

# PREDIKSI KEPUTUSAN KLIEN *TELEMARKETING* UNTUK DEPOSITO PADA BANK MENGGUNAKAN ALGORITMA *NAIVE BAYES* BERBASIS *BACKWARD ELIMINATION*

**Ruhmi Sulaehani**

ruhmisulaehani@yahoo.co.id  
Universitas Ichsan Gorontalo

## Abstrak

*Telemarketing* merupakan salah satu promosi yang dianggap paling efektif dalam mempromosikan produk, strategi pemasaran ini dilakukan oleh bank-bank untuk menawarkan produk pada klien, salah satu produk yang ditawarkan bank yaitu deposito berjangka. Sulitnya mengetahui keputusan klien *Telemarketing* untuk melakukan deposito berjangka pada bank, menyebabkan bank selalu menghadapi ancaman krisis keuangan. Oleh karena itu, *Telemarketing* bank harus dapat membuat target klien, klien mana yang berpotensi untuk melakukan deposito dengan melihat data-data klien yang ada. Dalam penelitian ini akan digunakan algoritma *Naive Bayes* untuk memprediksi keputusan klien *Telemarketing* dengan menggunakan dataset gudang data UCI Repository. Hasil pengujian menunjukkan bahwa nilai akurasi *Naive Bayes* sebesar 89,08%, setelah dilakukan pemilihan fitur dengan menggunakan *Backward Elimination* didapatkan hasil akurasi yang lebih tinggi yaitu sebesar 90,69%, dengan melihat nilai akurasi maka algoritma *Naive Bayes* berbasis *Backward Elimination* meningkatkan akurasi untuk memprediksi keputusan klien *Telemarketing*.

**Kata kunci :** *Telemarketing*, Deposito, *Naive Bayes*, *Backward Elimination*.



ILKOM Jurnal Ilmiah work is licensed under a CCA-SA 4.0 International License.

## 1. Pendahuluan

Bank merupakan badan usaha yang menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkannya kepada masyarakat dalam bentuk kredit dan atau bentuk-bentuk lainnya dalam rangka meningkatkan taraf hidup rakyat banyak [1].

Dengan persaingan antar bank dapat dilihat berbagai upaya bank dalam merebut nasabah sebanyak-banyaknya dalam berbagai kegiatan pemasaran. Jika dulu para pelaku usaha menawarkan barang dan jasa mereka kepada calon konsumen dengan cara bertatap muka langsung, sekarang dengan pemanfaatan teknologi yang canggih bisa menggunakan alat komunikasi jarak jauh seperti telepon dan fax serta media elektronik lainnya [2]. Strategi pemasaran ini dilakukan juga oleh bank-bank untuk menawarkan produk pada konsumen. Biasanya bank melakukan media promosi melalui brosur dan pegawai datang ke klien langsung. Namun hal ini kurang efektif jadi bank melakukan strategi *Telemarketing*. Strategi ini sangat membantu para pemasar maupun pelaku usaha untuk mengenalkan produk dan jasa mereka kepada masyarakat luas dengan waktu yang singkat.

Sulitnya mengetahui keputusan klien *Telemarketing* untuk melakukan deposito pada bank, menyebabkan bank selalu menghadapi ancaman krisis keuangan. Oleh karena itu bank ditekan untuk meningkatkan persyaratan modal dengan menawarkan deposito jangka panjang pada masyarakat. Salah satu cara yaitu menawarkan deposito secara langsung melalui panggilan telepon *Telemarketing* [3]. Seorang *Telemarketing* bank harus dapat membuat target klien, klien mana yang berpotensi untuk melakukan deposito dengan melihat data-data klien yang ada. Dalam mendukung *Telemarketing* untuk meningkatkan tingkat keberhasilan, dilakukan prediksi keputusan klien untuk meningkatkan keberhasilan *Telemarketing* untuk mengurangi biaya pengeluaran dalam melakukan panggilan telephone.

Menurut Larose [4] pendekatan Bayesian digunakan untuk menentukan kemungkinan terhadap asumsi disekitarnya. *Naive Bayes* bekerja sangat efektif saat diuji pada dataset yang besar, terutama bila dikombinasikan dengan beberapa prosedur seleksi atribut [5]. *Backward Elimination* digunakan sebagai seleksi fitur, dibandingkan dengan forward elimination, *Backward Elimination* membuang atribut-atribut yang tidak independen terhadap data yang diolah sedangkan forward elimination mempertahankan semua atribut baik yang independen maupun tidak independen sehingga atribut yang

tidak independen akan menjadi tidak signifikan dan mengurangi jumlah kuadrat residualnya [6] Tingkat akurasi dengan menggunakan seleksi fitur lebih baik daripada tidak menggunakan seleksi fitur, ini menunjukkan bahwa keberadaan fitur mempengaruhi hasil prediksi. Metode *Backward Elimination* memberikan kinerja yang lebih baik jika dibandingkan dengan metode statistik signifikan dalam tahap seleksi. Kinerja terbaik dibuktikan dengan *sensitivitas*, *spesifisitas* dan akurasi yang tinggi [7].

