

## **KLASIFIKASI PELANGGAN PRODUK INDIHOME MENGGUNAKAN NAIVE BAYES CLASSIFIER DENGAN SELEKSI FITUR ALGORITMA GENETIK**

**Indah Purnamasari**

Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri  
Jl. Damai No. 8, Warung Jati Barat, Pasar Minggu, Jakarta 12540  
<http://www.nusamandiri.ac.id>  
[indah.ih@nusamandiri.ac.id](mailto:indah.ih@nusamandiri.ac.id)

*Abstract - Classification is a method of data mining used to organize data systematically or according to some rules or rules that have been established. PT. TELKOM as a state-owned company engaged in telecommunications and its development has a variety of telecommunications products that need to be marketed. Marketing strategies are intended to market the company's products for profit especially when it comes to offering new products. However, too many inappropriate offers to customers will only make marketing inefficient and effective. To simplify the marketing strategy of PT. TELKOM then made the classification of telephone subscribers and speedy that already exist in PT. TELKOM Jakarta which has the potential to become a new product customer of Indihome so that the marketing done by PT. TELKOM becomes more effective and efficient. This research uses data mining classification with Naive Bayes Classifier with genetic algorithm as feature selection. This is because Naive Bayes Classifier tends to perform well and can learn quickly in various classification problems, as evidenced by the results of classification accuracy in this study included in the category of good classification.*

**Keyword :** Naive Bayes Classifier (NBC), Genetic Algorithm (GA), Indihome, data mining

### **I. PENDAHULUAN**

Klasifikasi merupakan metode untuk menyusun data secara sistematis atau menurut beberapa aturan atau kaidah yang telah ditetapkan. Sementara penggunaan metode seleksi fitur membantu memahami atribut yang relevan serta dapat meningkatkan akurasi klasifikasi.

Strategi pemasaran dengan menjual produk baru untuk pelanggan eksisting yang telah menggunakan produk lainnya. Strategi pemasaran ini untuk memperoleh keuntungan dari produk tambahan yang dijual (Jaroszewicz, 2008).

Salah satu strategi pemasaran PT. Telkom Jakarta dilakukan dengan menawarkan produk baru Indihome kepada pelanggan telepon dan pelanggan speedy eksisting. Tetapi terlalu banyak penawaran hanya akan membuat pelanggan menolaknya. Untuk itu agar pemasaran produk baru Indihome lebih efisien dan efektif maka dilakukan klasifikasi terhadap pelanggan telepon dan pelanggan speedy eksisting PT. Telkom Jakarta yang berpotensi untuk menjadi pelanggan produk Indihome.

Naive Bayes telah digunakan sebagai pengklasifikasi yang efektif bertahun-tahun. Naive Bayes mudah untuk dibangun karena strukturnya diberi atribut bersifat *independent* untuk memecahkan masalah (Gorunescu, 2011). Oleh karena sifat *independent* ini maka performa Naive Bayes Classifier menjadi tidak sempurna (Jun Li, dkk 2014) untuk meningkatkan kinerja Naive Bayes

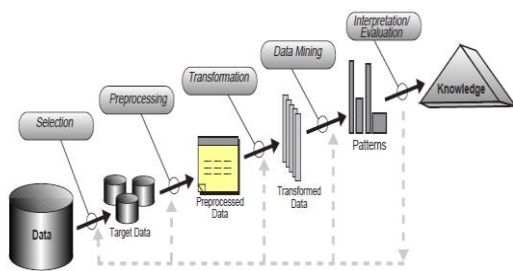
perlu dioptimasi dengan seleksi fitur (Triwulan, dkk 2017).

Pada penelitian ini mengklasifikasikan pelanggan telepon dan pelanggan speedy eksisting untuk pemasaran produk Indihome menggunakan Naive Bayes Classifier (NBC) dengan seleksi atribut algoritma genetika (GA).

