruhan solusi

4.5 Pengukuran

4.5.1 Algoritma *Naive Bayes Classifier*

Berikut ini merupakan model untuk pengujian performansi algoritma *Naive Bayes Classifier* pada Rapidminer 5.3 :



Gambar 4.5 Validasi model *Naive Bayes Classifier*

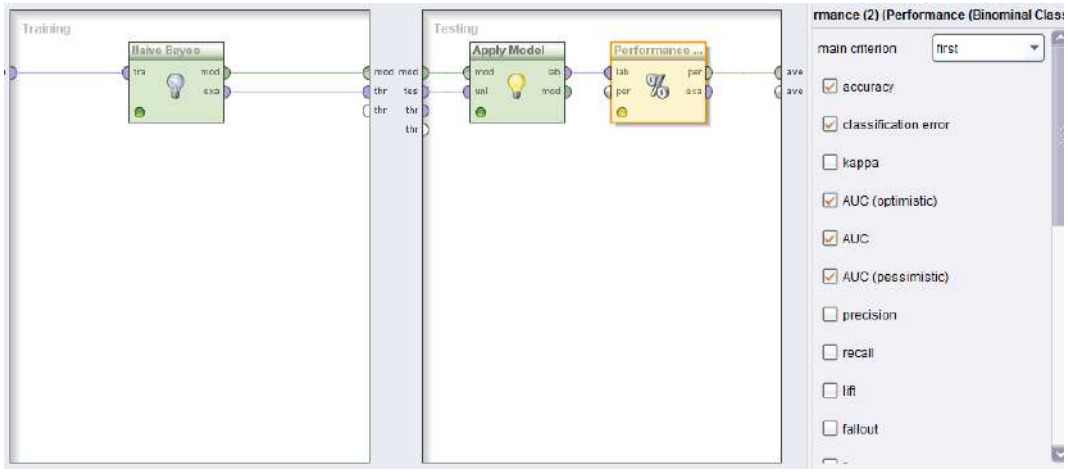
Gambar diatas menunjukkan model *Naive Bayes Classifier* dimana parameter yang digunakan untuk 10-*fold cross validation* yaitu :

- *average performance only* : aktif

  Average performance only merupakan parameter yang menunjukkan bahwa nilai kinerja adalah hasil rata-rata
- *number of validation* : 10

  Number of validation merupakan parameter yang menentukan jumlah subset yang terbagi sama dan jumlah yang sama dari iterasi yang akan berlangsung. Setiap iterasi akan melibatkan data *training* dan data *testing.*
- *sampling type* : *stratified sampling.*

  Stratified sampling adalah jenis sample yang dibangun secara acak dan distribusi kelas dalam himpunan sama.

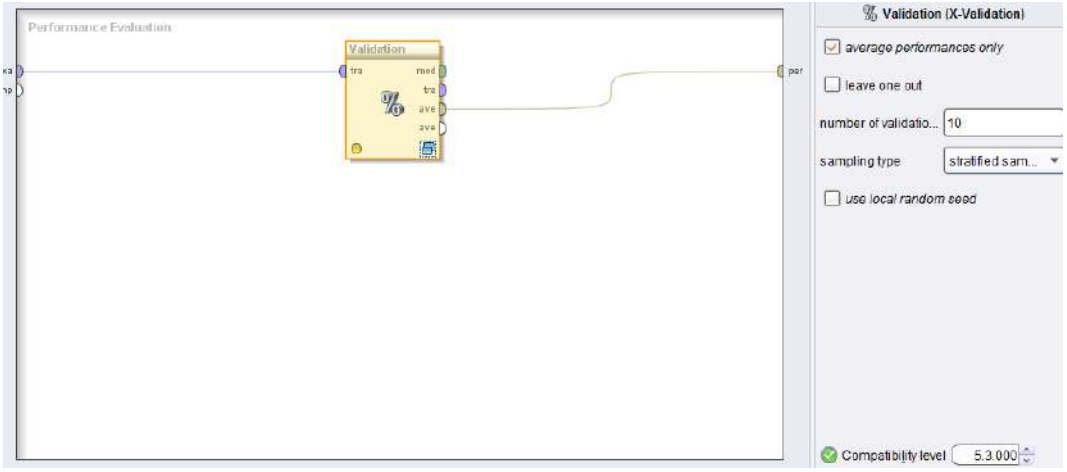


Gambar 4.6 Pengukuran model *Naive Bayes Classifier*

Parameter yang digunakan pada operator *performance* adalah *accuracy, classification error, Area Under Curve* (*AUC*) untuk menampilkan tingkat keakurasian model *Naive Bayes* ini.

4.5.2 Algoritma *Particle Swarm Optimization - Naive Bayes Classifier*

Berikut ini merupakan model untuk pengujian algoritma *Particle Swarm Optimization - Naive Bayes Classifier* pada Rapidminer 5.3 dengan 10-*fold cross validation.*



Gambar 4.7 Validasi model *Particle Swarm Optimization - Naive Bayes Classifier*

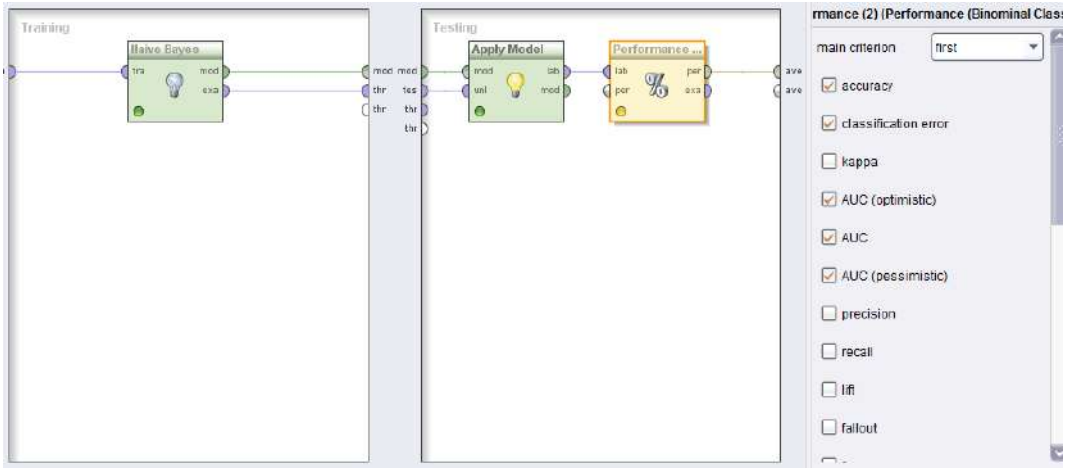
Gambar diatas menunjukkan pendekatan *Particle Swarm Optimization* pada *Naive Bayes Classifier* dimana parameter yang digunakan untuk 10-*fold cross validation* yaitu :

- *average performance only* : aktif

  Average performance only merupakan parameter yang menunjukkan bahwa nilai kinerja adalah hasil rata-rata
- *number of validation* : 10

  Number of validation merupakan parameter yang menentukan jumlah subset yang terbagi sama dan jumlah yang sama dari iterasi yang akan berlangsung. Setiap iterasi akan melibatkan data *training* dan data *testing.*
- *sampling type* : *stratified sampling.*

  Stratified sampling adalah jenis sample yang dibangun secara acak dan distribusi kelas dalam himpunan sama.