Untuk mengatasi masalah dalam mengambil keputusan klien mana yang akan dihubungi yang lebih mungkin untuk berminat deposito, penelitian ini menggunakan algoritma *Naive Bayes*. Pada penelitian sebelumnya, secara mandiri tingkat akurasi yang dihasilkan algoritma *Naive Bayes* rendah, tetapi dengan menambahkan seleksi fitur *Backward Elimination* diharapkan dapat meningkatkan akurasi. Rumusan masalah pada penelitian ini adalah 1) Sulitnya mengetahui keputusan klien *Telemarketing* untuk melakukan deposito pada bank yang menyebabkan bank selalu menghadapi ancaman krisis keuangan. 2) Secara mandiri tingkat akurasi yang dihasilkan algoritma *Naive Bayes* rendah, dengan menambahkan seleksi fitur *Backward Elimination* diharapkan dapat meningkatkan akurasi dalam memprediksi keputusan klien *Telemarketing*. Tujuan penelitian adalah Untuk memprediksi keputusan klien *Telemarketing* pada panggilan berikutnya agar dapat memudahkan bank dalam mengambil keputusan klien mana yang berpotensi untuk melakukan deposito berjangka dengan menerapkan algoritma *Naive Bayes* dan seleksi fitur *Backward Elimination*. Tujuan berikutnya Membandingkan algoritma *Naive Bayes* dengan algoritma *Naive Bayes* berbasis *Backward Elimination* untuk mendapatkan akurasi yang tinggi dalam memprediksi keputusan klien.

## 2. Landasan Teori

### 2.1. Penelitian Terkait

Moro. S *et.al* [8] membuat model keberhasilan berlangganan deposito jangka panjang, dengan menggunakan atribut yang dikenal sebelum melakukan panggilan. Dalam penelitian ini, penggunaan sistem pendukung keputusan di dasarkan pada model data-driven. Data yang dikumpulkan dari tahun 2008 sampai 2013, sehingga termasuk efek dari krisis keuangan. Peneliti menganalisis 150 fitur terkait dengan nasabah bank, produk, dan atribut sosial ekonomi. Dalam tahap pemodelan, fitur dipilih dan otomatis di eksplorasi, ini dilakukan pada data sebelum juli 2012 dan terjadi pengurangan dan dipilih 22 fitur. Empat model data mining dibandingkan : *Logistic Regresi*, *Decision Trees*, *Neural Network*, dan *Support Vector Machine*, dengan menggunakan dua metrik yaitu *AUC* dan *LIFT*. Empat model tersebut diuji pada set evaluasi dengan menggunakan data terbaru (setelah juli 2012). Dari hasil evaluasi *Neural Network* memperoleh hasil terbaik *AUC* 0,8 dan *LIFT* 0,7, dengan tingkat keberhasilan yang diperoleh 79% dari sukses menjual yang dapat dicapai dengan hanya menghubungi setengah dari klien, yang berarti terjadi peningkatan 29% bila dibandingkan dengan praktek perbankan saat ini yang menghubungi semua klien.

Dalam penelitian Masud. K & Rahman. R. M [9] mengungkapkan jika teknik data mining dapat membantu lembaga untuk menetapkan tujuan pemasaran. Teknik data mining memiliki prospek yang baik untuk target dan meningkatkan kemungkinan respon. Peneliti menggunakan dua teknik data mining yaitu *Naive Bayes* dan algoritma *Decission tree C4.5*. Tujuan penelitian ini yaitu untuk memprediksi apakah klien akan berlangganan deposito berjangka atau tidak dan membuat studi perbandingan kinerja kedua algoritma. Set pengujian data berisi 10% dari seluruh data (4521 *record*) dipilih secara acak dari kumpulan data secara keseluruhan, data diambil dari UCI Repository. Tingkat akurasi yang dihasilkan *C4.5* adalah 93,39% sedangkan *Naive Bayes* 84,91%. Keakuratan dari kedua model tinggi, algoritma *C4.5* membuktikan lebih baik dari *Naive Bayes*. Tetapi dari segi waktu untuk membangun model *Naive Bayes* lebih cepat dibanding *C4.5*.