### **1.1 Data Mining**

*Data mining* adalah proses yang menggunakan berbagai analisis data alat untuk menemukan pola dan hubungan dalam data itu dapat digunakan untuk membuat prediksi yang valid. Teknik dan algoritma ini sering digunakan untuk pemasaran. Penemuan pengetahuan dengan *data mining* adalah proses penemuan yang sebelumnya tidak diketahui dan pola dan hubungan yang berpotensi menarik dalam database besar. Prediksi dan keputusan masa depan bisa dilakukan berdasarkan penemuan pengetahuan melalui *data mining*. (Masud Karim & Rashedur M Rahman, 2013)

*Data mining* merupakan salah satu tahap dari proses *Knowledge Discovery in Databases* (KDD). Proses KDD menurut adalah sebagai berikut :



Sumber : Fayyad et al (1996)

Gambar 1. Proses KDD

1. Data Selection
2. Preprocessing
3. Transformation
4. Data Mining
5. Interpretation / Evaluation

### 1.2 Naive Bayes Classifier

Algoritma Naive Bayes adalah *classifier* probabilistik sederhana yang menghitung seperangkat probabilitas dengan menghitung frekuensi dan kombinasi nilai pada kumpulan data yang diberikan. Algoritma ini menggunakan teorema Bayes dan mengasumsikan semua atribut menjadi independen mengingat nilai variabel kelas. Ini kondisional dengan asumsi bahwa kemerdekaan jarang berlaku pada aplikasi dunia, maka karakterisasi diasumsikan sebagai Naif namun algoritma cenderung berkinerja baik dan dapat belajar dengan cepat dalam berbagai masalah klasifikasi. (Patil & Sherekar, 2013)

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

Keterangan :

$X$  : Data dengan class yang belum diketahui

$H$  : Hipotesis data  $X$  merupakan suatu kelas spesifik

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

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

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

$P(X)$  : Probabilitas  $X$

Proses klasifikasi memerlukan petunjuk untuk menentukan kelas. Karena itu, teorema Bayes di atas disesuaikan sebagai berikut :

$$P(H|X) = P(X|H) \cdot P(H) \quad (2)$$

Klasifikasi dengan data kontinyu digunakan rumus *Densitas Gauss* :

$$P(X, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3)$$

$$\sqrt{2\pi\sigma^2}$$

Keterangan :

$\mu$  : Mean, menyatakan rata – rata dari seluruh atribut

$\sigma$  : Deviasi standar, menyatakan varian dari seluruh atribut

$\pi$  = 3,1416

$e$  = 2,7183

$$\sigma^2 = \frac{1}{N-1} \sum (X_i - \mu)^2 \quad (4)$$

### 1.3 Algoritma Genetika

Algoritma genetika (GA) merupakan algoritma pengoptimalan metaheuristik atau teknik identifikasi yang mendekati solusi untuk masalah optimasi dan pencarian, menjadi kelas tertentu. GA berdasarkan populasi solusi potensial dan menggunakan mekanisme spesifik yang terinspirasi oleh evolusi biologis (natural genetika), seperti: individu (kromosom), reproduksi, mutasi, rekombinasi, seleksi, kelangsungan hidup yang terkuat (Gorunescu 2011).

Kromosom-kromosom berevolusi beberapa kali tahapan iterasi yang disebut dengan generasi. Generasi baru (offsprings) digenerate dengan teknik kawin silang (crossover) dan mutasi (mutation). Kromosom-kromosom tersebut kemudian berevolusi dengan suatu kriteria kesesuaian (fitness) yang telah ditetapkan, hasil yang terbaik akan dipilih sedangkan yang lainnya diabaikan (Suhartono 2015). Rumus fitness yang digunakan, (Lee dkk, 2001) adalah sebagai berikut :

$$Fitness = \frac{1}{1+(F1B1+F2B2+\dots)} \quad (5)$$

Keterangan :  $B_n$  = Bobot pelanggaran

$F_n$  = Banyaknya pelanggaran

$n = 1 \dots n$

### 1.4 Indihome

Indihome Internet Fiber merupakan produk baru PT. TELKOM yang berupa layanan *Triple Play* yaitu terdiri dari 3 layanan dalam satu *bundle* antara lain *High Speed Internet on Fiber* (Internet cepat), *Interactive TV* (Use TV) dan Telepon Rumah dengan menggunakan 100% fiber optic artinya fiber optic digelar sampai ke rumah pelanggan (*Fiber To The Home / FTTH*).

Fitur-fitur tambahan pada produk Indihome seperti *wifi.id Seamless*, Indihome Telkomsel Mania, Indohime *Global Call*, Indihome *view*, MelOn dan *Trend Micro Security System*.