Gambar 4.8 Pengukuran model *Particle Swarm Optimization - Naive Bayes Classifier*

Parameter yang digunakan pada operator *performance* adalah *accuracy*, *classification error*, *Area Under Curve* (*AUC*) untuk menampilkan tingkat keakurasian model *Naive Bayes* ini.

4.6 Hasil Analisa

4.6.1 Algoritma *Naive Bayes Classifier*

Hasil pendekatan model *Naive Bayes Clasifier* pada penelitian ini adalah sebagai berikut :

Tabel 4.7 Hasil performansi pendekatan *Naive Bayes Classifier*

Akurasi	Classification Error	AUC
85,08%	14,92%	0,841

4.6.2 Algoritma *Particle Swarm Optimization - Naive Bayes Classifier*

Hasil seleksi atribut dengan menggunakan *Naive Bayes Classifier* berbasis *Particle Swarm Optimization* adalah sebagai berikut :

Tabel 4.8 Hasil seleksi atribut *Particle Swarm Optimization*

ATRIBUT	WEIGHT
REGIONAL	0
WITEL	1
KWADRAN	0
PAKET EXISTING	0
R2BB	0
TYPE JARINGAN	0.9261941575329261
ZONA	0
MIGRASI	0

Tabel diatas merupakan hasil dari seleksi atribut menggunakan *Particle Swarm Optimization* dimana dari hasil menunjukkan bahwa atribut Regional, Kwadran, Paket Eksisting, R2BB, Zona, Migrasi bernilai 0. Sedangkan atribut Witel dan Tipe Jaringan bernilai 1 dan 0,926. Hal ini menunjukkan bahwa atribut-atribut yang bernilai lebih dari 0 terkait terhadap klasifikasi suatu data sehingga digunakan sebagai atribut prediksi klasifikasi pelanggan telepon dan speedy *cross seling* produk Indihome dan sebaliknya atribut-atribut yang bernilai 0 tidak memiliki kaitan terhadap klasifikasi suatu data dan dapat dihilangkan untuk mempercepat kinerja model. Hal ini memiliki arti bahwa dataset yang semula terdapat 8 atribut prediktor menjadi 2 atribut prediktor.

Hasil pendekatan *Naive Bayes Clasifier* dengan seleksi atribut *Particle Swarm Optimization* pada model adalah sebagai berikut :

Tabel 4.9 Hasil performansi pendekatan *Particle Swarm Optimization-Naive Bayes Clasifier*

Akurasi	Classification Error	AUC
89,31%	10,69%	0,845

4.7 Pembahasan

4.7.1 Peningkatan akurasi

Berdasarkan analisa hasil evaluasi data mining untuk klasifikasi *cross selling* dengan algoritma *Naive Bayes* dan *Particle Swarm Optimization – Naive Bayes* maka dapat dirangkum sebagai berikut :

Tabel 4.10 Nilai perbandingan performansi

	NB	PSO - NB
Accuracy	85,08 %	89,31%
Classification Error	14,92 %	10,69%
AUC	0.841	0,845

Tabel diatas menunjukan bahwa dengan pendekatan *Naive Bayes Classifier* menghasilkan nilai akurasi sebesar 85,08% sedangkan dengan pendekatan *Particle Swarm Optimization – Naive Bayes Classifier* terdapat peningkatan nilai akurasi menjadi 89,31%.

Hasil performansi terbaik diperoleh dengan pendekatan *Particle Swarm Optimization – Naive Bayes Classifier*. Hasil performansi dan ROC Curve adalah sebagai berikut :

PerformanceVector

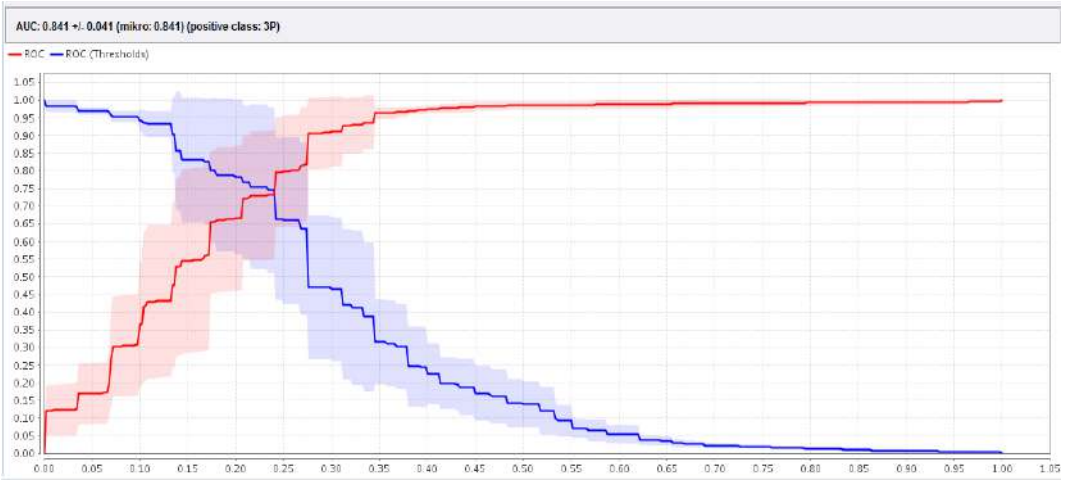
```
PerformanceVector:
accuracy: 85.08% +/- 3.47% (mikro: 85.08%)
ConfusionMatrix:
True:   DECLINE 3P
DECLINE:        213     53
3P:     78      534
classification_error: 14.92% +/- 3.47% (mikro: 14.92%)
ConfusionMatrix:
True:   DECLINE 3P
DECLINE:        213     53
3P:     78      534
AUC (optimistic): 0.842 +/- 0.041 (mikro: 0.842) (positive class: 3P)
AUC: 0.841 +/- 0.041 (mikro: 0.841) (positive class: 3P)
AUC (pessimistic): 0.840 +/- 0.041 (mikro: 0.840) (positive class: 3P)
```