## 2.2. Landasan Teori

### 2.2.1 Deposito

Deposito berjangka merupakan jenis simpanan yang dikeluarkan oleh bank, yang berbeda dengan jenis simpanan giro dan tabungan dimana simpanan deposito mengandung unsur jangka waktu (jatuh tempo) lebih panjang dan tidak dapat ditarik sewaktu waktu [10]. Bank-bank sepenuhnya didanai oleh deposito, persaingan bank yang ketat akan meningkatkan dana perbankan dengan menaikkan suku bunga deposito, suku bunga deposito sebagai ukuran dari intensitas deposito pada persaingan pasar. Dalam hal ini, bank dapat memanfaatkan depositan dengan menginvestasikan kekayaan aset pada perusahaan [11].

### 2.2.2 Telemarketing

*Telemarketing* merupakan pendekatan target via telepon dengan menjelaskan produk dan menggunakan alat komunikasi yakni telepon [12]. *Telemarketing* dapat didefinisikan sebagai strategi

pemasaran yang dilakukan melalui komunikasi jarak jauh, seperti telepon, untuk menargetkan klien yang memungkinkan untuk memperoleh produk atau jasa yang ditawarkan [8][3]. *Telemarketing* mengacu pada proses menggunakan telepon sebagai media interaktif untuk promosi penjualan. *Telemarketing* memegang banyak potensi untuk menjangkau pelanggan sebanyak mungkin. Jika digunakan secara efektif, *Telemarketing* dapat menjadi alat yang ampuh untuk ekspansi bisnis [13]. Dalam *Telemarketing*, telepon menunjukkan fungsinya dalam bentuk [14] :

1. Menghubungi langsung pembeli potensial,
2. Menjadi semacam ujung tombak survei pemasaran, untuk di follow up lebih lanjut oleh tenaga pemasaran di lapangan, dan
3. Menjadi alat pembuka pintu pembeli potensial agar bisa dikunjungi oleh tenaga pemasaran.

#### 2.2.4 Metode Naive Bayes

*Naive Bayes* merupakan teknik sederhana yang sangat cocok saat dimensi dari input tinggi. Meskipun sederhana, *Naive Bayes* seringkali dapat mengungguli metode lain yang lebih canggih. Untuk membangun model *Naive Bayes*, perlu menghitung probabilitas setiap nilai atribut yang diberi atribut class.

Persamaan dari teorema Bayes adalah [15]:

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

Keterangan:

$X$  : Data dengan class yang belum diketahui

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

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

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

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

$P(X)$  : Probabilitas  $X$

Untuk membangun model *Naive Bayes*, perlu menghitung probabilitas setiap nilai atribut yang diberi atribut class. Oleh karena itu, jika menggunakan atribut dengan jumlah banyak nilai diskrit mungkin memerlukan waktu komputasi yang lebih lama [16].

#### 2.2.5 Backward Elimination

*Backward Elimination* digunakan untuk memilih yang terbaik diantara semua kemungkinan untuk menghapus dan menganalisis statistik yang tidak sesuai. Metode *Backward Elimination* memberikan kinerja yang lebih baik jika dibandingkan dengan metode statistik signifikan dalam tahap seleksi. Kinerja terbaik dibuktikan dengan *sensitivitas*, *spesifisitas*, dan *akurasi* yang tinggi [7].

*Backward Elimination* merupakan metode yang dapat menghilangkan atribut yang tidak signifikan dari model [17]. Untuk prosedur *Backward Elimination*, model dimulai dengan semua atribut yang ada di dalamnya, dan atribut dengan persial statistik terkecil dihapus.

Atribut yang digunakan akan diimplementasikan dengan *Backward Elimination*, sebelumnya data yang digunakan dijadikan data numerik dan diregresikan, sebelum mendapatkan nilai  $F$  terlebih dahulu mencari nilai prediksi, rata-rata prediksi, MSR dan MSE, rumus yang digunakan yaitu :

Untuk mencari nilai prediksi dari masing-masing atribut.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i \quad (2)$$

Keterangan :

$y$  : Atribut terkait,

$x$  : Atribut bebas,

$\beta$  : Parameter regresi,

$\varepsilon$  : Nilai kesalahan / standar error estimasi

Mencari masing-masing nilai MSR dan MSE tiap atribut dengan rumus:

$$MSR = \sum (\hat{y}_t - \bar{y})^2 / p - 1 \quad (3)$$

$$MSE = \sum (y_t - \hat{y}_t)^2 / n - p \quad (4)$$

Keterangan :

$\hat{y}_t$  : Nilai prediksi,

$\bar{y}$  : Nilai rata-rata prediksi,

$p$  : Jumlah atribut,

$y_t$  : Nilai aktual,

$n$  : Jumlah record data

Mencari nilai  $F$  masing-masing atribut dengan rumus:

$$F_k^* = MSR(X_k) / MSE(X_k) \quad (5)$$

Prosedur *Backward Elimination* [18] :