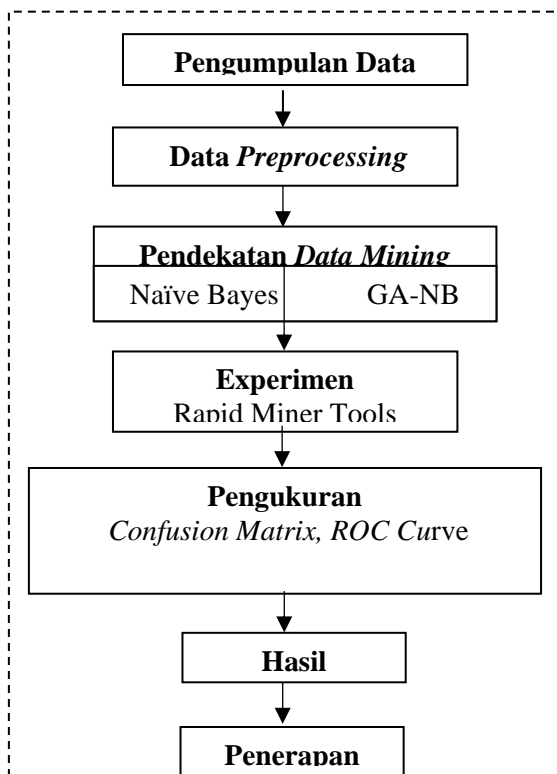
Use TV merupakan layanan televisi berbasis internet protokol (IPTV) diatur oleh Peraturan Menteri Komunikasi dan Informatika Nomor 15 Tahun 2014 yang merupakan perubahan dari

Peraturan Menteri Komunikasi dan Informatika Nomor: 11/PER/M.KOMINFO/07/2010 tentang penyelenggaraan layanan IPTV.

## II. METODOLOGI PENELITIAN

### Desain Penelitian

Tahapan penelitian adalah sebagai berikut:



Gambar 2. Desain Penelitian

### 1. Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data awal berupa data primer atau sekunder untuk menemukan wawasan awal tentang data dan mendeteksi subset menarik mengenai informasi penting yang tersembunyi.

### 2. Preprocessing Data

Beberapa teknik yang dilakukan untuk mendapatkan data berkualitas tinggi (Harb, 2014) yaitu :

- a. *Data Cleaning*  
 untuk membersihkan inkonsistensi data
- b. *Data Integration*  
 Untuk menggabungkan data warehouse
- c. *Data Reduction*  
 Untuk menghilangkan data yang tidak lengkap
- d. *Data Transformation*  
 Untuk mengubah data menjadi yang dibutuhkan dalam proses *data mining*.

### 3. Pendekatan Data Mining

Pada tahap ini dilakukan klasifikasi pelanggan eksisting menggunakan atribut-atribut dari data terkait dengan algoritma *Naive Bayes Classifier* (NBC).

### 4. Experimen

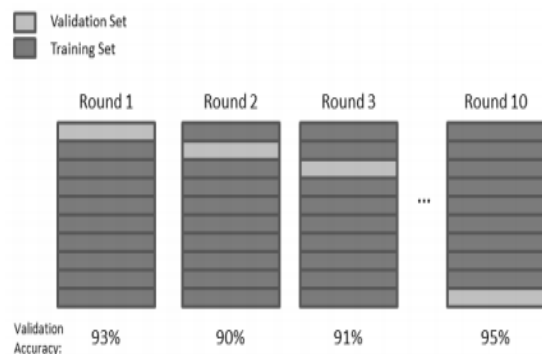
Eksperimen yang dilakukan menggunakan alat bantu *hardware* dan *software* untuk pemrosesan data.

### 5. Pengukuran

Dalam tahap ini akan dilakukan pengukuran untuk menguji model penelitian dan *data testing* dengan menggunakan *cross validation*, *confusion matrix*, *ROC curve*.

#### a. K - Fold Cross Validation

*K-fold Cross Validation* merupakan teknik validasi (Witten et al. 2011) membagi data secara acak ke dalam K bagian dan masing-masing bagian tersebut akan dilakukan proses klasifikasi.



Sumber : Witten et al. (2011)

Gambar 3. 10-fold cross validation

#### b. Confusion Matrix

*Confusion Matrix* berisi informasi tentang aktual dan prediksi klasifikasi dilakukan dengan klasifikasi sistem. Kinerja sistem seperti itu biasanya dievaluasi menggunakan data dalam matriks (Karim, M. & Rahman, R.M., 2013).