Gambar 4.9 Performansi *Naive Bayes Classifier*

Confusion Matrix

Nilai akurasi dengan pendekatan *Naive Bayes Classifier* adalah = 85,08% dapat dihitung dari hasil *Confusion Matrix* dimana *true positif* (tp) bernilai 534, *false negatif* (fn) bernilai 78, *false positif* (fp) bernilai 53 dan *true negatif* (tn) bernilai 213 sehingga berdasarkan persamaan 2.8 maka :

$$\text{Akurasi} = \frac{534 + 213}{534 + 53 + 213 + 78} = 0{,}8508$$



Gambar 4.10 *ROC Curve Naive Bayes Classifier*

ROC Curve

Hasil kurva ROC menujukkan arah garis melintang dari kiri bawah ke kanan atas didalam grafik dengan nilai AUC 0.841 sehingga dapat disimpulkan bahwa model *Naive Bayes* ini merupakan model klasifikasi yang Baik bahwa akurasi 0.80 – 0.90 = *Good classification.*

Sedangkan untuk pendekatan *Particle Swarm Optimization* - *Naive Bayes Classifier* menunjukan nilai akurasi dan AUC sebagai berikut :

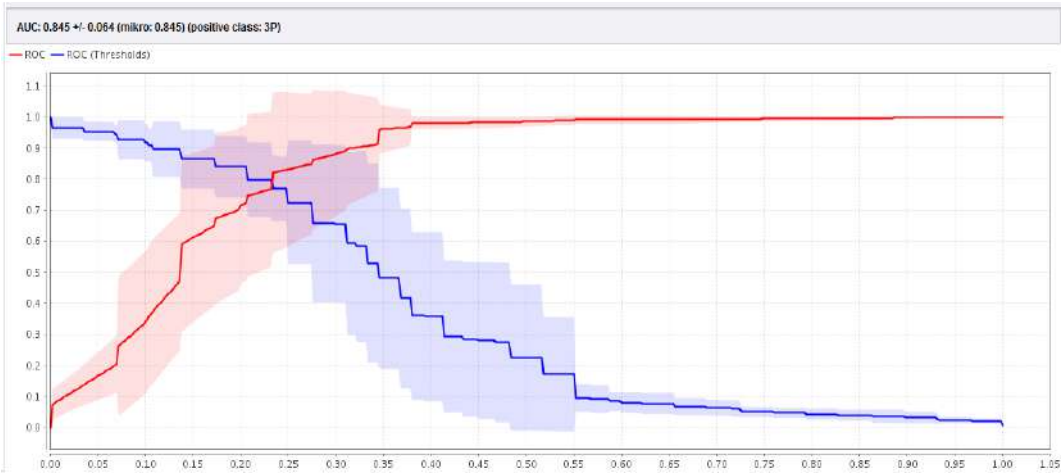
PerformanceVector

```
PerformanceVector:
accuracy: 89.31% +/- 3.51% (mikro: 89.29%)
ConfusionMatrix:
True:   DECLINE 3P
DECLINE:        216     19
3P:     75      568
classification_error: 10.69% +/- 3.51% (mikro: 10.71%)
ConfusionMatrix:
True:   DECLINE 3P
DECLINE:        216     19
3P:     75      568
AUC (optimistic): 0.865 +/- 0.059 (mikro: 0.865) (positive class: 3P)
AUC: 0.845 +/- 0.064 (mikro: 0.845) (positive class: 3P)
AUC (pessimistic): 0.825 +/- 0.070 (mikro: 0.825) (positive class: 3P)
```

Gambar 4.11 Performansi *Particle Swarm Optimization - Naive Bayes Classifier Confusion Matrix*

Nilai akurasi dengan pendekatan *Particle Swarm Optimization - Naive Bayes Classifier* adalah = 89,31% dapat dihitung dari hasil *Confusion Matrix* dimana *true positif* (tp) bernilai 568, *false negatif* (fn) bernilai 75, *false positif* (fp) bernilai 19 dan *true negatif* (tn) bernilai 216 sehingga berdasarkan persamaan 2.8 maka :

$$\text{Akurasi} = \frac{568 + 216}{568 + 216 + 75 + 19} = 0{,}8929$$



Gambar 4.12 *ROC Curve Particle Swarm Optimization - Naive Bayes Classifier*

ROC Curve

Sedangkan hasil kurva ROC menujukkan arah garis melintang dari kiri bawah ke kanan atas didalam grafik dengan nilai AUC 0.845 sehingga dapat disimpulkan bahwa model *Particle Swarm Optimization - Naive Bayes Classifier* ini merupakan model klasifikasi yang Baik bahwa akurasi 0.80 – 0.90 = *Good classification*.

4.7.2 Uji *T-test*

Prosedur *T-test* dilakukan untuk melihat perbedaan antara hasil algoritma *Naive Bayes Classifier* dan *Particle Swarm Optimization - Naive Bayes Classifier*. Berikut ini merupakan hasil uji statistik *T-test* :