1. Gunakan semua atribut yang tersedia. Sebagai contoh, model memiliki empat atribut,  $x_1, x_2, x_3, x_4$ .
2. Untuk setiap atribut dalam model, menghitung parsial F-statistik. Pada saat pertama melalui algoritma, akan menjadi  $F(x_1 | x_2, x_3, x_4)$ ,  $F(x_2 | x_1, x_3, x_4)$ ,  $F(x_3 | x_1, x_2, x_4)$ , dan  $F(x_4 | x_1, x_2, x_3)$ . Pilih atribut dengan parsial F-statistik terkecil. Menunjukkan nilai  $F_{min}$
3. Untuk uji pentingnya  $F_{min}$ . Jika  $F_{min}$  tidak signifikan, atribut menghapus yang terkait dengan  $F_{min}$  dari model, dan kembali kelangkah 2. Jika  $F_{min}$  signifikan, algoritma berhenti dan melaporkan model saat ini.

*K-Fold Cross Validation* merupakan teknik validasi yang membagi data ke dalam  $k$  bagian dan kemudian masing-masing bagian akan dilakukan proses klasifikasi. Dengan menggunakan *K-Fold Cross Validation* akan dilakukan percobaan sebanyak  $k$ . Tiap percobaan akan menggunakan satu data testing dan  $k-1$  bagian akan menjadi data training, kemudian data testing itu akan ditukar dengan satu buah data training sehingga untuk tiap percobaan akan didapatkan data testing yang berbeda-beda. Data training merupakan data yang akan dipakai dalam melakukan pembelajaran sedangkan data testing merupakan data yang belum pernah dipakai sebagai pembelajaran dan akan berfungsi sebagai data pengujian kebenaran atau keakurasian hasil pembelajaran. Dalam penelitian ini nilai  $k$  yang digunakan berjumlah 10 atau *10Fold Cross Validation* [20].

Dalam penelitian ini dipilih alat ukur evaluasi berupa *Confusion Matrix* dengan tujuan untuk mempermudah dalam menganalisis performa algoritma karena *Confusion Matrix* memberikan informasi dalam bentuk angka sehingga dapat dihitung rasio keberhasilan. Evaluasi model didasarkan pada pengujian untuk memperkirakan obyek yang benar dan salah [19], *Confusion Matrix* adalah salah satu alat ukur berbentuk matrik 2x2 yang digunakan untuk mendapatkan jumlah ketepatan klasifikasi dataset terhadap kelas aktif dan tidak aktif pada kedua algoritma yang dipakai.

Tabel 1. *Confusion Matrix*

CLASSIFICATION	Predicted Class		
		Class = Yes	Class = No
OBERVED CLASS	Class = Yes	A (True Positive-TP)	B (False Negative-FN)
	Class = No	C (False Positive-FP)	D (True negative-TN)

Keterangan :

- True Positive (tp)* = Proporsi sampel bernilai *true* yang diprediksi secara benar  
*True positive (tp)* = Proporsi sampel bernilai *false* yang diprediksi secara benar  
*False Positive (fp)* = Proporsi sampel bernilai *false* yang salah diprediksi sebagai sampel bernilai *true*.  
*False Negative (fn)* = Proporsi sampel bernilai *true* yang salah diprediksi sebagai sampel bernilai *true*.

Untuk menghitung nilai *accuracy*, *precision*, dan *recall* dengan rumus perhitungan sebagai berikut :

$$Accuracy = \left( \frac{a+d}{a+b+c+d} \right) = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

*ROC (Receiver Operating Characteristic)* merupakan grafik yang dapat digunakan untuk menilai model. Untuk klasifikasi data mining, nilai *AUC* dapat dibagi menjadi beberapa kelompok

- 0.90 – 1.00 = *Excellent classification*
- 0.80 – 0.90 = *Good classification*
- 0.70 – 0.80 = *Fair classification*
- 0.60 – 0.70 = *Poor classification*
- 0.50 – 0.60 = *Failure*

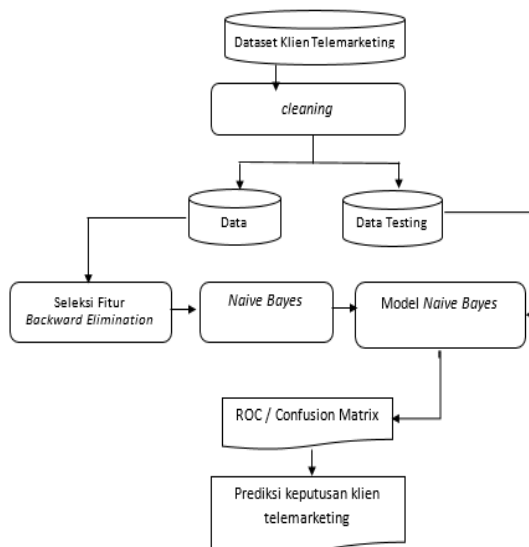
### 3. Metode

Data yang digunakan pada penelitian ini adalah dataset global, dataset ini dapat diunduh secara gratis diwebsite : <http://archive.ics.uci.edu/ml/datasets/%28UPDATE%29+Bank+Marketing> di publikasi pada bulan mey 2014. Atribut-atribut yang digunakan yaitu *age, job, marital, education, default, housing, loan, contact, month, day of week, duration, campaign, previous, poutcome* dan  $y$  sebagai labelnya.