Tabel 1. *Confusion Matrix*

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

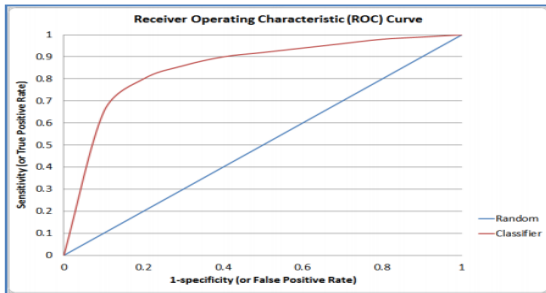
Sumber : Masud Karim & Rahman (2013)

Nilai akurasi dapat dihitung dengan persamaan berikut ini :

$$\text{Akurasi} = \frac{tp + tn}{tp + tn + fp + fn} \quad (6)$$

c. *ROC Curve*

*ROC curve (Receiver Operating Characteristic)* adalah salah satu untuk mengevaluasi akurasi dari klasifikasi secara visual.



Sumber : Jiawei & Micheline (2006)  
Gambar 4. *ROC curve*

**6. Hasil**

Setelah dilakukan pengukuran terhadap model dengan pendekatan *Naive Bayes Classifier (NBC)* dan model dengan pendekatan *GA-NBC* maka diperoleh hasil akurasi dari pengukuran *cross validation, confusion matriks, ROC curve* untuk klasifikasi pelanggan telepon dan speedy eksisting.

**7. Penerapan**

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

**III. HASIL DAN PEMBAHASAN**

**1. Pengumpulan Data**

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 atribut* (atribut output) yaitu atribut Paket. 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 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, 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, 254xxxxxxxx)
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)

**2. Data Preprocessing**

Data penelitian memiliki 14 atribut termasuk *class label atribut* (atribut output) yaitu atribut Paket. Akan tetapi terdapat 5 atribut NCLI, Nama, Alamat, POTS, Speedy mempunyai nilai yang tidak akan sama (unik) untuk setiap pelanggan sehingga pada penelitian ini hanya menggunakan 9 atribut.

Data yang mengandung *missing value* diisi dengan nilai rata-rata.

**3. Pendekatan Data Mining Naive Bayes Classifier**

Perhitungan probabilitas prior algoritma Naive Bayes

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

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

Tabel 2. Probabilitas Prior

ATRIBUT	JML KASUS	3P	DECLINE	P(H X)		
				3P	DECLINE	
TOTAL	878	587	291	0,67	0,33	
REGIONAL	2	878	587	1	1	
WITEL	BANTEN					
	BARAT (SERANG)	12	12		0,02	0
	BANTEN TIMUR (TANGERANG)	55	1	54	0	0,19
	JABAR BARAT (BOGOR)	89	12	77	0,02	0,26
	JABAR BARAT UTARA (BEKASI)	91	6	85	0,01	0,29
	JAKARTA BARAT	230	202	28	0,34	0,1
	JAKARTA PUSAT	21	17	4	0,03	0,01
	JAKARTA SELATAN	96	85	11	0,14	0,04
	JAKARTA TIMUR	157	137	20	0,23	0,07
	JAKARTA UTARA	127	115	12	0,2	0,04
KWADRAN	4	878	587	1	1	
TYPE JARINGAN	COPPER	509	386	123	0,66	0,42
	FTTH	231	110	121	0,19	0,42
	MSAN	138	91	47	0,16	0,16
ZONA	ATTACK_PROMO	16	10	6	0,02	0,02
	SUPER_WINNING	745	493	252	0,84	0,87
	WINNING	117	84	33	0,14	0,11
MIGRASI	2P	878	587	291	1	1
R2BB	GIPTV	228	107	121	0,18	0,42
	GSP5	1		1	0	0