T-Test Significance

	0.893 +/- 0.036	0.851 +/- 0.035
0.893 +/- 0.036		**0.017**
0.851 +/- 0.035		

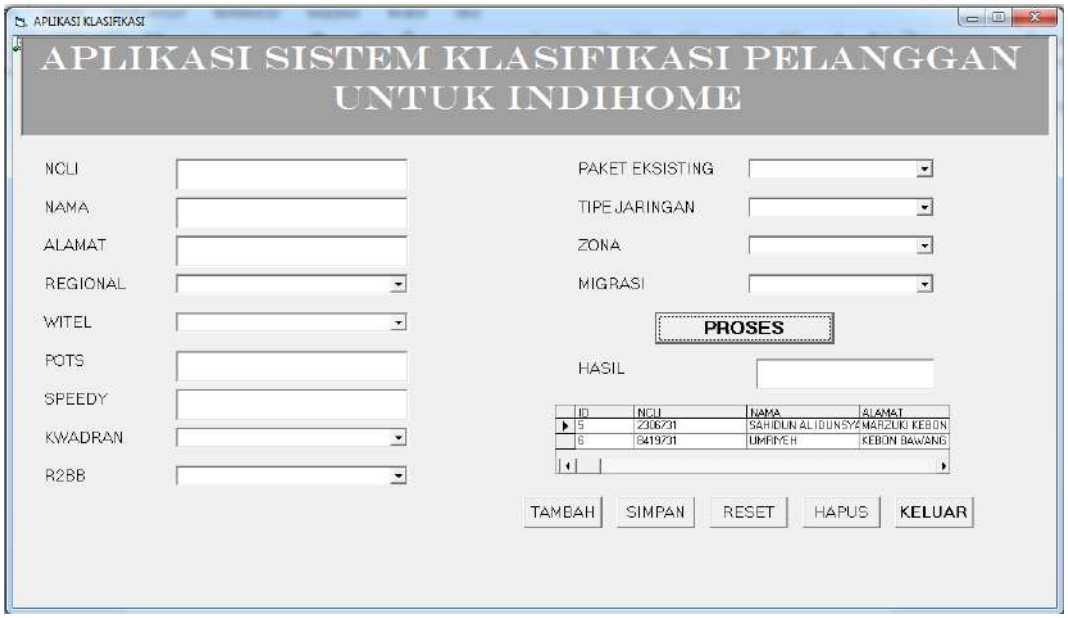
Probabilities for random values with the same result.
Bold values are smaller than alpha=0.050 which indicates a probably significant difference between the actual mean values!

Gambar 4.13 Uji T-test

Dari hasil pengujian dengan T-test diperoleh nilai $\alpha < 0.05$ yaitu 0,017 yang menunjukkan bahwa terdapat perubahan yang signifikan terhadap dua algoritma klasifikasi sehingga algoritma *Particle Swarm Optimization - Naive Bayes Classifier* merupakan algoritma yang signifikan untuk *data mining* klasifikasi ini.

4.8 Penerapan

Berdasarkan hasil penelitian yang telah dilakukan untuk data mining klasifikasi pelanggan telepon dan speedy eksisting PT Telkom Jakarta dalam *cross selling* produk Indihome maka diperoleh *rule* terbaik untuk diterapkan dalam pembuatan interface dengan algoritma *Naive Bayes Classifier* yang dapat memudahkan pemasaran PT. Telkom Jakarta. Interface yang digunakan dalam penelitian ini yaitu dengan menggunakan Microsoft Visual Basic 6.0.



Gambar 4.14 GUI Sistem Klasifikasi *Cross Selling* Produk Indihome

Gambar 4.15 Hasil klasifikasi GUI

4.9 Implikasi Penelitian

Berdasarkan tinjauan pustaka yang telah dilakukan menyatakan bahwa salah satu algoritma *data mining* untuk klasifikasi adalah *Naive Bayes Classifier.* Dalam tinjauan studi terdapat penelitian - penelitian *cross selling* menggunakan algoritma *Naive Bayes Classifier*. *Naive Bayes Classifier* merupakan classifier yang sederhana namun asumsi atribut *independent* Naive Bayes sering dilanggar dan sebagai hasilnya perkiraan probabilitas sering kurang optimal sehingga tingkat akurasi yang dihasilkan pun menjadi kurang optimal. Oleh sebab itu, untuk memperoleh tingkat akurasi yang optimal maka diperlukan seleksi atribut.

Pada penelitian klasifikasi *cross selling* produk Indihome ini menggunakan algoritma *Naive Bayes* dan *Particle Swarm Optimization* sebagai seleksi atribut. Penelitian ini dapat diterapkan pada strategi pemasaran perusahaan-perusahaan yang hendak memasarkan produk barunya kepada pelanggan yang telah membeli produk lain sebelumnya.

Berdasarkan pengujian dari dataset penelitian ini terbukti bahwa algoritma *Naive Bayes* dan *Particle Swarm Optimization* sebagai seleksi atribut akan lebih akurat bila dibandingkan dengan algoritma *Naïve Bayes* tanpa seleksi atribut. Hal ini dapat terlihat dari hasil penelitian yang telah dilakukan. Dengan hasil penelitian ini, menunjukkan bahwa *Naive Bayes* dan *Particle Swarm Optimization* sebagai seleksi atribut merupakan metode yang baik dalam klasifikasi pelanggan sehingga dapat memberikan pemecahan untuk permasalahan klasifikasi pada *cross selling*.

Untuk mendukung pengambilan keputusan dan pengembangan sistem pengetahuan berbasis komputer, hasil penelitian ini memiliki implikasi pada beberapa aspek, yaitu :

1. Aspek Sistem

   Dari hasil penelitian yang telah dilakukan menunjukan bahwa metode *Naive Bayes* dan *Particle Swarm Optimization* sebagai seleksi atribut dapat meningkatkan keakuratan. Adanya program pengolahan data mining dengan model ini sebagai alat bantu pertimbangan suatu perusahaan digunakan pada strategi pemasaran *cross selling* sehingga efektif dan efisien.

2. Aspek Manajerial

   Pada aspek manajerial, penelitian ini memiliki implikasi untuk membantu pengambil keputusan suatu perusahaan dalam mengambil keputusan terutama pada strategi pemasaran *cross selling* agar lebih efisien dan efektif.

3. Aspek Penelitian Lanjutan

   Penelitian ini perlu untuk terus dikembangkan, diantaranya :

   a. *Extrapolasi*

   Hasil penelitian ini dapat dikembangkan pada perusahaan – perusahaan yang bergerak di bidang lain selain industri telekomunikasi.

   b. *Extention*

   Penelitian selanjutnya dapat mengembangkan untuk *cross selling* dengan menggunakan algoritma klasifikasi *data mining* lain atau metode seleksi atribut lain.

BAB 5 PENUTUP

5.1 Kesimpulan

Penelitian ini menggunakan *data mining* klasifikasi dengan algoritma *Naive Bayes* dan *Particle Swarm Optimization* sebagai seleksi atribut. Dari hasil penelitian diperoleh :

1. Akurasi yang dihasilkan algoritma *Naive Bayes Classifier* yaitu 85,08% , *classification error* = 14,92% dan AUC = 0,841
2. Akurasi yang dihasilkan algoritma *Naive Bayes Classifier* dengan *Particle Swarm Optimization* sebagai seleksi atribut meningkat menjadi yaitu 89,31% , *classification error* = 10,69% dan AUC = 0,845. Hal ini menunjukkan bahwa algoritma *Naive Bayes* dapat meningkatkan akurasinya dengan menggunakan *Particle Swarm Optimization* sebagai seleksi atribut.

Dengan demikian algoritma *Naive Bayes* dengan *Particle Swarm Optimization* sebagai seleksi atribut dapat memberikan pemecahan untuk permasalahan klasifikasi pelanggan *cross selling*.