Dataset yang digunakan berkaitan dengan semua aspek dari klien *Telemarketing*, termasuk informasi pribadi, informasi kontak dan informasi sosial dan ekonomi. Pada penelitian ini atribut yang berkaitan dengan informasi sosial ekonomi tidak digunakan, dalam penelitian Valle. M. V *et.al* [20] mengatakan bahwa atribut sosial ekonomi tidak mampu untuk membedakan secara efektif atribut kelas karena atribut ini bukan merupakan indikasi kerja masa depan.

Beberapa tahapan dalam pengolahan eksperimen ini, dengan tahapan sebagai berikut:

- *Praprocessing* : Untuk meningkatkan keakurasian data maka pada tahap ini akan dilakukan reduction dan cleaning data.
- Desain Model : Tahap pertama menghitung jumlah class / label, tahap kedua menghitung jumlah kasus yang sama dengan class yang sama, tahap ketiga hitung probabilitas untuk atribut  $y = \text{yes}$  dan  $y = \text{no}$  dan tahap keempat membandingkan hasil class *yes* dan *no* untuk berlangganan deposito.
- Evaluasi : Proses eksperimen yang dilakukan oleh peneliti yaitu dengan menggunakan tools Rapid Miner dengan tahapan-tahapan seperti gambar berikut



Gambar 1. Desain eksperimen

- *Prediksi Keputusan Klien Telemarketing* : *Prediksi keputusan klien Telemarketing* untuk deposito berjangka penting dilakukan oleh suatu bank untuk mengetahui apakah *Telemarketing* suatu bank dapat mencapai tingkat keberhasilan yang ditargetkan, tujuannya untuk meningkatkan persyaratan modal agar suatu bank tidak menghadapi ancaman krisis keuangan apabila dapat diprediksi dan klien melakukan deposito. Melakukan prediksi terhadap keputusan klien menjadi penting karena menjadi salah satu landasan kebijakan masalah. Algoritma *Naive Bayes* berbasis *Backward Elimination* dipilih untuk memperoleh akurasi dalam memprediksi keputusan klien untuk melakukan deposito. Peneliti akan menggunakan data sebanyak 411 *record* atau 10% dari keseluruhan data untuk mendapatkan contoh hasil prediksi.

## 4. Hasil

### 4.1. Evaluasi Menggunakan Algoritma *Naive Bayes* Berbasis *Backward Elimination*

Pada penelitian ini, data yang digunakan adalah data bersih yang telah melalui preprocessing. Pengujian menggunakan model 10 *fold cross validation* dengan menggunakan fitur seleksi, yang akan secara acak mengambil 10% dari data training untuk data testing, proses ini diulang sebanyak 10 kali dan hasil pengujian model berupa *accuracy*, *precision*, dan *recall*. Seleksi fitur dengan menggunakan *Backward Elimination* dilakukan untuk menentukan atribut-atribut mana yang berpengaruh terhadap dataset. Proses seleksi fitur dilakukan untuk menghasilkan atribut-atribut yang sangat bermutu atau memiliki weight yang paling berpengaruh

Atribut yang digunakan yaitu *age*, *job*, *marital*, *education*, *default*, *housing*, *loan*, *contact*, *month*, *day of week*, *duration*, *campaign*, *previous*, *poutcome*, dan 1 atribut sebagai label yaitu *y*. Dari hasil eksperimen dengan menggunakan algoritma *Naive Bayes* berbasis *Backward Elimination* diperoleh hasil atribut weight yang berpengaruh seperti *Duration Campaign*, *Contact* dan *Poutcome* yang dianggap mempengaruhi tingkat akurasi dari 14 atribut. Pengolahan dengan menggunakan atribut tersebut dapat memudahkan pemrosesan dalam menentukan keputusan klien untuk berlangganan deposito atau tidak, karena ada 1 atribut sebagai label yang menentukan *yes* or *no*. Berikut contoh hasil data setelah seleksi atribut dengan menggunakan *Backward Elimination* :

Tabel 1. Contoh Hasil Data

Duration	Campaign	contact	poutcome	y
range3 [402 - 681]	range1 [-∞ - 4]	cellular	nonexistent	no

range2 [147 - 402]	range1 [-∞ - 4]	telephone	nonexistent	no
range2 [147 - 402]	range1 [-∞ - 4]	telephone	nonexistent	no
range1 [-∞ - 147]	range1 [-∞ - 4]	telephone	nonexistent	no
range1 [-∞ - 147]	range1 [-∞ - 4]	cellular	nonexistent	no

Hasil akurasi yang telah didapatkan pada data set klien *Telemarketing* dengan menggunakan algoritma *Naive Bayes* berbasis *Backward Elimination* ditampilkan pada tabel 4.3 yang menunjukkan tingkat akurasi sebesar 90,69%.