GTLP	2	2		0	0	
IPTV	137	93	44	0,16	0,15	
SP2	66	49	17	0,08	0,06	
SP3	94	75	19	0,13	0,07	
SP4	128	99	29	0,17	0,1	
SP5	56	45	11	0,08	0,04	
SP7	166	117	49	0,2	0,17	
PAKET EKSTING	INET L15H	6	6		0,01	0
	INET L50H	3	1	2	0	0,01
	INET U1M	15	10	5	0,02	0,02
	INET U2M	5	4	1	0,01	0
	INET U384K	25	20	5	0,03	0,02
	INET U384R	6	3	3	0,01	0,01
	INET U3M	1	1		0	0
	INET U512K	4	4		0,01	0
	INET U512R	2	2		0	0
	INET_1 MB_H	6	5	1	0,01	0
	INET_2 MB_H	2	1	1	0	0
	INET100 OS	28	18	10	0,03	0,03
	INETG1 0M1	1	1		0	0
	INETG2 M1	2	2		0	0
	INETR10 M1	43	26	17	0,04	0,06
	INETR10 M3	6	1	5	0	0,02
	INETR1 M1	299	200	99	0,34	0,34
	INETR1 M2	3		3	0	0,01
	INETR1 M3	166	107	59	0,18	0,2
	INETR2 M1	104	82	22	0,14	0,08
	INETR2 M3	6	6		0,01	0
INETR3 M1	71	37	34	0,06	0,12	
INETR3 M2	1	1		0	0	
INETR3 M3	6	5	1	0,01	0	
INETR50 M1	1	1		0	0	
INETR51 2K1	35	18	17	0,03	0,06	
INETR51 2K2	4	3	1	0,01	0	
INETR51 2K3	3	3		0,01	0	
INETR5 M1	22	17	5	0,03	0,02	

INETR5 M3	1	1	0	0
GIPTV	1	1	0	0

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

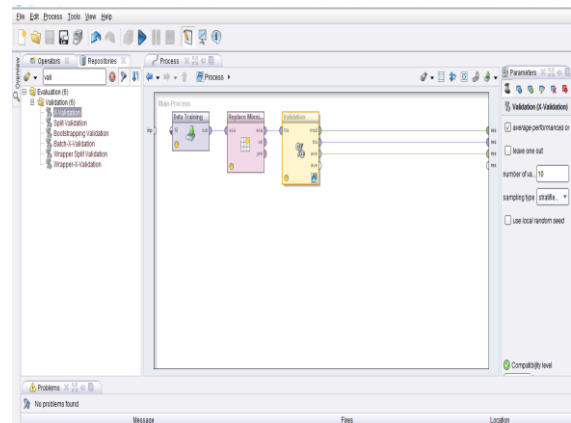
Tabel 3. Probabilitas Posterior

ATRIBUT	NILAI	P(H X)	
		3P	DECLINE
REGIONAL	2	1	1
WITEL	Jakarta Utara	0,2	0,04
KWADRAN	4	1	1
PAKET EKSISTING	INETR1M1	0,34	0,34
TYPE JARINGAN	SP7	0,2	0,17
ZONA	Winning	0,14	0,11
MIGRASI	2P	1	1
R2BB	MSAN	0,16	0,16

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

### 5. Eksperimen

Eksperimen klasifikasi pelanggan eksisting dilakukan menggunakan *software Rapidminer 5.3* berikut ini :



Gambar 5. Model NBC

$$P(H|3P) = 1 \times 0,20 \times 1 \times 0,34 \times 0,20 \times 0,16 \times 1 \times 0,14 = \mathbf{0,000304}$$

$$P(H|Decline) = 1 \times 0,04 \times 1 \times 0,34 \times 0,17 \times 0,16 \times 1 \times 0,11 = \mathbf{0,000040}$$

$$P(H| \text{paket} = 3P) P(3P) = \mathbf{0,000304} \times 0,66856492 = \mathbf{0,000203}$$

$$P(H| \text{paket} = Decline) P(Decline) = \mathbf{0,000040} \times 0,33143508 = \mathbf{0,000013}$$

sehingga

$$P(H| \text{paket} = 3P) P(3P) > P(H| \text{paket} = Decline) P(Decline)$$

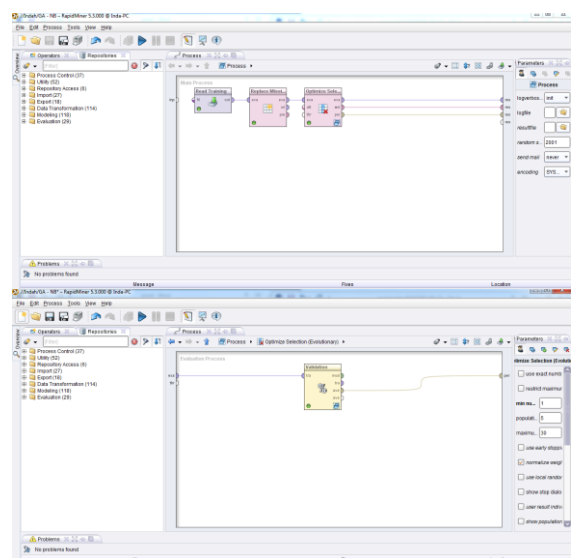
Dari hasil perhitungan tersebut diketahui nilai  $P(H|3P)$  lebih besar dari pada nilai  $P(H|Decline)$ , sehingga dapat disimpulkan bahwa untuk kasus tersebut masuk kedalam klasifikasi **3P**.