5.2 Saran

Meskipun pada penelitian algoritma *data mining* ini menghasilkan *Naive Bayes* dengan *Particle Swarm Optimization* sebagai seleksi atribut dapat meningkatkan akurasi namun ada beberapa penambahan yang sebaiknya dilakukan, antara lain adalah:

a. Menambahkan lagi beberapa algoritma klasifikasi *data mining* untuk dikomparasikan seperti *Decision Tree, Logistic Regression, Artificial Neural Network* (ANN), *Support Vector Machines* (SVM) dan lain lain.
b. Memperbanyak jumlah atribut dan memanfaatkan metode penyeleksian atribut seperti *Chi-Square* dan lain-lain.

DAFTAR REFERENSI

Ahn, H. et al., 2011. Facilitating cross-selling in a mobile telecom market to develop customer classification model based on hybrid data mining techniques. *Expert Systems with Applications*, 38(5), pp.5005–5012.

Balaji, S. & K. Srivatsa, S., 2012. Naïve Bayes Classification Approach for Mining Life Insurance Databases for Effective Prediction of Customer Preferences over Life Insurance Products. *International Journal of Computer Applications*, 51(3), pp.22–26.

Bramer, M., 2007. *Principles of Data Mining, ser. Undergraduate Topics in Computer Science.*

Correa, E.S., Freitas, A. a. & Johnson, C.G., 2008. Particle Swarm for Attribute Selection in Bayesian Classification: An Application to Protein Function Prediction. *Journal of Artificial Evolution and Applications*, 2008, pp.1–12.

Elliott, A.C. & Woodward, W. a., 2007. Comparing One or Two Means Using the t -Test. *Statistical Analysis Quick Reference Guidebook: With SPSS examples*, pp.46–76.

Elsalamony, H. a, 2014. Bank Direct Marketing Analysis of Data Mining Techniques. , 85(7), pp.12–22.

Escalante, H.J., 2009. Particle Swarm Model Selection. , 10, pp.405–440.

Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P., 1996. From data mining to knowledge discovery in databases. *AI magazine*, pp.37–54.

Ferdousy, E.Z., Islam, M.M. & Matin, M.A., 2013. Combination of Naïve Bayes Classifier and K-Nearest Neighbor (cNK) in the Classification Based Predictive Models. *Computer and Information Science*, 6(3).

Ghanad, N.K. & Ahmadi, S., 2015. Combination of PSO Algorithm and Naive Bayesian Classification for Parkinson Disease Diagnosis. , 4(4), pp.119–125.

Gorunescu, F., 2011. *Data mining: concepts and techniques*.

Harb, H.M., 2014. Feature Selection on Classification of Medical Datasets based on Particle Swarm Optimization. , 104(5), pp.14–17.

Jaroszewicz, S., 2008. Cross-selling models for telecommunication services. *Praca wys\lana do JTIT*, (December).

Jiawei, H. & Micheline, K., 2006. Data Mining Concepts and Techniques. *The Morgan Kaufman Series in Data Management systems*, p.770.

Kabir, M.F.M. et al., 2011. Enhanced Classification Accuracy on Naive Bayes Data Mining Models. *International Journal of Computer Applications (0975–8887) Volume*, 28(3), pp.9–16.

Karim, M. & Rahman, R.M., 2013. Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing. *Journal of Software Engineering and Applications*, 06(04), pp.196–206.

Kennedy, J. & Eberhart, R., 1995. Particle swarm optimization. *Neural Networks, 1995. Proceedings., IEEE International Conference on*, 4, pp.1942–1948 vol.4.

Kothari, C., 2004. *Research methodology: methods and techniques.*

Larose, D.T., 2005. *Discovering Knowledge in Data an introduction to data mining.*

Li, J., Ding, L. & Li, B., 2014. A Novel Naive Bayes Classification Algorithm Based on Particle Swarm Optimization. , pp.747–753.

Lin, S.-W. et al., 2008. Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Systems with Applications*, 35(4), pp.1817–1824.

Madyatmadja, E.D. & Aryuni, M., 2014. Comparative Study of Data Mining Model for. , 59(2), pp.269–274.

Moro, S. & Laureano, R.M.S., 2011. Using Data Mining for Bank Direct Marketing: An application of the CRISP-DM methodology. *European Simulation and Modelling Conference*, (Figure 1), pp.117–121.

Ren, S., Lian, Y. & Zou, X., 2014. Incremental Naïve Bayesian Learning Algorithm based on Classification Contribution Degree. *Journal of Computers*, 9(8), pp.1967–1974.

Sharma, A. & Panigrahi, P., 2011. A neural network based approach for predicting customer churn in cellular network services. *International …*, 27(11), pp.26–31.

Sheng, L.K. & Wah, T.Y., 2011. A comparative study of data mining techniques in predicting consumers' credit card risk in banks. *African Journal of Business Management*, 5(20), pp.8307–8312.

Taha, A.M., Mustapha, A. & Chen, S.-D., 2013. Naive Bayes-guided bat algorithm for feature selection. *TheScientificWorldJournal*, 2013, p.325973.

Taheri, S. & Mammadov, M., 2013. Learning the naive Bayes classifier with optimization models. *International Journal of Applied Mathematics and Computer Science*, 23(4), pp.787–795.

Ting, S.L., Ip, W.H. & Tsang, A.H.C., 2011. Is Naïve bayes a good classifier for document classification? *International Journal of Software Engineering and its Applications*, 5(3), pp.37–46.

Vihar, V. et al., 2009. Knowledge Discovery from Bio-medical Data Using a Hybrid PSO / Bayesian Classifier. , 2(1), pp.364–371.

Visa, S. et al., 2011. Confusion matrix-based feature selection. *CEUR Workshop Proceedings*, 710, pp.120–127.

Witten, I.H., Frank, E. & Hall, M. a., 2011. *Data Mining: Practical Machine Learning Tools and Techniques, Third Edition.*

Zaidi, N. & Cerquides, J., 2013. Alleviating Naive Bayes attribute independence assumption by attribute weighting. *The Journal of Machine …*, 14, pp.1947–1988.