Tabel 2. Hasil Confusion Matrix Menggunakan Algoritma *Naive Bayes* berbasis *Backward Elimination*

	true no	true yes	class precision
pred. No	3264	273	92,28%
pred.yes	79	166	67,76%
class recall	97,64%	37,81%	

Eksperimen ini menggunakan data sebanyak 3783 *record*. Berdasarkan *confusion matrix* terlihat bahwa 3264 *record* diprediksi *no* sebagai kelompok data *no* dan sebanyak 273 *record* diprediksi *no* sebagai kelompok data *yes*. Selanjutnya terlihat bahwa 166 *record* diprediksi *yes* sebagai kelompok data *yes* dan sebanyak 79 *record* di diprediksi *yes* sebagai kelompok data *no*.

*Precision* adalah proporsi data yang benar-benar kelas "no" diantara data yang diklasifikasikan sebagai class "no":

Precision : Proporsi jumlah sampel bernilai true yang berhasil diprediksi secara tepat.

$$Precision = \frac{TP}{TP+FP} = \frac{3264}{3264+273} \times 100\% = 92,28\%$$

Recall : Proporsi sampel bernilai true yang diprediksi secara benar.

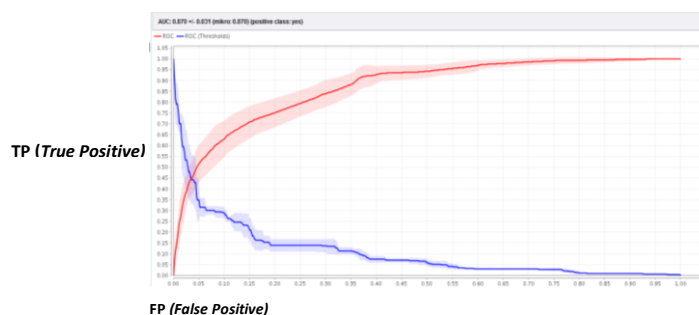
$$Recall = \frac{TP}{TP+FN} = \frac{3264}{3264+79} \times 100 = 97,636 \%$$

Dari hasil ini, dapat dihitung nilai akurasinya.

$$Akurasi = \frac{\text{banyaknya prediksi yang benar}}{\text{total banyaknya prediksi}} \times 100\% \\ = \frac{3264 + 166}{3264 + 273 + 79 + 166} \times 100\% = 90,69\%$$

Dapat disimpulkan bahwa perhitungan persentase tingkat akurasi pada *confusion matrix* mencapai nilai persentase sebesar 90,69%, sehingga dataset klien *Telemarketing* tersebut dinyatakan akurat menggunakan metode *Naive Bayes* berbasis *Backward Elimination*.

Hasil pengujian ini telah menghasilkan nilai *accuracy*, *precision* dan *recall* disertai dengan *ROC* (*Receiver Operating Characteristic*) curve. *AUC* (*Area Under the ROC Curve*) merupakan grafik yang dapat digunakan untuk menilai model. Karena *AUC* adalah bagian dari daerah unit persegi, nilainya akan selalu antara 0,0 dan 1,0. Pada gambar 4.1 garis berwarna merah merupakan kurva *ROC* dengan sebesar 0,870 termasuk *Good Classification*, sedangkan garis berwarna biru merupakan kurva ambang (*thresholds*).



Gambar 2. Kurva ROC

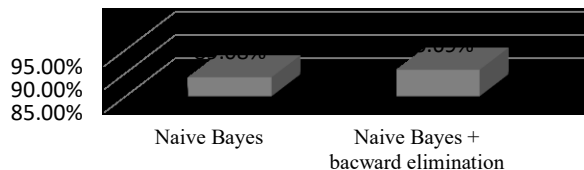
Berdasarkan hasil dengan *Confusion Matrix* dan kurva *ROC* menunjukkan *rule* hasil klasifikasi untuk memprediksi keputusan klien termasuk *Good classification* sehingga dapat digunakan untuk memprediksi keputusan klien *Telemarketing*.