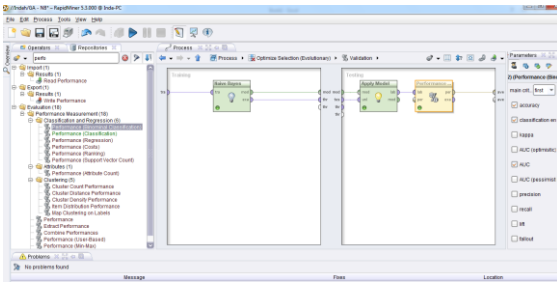
### 4. Algoritma Genetika- Naive Bayes Classifier

Model GA-NBC akan menseleksi atribut prediksi yang ada. Berikut ini adalah hasil seleksi atribut :

Tabel 4. Hasil GA

ATRIBUT	WEIGHT
---------	--------





Gambar 6. Model GA-NBC

### 6. Pengukuran

Parameter yang digunakan pada operator performance adalah *accuracy*, *classification error*, *Area Under Curve* (AUC) untuk menampilkan tingkat keakurasian model NBC dan model GA-NBC.

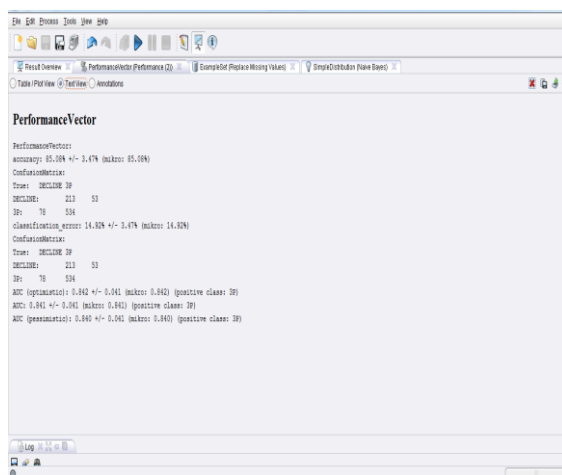
### 7. Hasil Analisis dan Pembahasan

Berdasarkan analisa hasil evaluasi data mining untuk klasifikasi dengan algoritma *Naive Bayes* maka dapat dirangkum sebagai berikut :

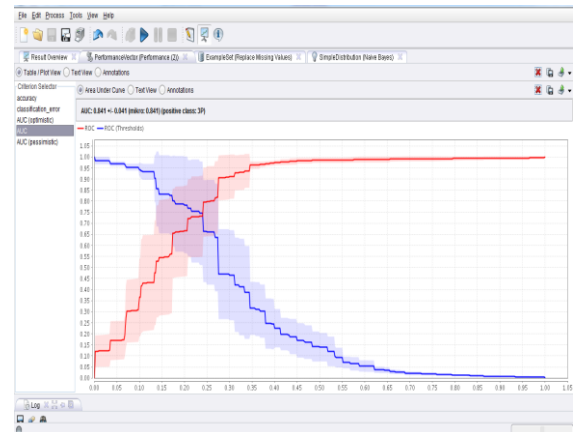
Tabel 5. Perbandingan Performansi

Pengukuran	Hasil	
	NBC	GA - NBC
Accuracy	85,08%	89,31%
Classification Error	14,92%	10,69%
AUC	0.841	0.843

Tabel diatas menunjukkan pendekatan *Naive Bayes* menghasilkan nilai akurasi sebesar 85,08% sementara jika menggunakan seleksi fitur algoritma genetika maka akurasi model meningkat menjadi 89,31%.