DAFTAR RIWAYAT HIDUP

A. Biodata Mahasiswa

NIM	: 14001240
Nama	: Indah Purnamasari
Tempat dan Tanggal Lahir	: Jakarta, 28 April 1980
Jenis Kelamin	: Perempuan
Agama	: Islam
Alamat	: Jl. Sedap Malam IV No. 2 Komp. Atsiri Permai Bojong Gede – Kab. Bogor

B. Pendidikan Formal

1. SDN 08 Jakarta, Lulus Tahun 1992
2. SLTPN 85 Jakarta, Lulus Tahun 1995
3. SMK Telkom Jakarta, Lulus Tahun 1998
4. Strata 1 STT Telkom Bandung, Lulus Tahun 2002

Jakarta, 20 Febuari 2016
Saya yang bersangkutan,

Indah Purnamasari

SURAT KETERANGAN RISET

Telkom Indonesia
the world in your hand

No. : Tel. 1/UM 000/DR2-10120000/2016 Jakarta, 25 Januari 2016

Lampiran : -

Hal : Surat Keterangan Riset

SURAT KETERANGAN

Yang bertanda tangan di bawah ini :

Nama : Ilham Fadli

NIK : 730554

Jabatan : Manager Performance & War Room

Dengan ini menerangkan bahwa yang tersebut di bawah ini :

Nama : Indah Purnamasari

NIM : 14001240

Perguruan Tinggi : Program Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri

Adalah benar telah selesai melakukan Riset pada Kantor PT. Telekomunikasi Indonesia, Tbk terhitung sejak tanggal 4 Desember 2015 s/d 15 Januari 2016, dan yang bersangkutan telah melaksanakan tugasnya dengan baik dan penuh tanggung jawab.

Demikian surat keterangan ini dibuat dengan benar, agar dapat dipergunakan sebagaimana mestinya.

PT. Telekomunikasi Indonesia, Tbk

Manager Performance & War Room

ILHAM FADLI

NIK. 730554

Telkom Regional II (JABOTABEK)
PT Telekomunikasi Indonesia, Tbk

KARTU BIMBINGAN TESIS

STMIK - STIBA NUSA MANDIRI	**Kartu Bimbingan Tesis** **STMIK Nusa Mandiri Jakarta**

Nama : Indah Purnamasari

NIM : 14001240

Judul Tesis : ANALISIS ALGORITMA *NAIVE BAYES CLASSIFIER* DENGAN SELEKSI ATRIBUT BERBASIS *PARTICLE SWARM OPTIMIZATION* UNTUK KLASIFIKASI PELANGGAN *CROSS SELLING* PRODUK INDIHOME PT TELKOM JAKARTA

No.	Tanggal Bimbingan	Pokok Bahasan	Paraf Dosen Pembimbing
1.	7 November 2015	Bimbingan Perdana Pengajuan Judul Tesis	
2.	21 November 2015	Bimbingan Data Tesis	
3.	5 Desember 2015	Bimbingan BAB I	
4.	19 Desember 2015	ACC BAB I dan Revisi BAB II	
5.	9 Januari 2016	ACC BAB II dan Revisi BAB III	
6.	23 Januari 2016	ACC BAB III dan Revisi BAB IV & V	
7.	6 Febuari 2016	ACC BAB I – V	

Dimulai pada tanggal : 7 November 2015

Diakhiri pada tanggal : 6 Febuari 2016

Jumlah Pertemuan Bimbingan : 7 Pertemuan

Disetujui oleh,

Dosen Pembimbing

(Dr. Sfenrianto, M.Kom.)

LAMPIRAN

Lampiran 1 Tarif berdasarkan Paket eksisting dan Zona

Paket Eksisting		Zona		
		Attack	Winning	Super winning
INET L15H	Internet Paket MAIL	58000		
INET L50H	Internet Paket CHAT	113000		
INET OFFIC	Internet Internet Unltd Office	233000		
INET SOD	Internet Instan	77000		
INET TBASE	Internet Time Based Personal			
INET U1M	Internet Paket FAMILIA	452000		
INET U256K	Internet Prabayar U256K			
INET U25AL	Internet Prepay (Unlimited 256kbps)			
INET U2M	Internet Paket EXECUTIVE	695000		
INET U384K	Internet Paket SOCIALIA - 384K/96K	234000		
INET U384M	Internet Flash Multi Session 384Kbps	436000		
INET U384R	Internet Paket Socialia Real (Unlimited 384Kbps)	230000		
INET U3M	Internet Paket BIZ	919000		
INET U512K	Internet Paket LOAD	255000		
INET U512R	Internet Paket Load Real (Unlimited 512 Kbps)	300000		
INET_1MB_H	Internet SOCIALIA 1 MB - HSSP 2	311000		
INET_2MB_H	Internet LOAD 2 MB - HSSP 2	491000		
INET_3MB_H	Internet FAMILIA 3MB - HSSP 2	853000		
INET1000H	Internet LOAD 1 MB- HSSP	194000		
INET1000S	Internet SOCIALIA 1 MB – HSSP	260000		
INET10Q103	INET Speed 10 M Kuota 100 Gb Zona 3	330000		
INET10Q21	INET Speed 10 M Kuota 20 Gb Zona 1	182000		
INET10Q22	INET Speed 10 M Kuota 20 Gb Zona 2	83000		
INET10Q23	INET Speed 10 M Kuota 20 Gb Zona 3	201000		
INET2000H	Internet FAMILIA 2MB – HSSP	323000		
INET512RH	Internet SOCIALIA 512 Kb- HSSP	259000		
INETG100M	New Internet - Gold 100M	7325000	5700000	1805000
INETG10M	New Internet - Gold 10M	1913000	1049000	757000
INETG1M	New Internet - Gold 1M	477000	341000	296000
INETG20M	New Internet - Gold 20M	6485000	1851000	687000
INETG2M	New Internet - Gold 2M	538000	335000	295000
INETG3M	New Internet - Gold 3M	855000	513000	457000
INETG50M	New Internet - Gold 50M	10223000	2838000	1310000
INETG512K	New Internet - Gold 512K	271000	247000	228000
INETG5M	New Internet - Gold 5M	1376000	705000	331000
Paket Eksisting		**Zona**		

		Attack	Winning	Super_winning
INETR100M	New Internet - Reguler 100M	2928000	2885000	1566000
INETR10M	New Internet - Reguler 10M	519000	399000	389000
INETR1M	New Internet - Reguler 1M	289000	257000	245000
INETR20M	New Internet - Reguler 20M	1103000	649000	598000
INETR2M	New Internet - Reguler 2M	446000	399000	310000
INETR3M	New Internet - Reguler 3M	427000	375000	351000
INETR50M	New Internet - Reguler 50M	1671000	1382000	1139000
INETR512K	New Internet - Reguler 512K	243000	207000	189000
INETR5M	New Internet - Reguler 5M	436000	391000	343000