#### 4.2. Perbandingan Hasil Metode *Naive Bayes* dan *Naive Bayes* berbasis *Backward Elimination*

Berdasarkan hasil analisis dapat dibandingkan metode algoritma *Naive Bayes* berbasis *Backward Elimination* lebih baik daripada hanya menggunakan metode *Naive Bayes* saja, hasil

accuracy yang didapatkan dengan menggunakan *Naive Bayes* yaitu 89,08% sedangkan *Naive Bayes* berbasis *Backward Elimination* 90,69%, dapat disimpulkan bahwa metode atau algoritma yang paling akurat adalah metode *Naive Bayes* berbasis *Backward Elimination* tingkat accuracy yang dihasilkan tinggi, dapat di lihat perbandingannya pada gambar 3.

Perbandingan Nilai Akurasi *Naive Bayes* dan *Naive Bayes + Backward Elimination*



Gambar 3. Perbandingan *Naive Bayes* dan *Naive Bayes + Backward Elimination*

#### 4.2 Prediksi Keputusan Klien dengan Menggunakan Algoritma *Naive Bayes* berbasis *Backward Elimination*

Prediksi keputusan klien *Telemarketing* untuk deposito penting dilakukan oleh suatu bank untuk mengetahui apakah *Telemarketing* suatu bank dapat mencapai tingkat keberhasilan yang ditargetkan, tujuannya untuk meningkatkan persyaratan modal agar suatu bank tidak menghadapi ancaman krisis keuangan, apabila dapat diprediksi dan klien melakukan deposito. Melakukan prediksi terhadap keputusan klien menjadi penting karena menjadi salah satu landasan kebijakan masalah. Algoritma *Naive Bayes* berbasis *Backward Elimination* dipilih untuk memperoleh akurasi dalam memprediksi keputusan klien untuk melakukan deposito.

Berdasarkan evaluasi model yang telah dilakukan dapat diketahui bahwa algoritma *Naive Bayes* berbasis *Backward Elimination* dapat diterapkan berdasarkan atribut klien *Telemarketing*, serta mampu secara optimal untuk memprediksi keputusan klien.

Tabel 4.4 merupakan contoh data klien *Telemarketing* yang keputusannya untuk berlangganan belum diketahui, dapat dilihat pada atribut *y* yang prediksi keputusannya belum ada. Peneliti menggunakan data sebanyak 411 *record* atau 10% dari keseluruhan data untuk mendapatkan contoh hasil prediksi.

Tabel 3. Data Tes untuk Prediksi

age	job	marital	Education	default	housing	Loan	Contact	Mont h	day_of_week	durati on	camp aign	previo us	poutcome	y
49	admin	divorced	high.school	no	no	no	cellular	Jul	Thu	602	1	0	nonexistent	
51	admin	divorced	university.degree	unknown	no	no	cellular	Aug	Fri	59	2	0	nonexistent	
34	admin	marrid	university.degree	no	yes	no	cellular	may	mon	351	1	0	nonexistent	
32	admin	marrid	basic.9y	no	no	no	cellular	Jul	mon	77	2	0	nonexistent	

Dengan menggunakan metode *Naive Bayes* berbasis *Backward Elimination* maka dapat diprediksi keputusan klien pada panggilan berikutnya. Pada Tabel 4.5 dapat dilihat pada kolom *y* prediksi keputusan belum ada, setelah diprediksi keputusan yang dihasilkan dapat dilihat pada kolom *prediction(y)*, ini gambaran dari hasil prediksi yang dilakukan dengan menggunakan algoritma *Naive Bayes* berbasis *Backward Elimination*, contoh data hasil prediksi dapat dilihat pada Tabel 4.

Tabel 4. Hasil Prediksi Menggunakan Algoritma *Naive Bayes* Berbasis *Backward Elimination*

y	Prediction(y)	age	job	marital	education	default	housing	loan	contact	month	day_of_week	duration	campaign	previous	poutcome
	no	49	admin.	divorced	high.school	no	no	no	cellular	jul	thu	602.0	1.0	0.0	nonexistent
	no	51	admin.	divorced	university.degree	unknown	no	no	cellular	aug	fri	59.0	2.0	0.0	nonexistent
	no	34	admin.	married	university.degree	no	yes	no	cellular	may	mon	351.0	1.0	0.0	nonexistent
	no	32	admin.	married	basic.9y	no	no	no	cellular	jul	mon	77.0	2.0	0.0	nonexistent

### 5. Kesimpulan dan Saran

#### 5.1 Kesimpulan

1. Algoritma *Naive Bayes* berbasis *Backward Elimination* dapat digunakan untuk memprediksi keputusan klien *Telemarketing* dengan melihat data dari klien sebelumnya, dalam prediksi

keputusan klien untuk berlangganan deposito dapat dilihat dari nilai perhitungan probabilitas tertinggi pada label.