Gambar 7. Performansi NBC



Gambar 8. ROC Curve NBC

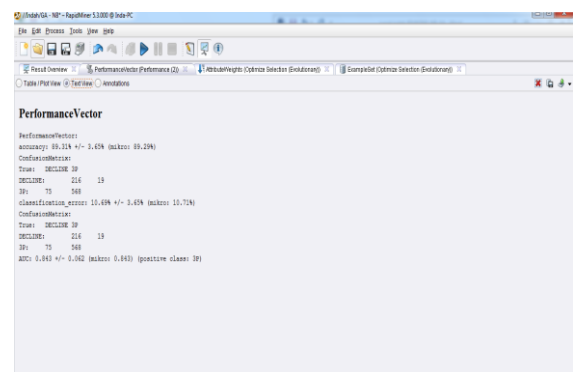
#### a. Confusion Matrix

Nilai akurasi yang dihasilkan adalah 85,08%. Berdasarkan persamaan 6, perhitungan nilai akurasi yaitu :

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

#### b. ROC Curve

Hasil AUC adalah 0.841 sehingga dapat disimpulkan bahwa model *Naive Bayes* ini merupakan model yang Baik untuk digunakan sebagai klasifikasi pelanggan bahwa akurasi 0.80 – 0.90 = *Good classification*.



Gambar 9. Performansi GA-NBC



Gambar 9. ROC Curve GA- NBC

#### a. Confusion Matrix

Nilai akurasi yang dihasilkan adalah 89,31%. Berdasarkan persamaan 6, perhitungan nilai akurasi yaitu :

$$\text{Akurasi} = \frac{216 + 568}{216 + 19 + 568 + 75} = 0,8931$$

b. *ROC Curve*

Hasil AUC adalah 0.843 sehingga dapat disimpulkan bahwa model *Naive Bayes* dengan seleksi fitur Algoritma Genetika ini merupakan model yang Baik untuk digunakan sebagai klasifikasi pelanggan bahwa akurasi 0.80 – 0.90 = *Good classification*

### 8. Penerapan

Berdasarkan hasil penelitian diperoleh hasil yang baik untuk diterapkan dalam pembuatan sistem aplikasi klasifikasi pelanggan menggunakan NBC dengan seleksi fitur Algoritma Genetika dapat memudahkan dalam strategi pemasaran PT. Telkom Jakarta. Sistem aplikasi yang digunakan dalam penelitian ini yaitu dengan menggunakan Microsoft Visual Basic 6.0.

Gambar 8. Hasil klasifikasi GUI

### IV. KESIMPULAN

Hasil penelitian menunjukkan bahwa dengan seleksi fitur Algoritma Genetika maka model *Naive Bayes Classifier* menjadi model klasifikasi yang lebih baik untuk diterapkan pada klasifikasi pelanggan eksisting PT. TELKOM dalam pemasaran produk barunya yaitu Indihome.

1. Akurasi yang dihasilkan model *Naive Bayes Classifier* yaitu 85,08% ,
2. Akurasi yang dihasilkan model *Naive Bayes Classifier* dengan seleksi fitur Algoritma Genetika meningkat menjadi 89,31%

### Saran

Meskipun hasil penelitian ini dengan seleksi fitur Algoritma Genetika maka model *Naive Bayes Classifier* menjadi model klasifikasi yang lebih baik dimana tercapai peningkatan akurasi 89,31%, namun penelitian ini dapat dikomparasikan dengan algoritma klasifikasi lain seperti *Decision Tree*, *Logistic Regression*, *Artificial Neural Network* (ANN), *Support Vector Machines* (SVM) atau dengan menggunakan seleksi atribut lainnya.

### 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 & Srivatsa, S. K., 2012. Naive Bayes Classification Approach for Mining Life Insurance Databases for Effective Prediction of Customer Preferences over Life Insurance Products
- Bramer, M., 2007. *Principles of Data Mining, ser. Undergraduate Topics in Computer Science*.
- 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 wysłana do JTIT*, (December).
- Jun Li, Lixin Ding & Bo Li, 2014. A Novel Naive Bayes Classification Algorithm Based on Particle Swarm Optimization. *The Open Automation and Control Systems Journal*, 747-753
- 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 Naive Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing. *Journal of Software Engineering and Applications*, 06(04), pp.196–206.
- Larose, D.T., 2005. *Discovering Knowledge in Data an introduction to data mining*.
- 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.
- Patil, Tina R. & Sherekar, S., 2013. Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification. *International Journal Of Computer Science And Applications Vol. 6, No.2, Apr 2013*
- 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.