Lampiran 2 Perhitungan Probabilitas Prior NBC

Berikut ini merupakan tabel hasil perhitungan probabilitas prior dengan menggunakan Naive Bayes yaitu :

ATRIBUT		JML KASUS	3P	DECLINE	P(H|X)	
					3P	DECLINE
TOTAL		878	587	291	0,66856492	0,33143508
REGIONAL	2	878	587	291	1	1
WITEL	BANTEN BARAT (SERANG)	12	12		0,02044293	0
	BANTEN TIMUR (TANGERANG)	55	1	54	0,00170358	0,18556701
	JABAR BARAT (BOGOR)	89	12	77	0,02044293	0,264604811
	JABAR BARAT UTARA (BEKASI)	91	6	85	0,01022147	0,29209622
	JAKARTA BARAT	230	202	28	0,34412266	0,096219931
	JAKARTA PUSAT	21	17	4	0,02896082	0,013745704
	JAKARTA SELATAN	96	85	11	0,14480409	0,037800687
	JAKARTA TIMUR	157	137	20	0,23339012	0,068728522
	JAKARTA UTARA	127	115	12	0,19591141	0,041237113
KWADRAN	4	878	587	291	1	1
TYPE JARINGAN	COPPER	509	386	123	0,65758092	0,422680412
	FTTH	231	110	121	0,18739353	0,41580756
	MSAN	138	91	47	0,15502555	0,161512027
ZONA	ATTACK_PROMO	16	10	6	0,01703578	0,020618557
	SUPER_WINNING	745	493	252	0,83986371	0,865979381
	WINNING	117	84	33	0,14310051	0,113402062
MIGRASI	2P	878	587	291	1	1
R2BB	GIPTV	228	107	121	0,18228279	0,41580756
	GSP5	1		1	0	0,003436426
	GTLP	2	2		0,00340716	0
	IPTV	137	93	44	0,15843271	0,151202749
	SP2	66	49	17	0,0834753	0,058419244
	SP3	94	75	19	0,12776831	0,065292096
	SP4	128	99	29	0,16865417	0,099656357
	SP5	56	45	11	0,07666099	0,037800687
	SP7	166	117	49	0,19931857	0,16838488

ATRIBUT		JML KASUS	3P	DECLINE	P(H|X)	
					3P	DECLINE
PAKET EKSISTING	INET L15H	6	6		0,01022147	0
	INET L50H	3	1	2	0,00170358	0,006872852
	INET U1M	15	10	5	0,01703578	0,017182131
	INET U2M	5	4	1	0,00681431	0,003436426
	INET U384K	25	20	5	0,03407155	0,017182131
	INET U384R	6	3	3	0,00511073	0,010309278
	INET U3M	1	1		0,00170358	0
	INET U512K	4	4		0,00681431	0
	INET U512R	2	2		0,00340716	0
	INET_1MB_H	6	5	1	0,00851789	0,003436426
	INET_2MB_H	2	1	1	0,00170358	0,003436426
	INET1000S	28	18	10	0,0306644	0,034364261
	INETG10M1	1	1		0,00170358	0
	INETG2M1	2	2		0,00340716	0
	INETR10M1	43	26	17	0,04429302	0,058419244
	INETR10M3	6	1	5	0,00170358	0,017182131
	INETR1M1	299	200	99	0,3407155	0,340206186
	INETR1M2	3		3	0	0,010309278
	INETR1M3	166	107	59	0,18228279	0,202749141
	INETR2M1	104	82	22	0,13969336	0,075601375
	INETR2M3	6	6		0,01022147	0
	INETR3M1	71	37	34	0,06303237	0,116838488
	INETR3M2	1	1		0,00170358	0
	INETR3M3	6	5	1	0,00851789	0,003436426
	INETR50M1	1	1		0,00170358	0
	INETR512K1	35	18	17	0,0306644	0,058419244
	INETR512K2	4	3	1	0,00511073	0,003436426
	INETR512K3	3	3		0,00511073	0
	INETR5M1	22	17	5	0,02896082	0,017182131
	INETR5M3	1	1		0,00170358	0
	GIPTV	1	1		0,00170358	0

Lampiran 3 Probabilitas Prior PSO-NBC

Berikut ini merupakan tabel hasil perhitungan probabilitas prior dengan menggunakan *Particle Swarm Optimization - Naive Bayes Classifier* yaitu :

ATRIBUT		JML KASUS	3P	DECLINE	P(H|X)	
					3P	DECLINE
TOTAL		878	587	291	0,66856492	0,33143508
WITEL	BANTEN BARAT (SERANG)	12	12		0,02044293	0
	BANTEN TIMUR (TANGERANG)	55	1	54	0,001703578	0,18556701
	JABAR BARAT (BOGOR)	89	12	77	0,02044293	0,264604811
	JABAR BARAT UTARA (BEKASI)	91	6	85	0,010221465	0,29209622
	JAKARTA BARAT	230	202	28	0,344122658	0,096219931
	JAKARTA PUSAT	21	17	4	0,028960818	0,013745704
	JAKARTA SELATAN	96	85	11	0,144804089	0,037800687
	JAKARTA TIMUR	157	137	20	0,233390119	0,068728522
	JAKARTA UTARA	127	115	12	0,195911414	0,041237113
TYPE JARINGAN	COPPER	509	386	123	0,65758092	0,422680412
	FTTH	231	110	121	0,187393526	0,41580756
	MSAN	138	91	47	0,155025554	0,161512027

Lampiran 4 Data pelanggan eksisting dan Hasil penelitian

NO	REGIONAL	WITEL	KWADRAN	PAKET EKSISTING	R2BB	TYPE JARINGAN	ZONA	MIGRASI	PAKET	NB	PSO-NB
1	2	JAKARTA UTARA	4	INETR1M1	SP5	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
2	2	JAKARTA UTARA	4	INETR1M1	SP5	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
3	2	JAKARTA BARAT	4	INETR1M1	IPTV	MSAN	SUPER_WINNING	2P	DECLINE	3P	3P
4	2	JAKARTA BARAT	4	INETR1M1	IPTV	MSAN	SUPER_WINNING	2P	DECLINE	3P	3P
5	2	JAKARTA BARAT	4	INETR1M1	SP3	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
6	2	JAKARTA BARAT	4	INETR1M1	SP3	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
7	2	JAKARTA BARAT	4	INETR3M1	IPTV	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
8	2	JAKARTA BARAT	4	INETR3M1	IPTV	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
9	2	JAKARTA BARAT	4	INET U1M	IPTV	COPPER	ATTACK_PROMO	2P	DECLINE	3P	3P
10	2	JAKARTA BARAT	4	INET U1M	IPTV	COPPER	ATTACK_PROMO	2P	DECLINE	3P	3P
11	2	JAKARTA TIMUR	4	INETR1M1	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE	3P	3P
12	2	JAKARTA TIMUR	4	INETR1M1	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE	3P	3P
13	2	JAKARTA TIMUR	4	INETR1M1	SP4	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
14	2	JAKARTA TIMUR	4	INETR1M1	SP4	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
15	2	JAKARTA BARAT	4	INETR2M1	SP5	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
16	2	JAKARTA BARAT	4	INETR2M1	SP5	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P