2. Dalam pemilihan atribut yang berpengaruh dilihat dari nilai  $F$  terkecil dan nilai *signifikan*, jika nilai  $F$  terkecil dari suatu atribut memiliki nilai *signifikan* dari nilai  $\alpha \leq 0,05$  maka atribut tersebut berpengaruh, dari 14 atribut yang digunakan yaitu *age, duration, campaign, previous, job, marital, education, default, housing, loan, contact, month, day\_of\_week* dan *poutcome* hanya 4 atribut yang berpengaruh dari hasil evaluasi yaitu *duration, campaign, contact* dan *poutcome*. Hasil evaluasi *confusion matrix* dengan menggunakan algoritma *Naive Bayes* berbasis *Backward Elimination* di dapatkan akurasi yang tinggi yaitu 90,69% dibanding dengan hanya menggunakan algoritma *Naive Bayes* saja yaitu 89,08%.

## 5.2 Saran

1. Menggunakan dataset lain sebagai perbandingan tingkat akurasi yang dihasilkan *Naive Bayes*.
2. Untuk meningkatkan akurasi prediksi keputusan klien *Telemarketing* dapat dilakukan penggabungan beberapa algoritma dan dapat juga menggunakan seleksi fitur yang lain.

## Daftar Pustaka

- [1] Undang-undang Nomor 10 tahun, 1998.
- [2] Kotler. P and Keller. K. L, "Framework for Marketing Management", 2012.
- [3] Laureano. M. S. R., Moro. S and Paulo. C, "A Data Mining Approach For Bank Telemarketing Using the rminer Package and R Tool.
- [4] Larose. D. T, Data Mining Methods and Models, Canada: Simultaneously in Canada, 2007.
- [5] Witten. I. H, E. Frank and M. A. Hall " Data Mining: Practical Machine Learning Tools and Techniques," Published is an Imprint of Elsevier, 2011.
- [6] N. Sunengsih, " Seleksi Variabel Dalam Analisis Regresi Multivariat Multipel," Staf Jurusan Statistika FMIPA UNPAD, Seminar Nasional Matematika dan Pendidikan Matematika, 5 Desember 2009
- [7] Ali. N, Icle and. Y, Ozer. M, "Investigating The Performance Improvement Of HRV Indices in CHF Using Feature Selection Methods Based On Backward Elimination and Statistical Significance," Computer in Biology and Medicine, 2014.
- [8] Moro. S, Paulo. C and Rita. P, "A Data-driven Approach To Predict The Success of Bank Telemarketing," Decision Support System, 2014.
- [9] Masud. K and Rahman. R. M, "Decision Tree and Naive Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing," Journal of Software Engineering and Applications, February 2013.
- [10] Hamonangan. T. A, " Deposito Berjangkan Sebagai Jaminan Kredit Pada PT. Bank Danamon Indonesia, Tbk Cabang Pemuda Semarang," M.K thesis, Universitas Diponegoro, 2009.
- [11] Mastura Abdul Karim, M. Kabir Hassan, Taufiq Hassan & Shamsheer Mohammad, " Capital Adequacy and Lending and Deposit Behaviors of Conventional and Islamic Banks," Pacific Basin Finance Journal, 2013.
- [12] Artikel Strategi Pemasaran menggunakan Telemarketing, "<http://www.kolombloggratis.org/2013/05/strategi-pemasaran-Telemarketing.html>" ,
- [13] Artikel dan Kumpulan Tips, " <http://www.artikelcara.com/3189/tips-teknik-efektif-dalam-Telemarketing/> "
- [14] Artikel Marketing, "<http://rajapresentasi.com/2009/05/Telemarketing-untuk-menambah-pelanggan-baru/> ".
- [15] Bustami, " Penerapan Algoritma Naive Bayes Untuk Mengklasifikasikan Data Nasabah Asuransi," Jurnal Teknik Informatika, 2012.
- [16] Aggarwal. C & Philip S.Yu, " Privacy-Preserving Data mining : Models and Algorithms," Springer, 2008.
- [17] Chana. K. Y, Kwong. C. K, Dillona. T. S and Tsimb. Y. C, " Reducing Overfitting in Manufacturing Process Modeling Using a Backward Elimination Based Genetic Programming," Applied Soft Computing, 2011
- [18] Larose. D. T, " Data Mining : Methods and Models," Published simultaneously, 2006.
- [19] Gonescu. F, " Data Mining : Concepts, Method and Techniques", Spinger, 2011
- [20] Valle. M. V, Varas. S & Ruz. G. A, "Job Performance Prediction in a Call Center Using a Naive Bayes Classifier," Expert System With Applications, 2012.