NO	REGIONAL	WITEL	KWADRAN	PAKET EKSISTING	R2BB	TYPE JARINGAN	ZONA	MIGRASI	PAKET	NB	PSO-NB
17	2	JABAR BARAT UTARA (BEKASI)	4	INETR3M1	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE	DECLINE	DECLINE
18	2	JABAR BARAT UTARA (BEKASI)	4	INETR3M1	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE	DECLINE	DECLINE
19	2	JAKARTA TIMUR	4	INETR1M3	SP3	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
20	2	JAKARTA TIMUR	4	INETR1M3	SP3	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
21	2	JAKARTA UTARA	4	INETR1M1	SP7	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
22	2	JAKARTA UTARA	4	INETR1M1	SP7	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
23	2	JAKARTA TIMUR	4	INETR1M1	SP4	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
24	2	JAKARTA TIMUR	4	INETR1M1	SP4	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
25	2	JABAR BARAT UTARA (BEKASI)	4	INETR1M1	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE	DECLINE	DECLINE
26	2	JABAR BARAT UTARA (BEKASI)	4	INETR1M1	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE	DECLINE	DECLINE
27	2	JAKARTA BARAT	4	INETR1M3	SP7	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
28	2	JAKARTA BARAT	4	INETR1M3	SP7	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
29	2	JAKARTA BARAT	4	INETR2M1	SP7	MSAN	WINNING	2P	DECLINE	3P	3P
30	2	JAKARTA BARAT	4	INETR2M1	SP7	MSAN	WINNING	2P	DECLINE	3P	3P
31	2	JAKARTA BARAT	4	INETR1M3	IPTV	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
32	2	JAKARTA BARAT	4	INETR1M3	IPTV	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P

NO	REGIONAL	WITEL	KWADRAN	PAKET EKSISTING	R2BB	TYPE JARINGAN	ZONA	MIGRASI	PAKET	NB	PSO-NB
33	2	JAKARTA TIMUR	4	INETR512K1	IPTV	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
34	2	JAKARTA TIMUR	4	INETR512K1	IPTV	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
35	2	JAKARTA BARAT	4	INETR1M1	SP4	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
36	2	JAKARTA BARAT	4	INETR1M1	SP4	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
37	2	JABAR BARAT (BOGOR)	4	INETR1M1	SP2	COPPER	SUPER_WINNING	2P	DECLINE	DECLINE	DECLINE
38	2	JABAR BARAT (BOGOR)	4	INETR1M1	SP2	COPPER	SUPER_WINNING	2P	DECLINE	DECLINE	DECLINE
39	2	JAKARTA TIMUR	4	INETR3M1	SP5	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
40	2	JAKARTA TIMUR	4	INETR3M1	SP5	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
41	2	JAKARTA TIMUR	4	INET U384R	SP3	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
42	2	JAKARTA TIMUR	4	INET U384R	SP3	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
43	2	JAKARTA UTARA	4	INETR1M1	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE	3P	3P
44	2	JAKARTA UTARA	4	INETR1M1	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE	3P	3P
45	2	JAKARTA PUSAT	4	INETR10M1	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE	DECLINE	3P
46	2	JAKARTA PUSAT	4	INETR10M1	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE	DECLINE	3P
47	2	JAKARTA PUSAT	4	INETR3M1	GIPTV	FTTH	WINNING	2P	DECLINE	DECLINE	3P
48	2	JAKARTA PUSAT	4	INETR3M1	GIPTV	FTTH	WINNING	2P	DECLINE	DECLINE	3P
49	2	JAKARTA UTARA	4	INETR3M1	SP7	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P
50	2	JAKARTA UTARA	4	INETR3M1	SP7	COPPER	SUPER_WINNING	2P	DECLINE	3P	3P

NO	REGIONAL	WITEL	KWADRAN	PAKET EKSISTING	R2BB	TYPE JARINGAN	ZONA	MIGRASI	PAKET	NB	PSO-NB
51	2	BANTEN BARAT (SERANG)	4	INETR1M3	GTLP	FTTH	SUPER_WINNING	2P	3P	3P	3P
52	2	BANTEN BARAT (SERANG)	4	INETR1M3	GTLP	FTTH	SUPER_WINNING	2P	3P	3P	3P
53	2	JAKARTA TIMUR	4	INETR1M1	GIPTV	FTTH	SUPER_WINNING	2P	3P	3P	3P
54	2	JAKARTA TIMUR	4	INETR1M1	GIPTV	FTTH	SUPER_WINNING	2P	3P	3P	3P
55	2	JABAR BARAT UTARA (BEKASI)	4	INETR1M1	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE	DECLINE	DECLINE
56	2	JABAR BARAT UTARA (BEKASI)	4	INETR1M1	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE	DECLINE	DECLINE
57	2	JAKARTA BARAT	4	INETR3M1	GIPTV	FTTH	SUPER_WINNING	2P	3P	DECLINE	3P
58	2	JAKARTA BARAT	4	INETR3M1	GIPTV	FTTH	SUPER_WINNING	2P	3P	DECLINE	3P
59	2	JAKARTA TIMUR	4	INETR1M3	SP4	COPPER	SUPER_WINNING	2P	3P	3P	3P
60	2	JAKARTA TIMUR	4	INETR1M3	SP4	COPPER	SUPER_WINNING	2P	3P	3P	3P
61	2	JABAR BARAT UTARA (BEKASI)	4	INETR1M3	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE	DECLINE	DECLINE
62	2	JABAR BARAT UTARA (BEKASI)	4	INETR1M3	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE	DECLINE	DECLINE
63	2	JABAR BARAT UTARA (BEKASI)	4	INETR1M1	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE	DECLINE	DECLINE
64	2	JABAR BARAT UTARA (BEKASI)	4	INETR1M1	GIPTV	FTTH	SUPER_WINNING	2P	DECLINE	DECLINE	DECLINE