ISSN print 2087-1716
ISSN online 2548-7779

ILKOM Jurnal Ilmiah Volume 9 Nomor 1 April 2017

KLASIFIKASI NASABAH ASURANSI JIWA MENGGUNAKAN ALGORITMA NAIVE BAYES BERBASIS BACKWARD ELIMINATION

Betrisandi
betris.sin@gmail.com
Universitas Ichsan Gorontalo

Abstrak

Pendapatan untuk perusahaan asuransi ditentukan oleh jumlah premi yang dibayar oleh nasabah. Banyaknya nasabah yang tidak lancar membayar premi berpengaruh terhadap kinerja serta eksistensi perusahaan sehari-hari. Algoritma *Naive Bayes* berbasis *Backward Elimination* bertujuan untuk melakukan klasifikasi nasabah asuransi dengan hasil akurasi 85,89 % dengan delapan atribut *weight* yaitu umur, jangka waktu, cara bayar, premi, jumlah hari, pekerjaan, penghasilan dan mata uang.
Kata kunci : Asuransi, *Naive Bayes, Backward Elimination*.

1. Pendahuluan

Asuransi atau pertanggungan adalah perjanjian antara dua pihak atau lebih, dengan mana pihak penanggung mengikatkan diri kepada tertanggung, dengan menerima premi asuransi, untuk memberikan penggantian kepada tertanggung karena kerugian, kerusakan atau kehilangan keuntungan yang diharapkan, atau tanggung jawab hukum kepada pihak ketiga yang mungkin akan diderita tertanggung, yang timbul dari suatu peristiwa yang tidak pasti, atau untuk memberikan suatu pembayaran yang didasarkan atas meninggal atau hidupnya seseorang yang dipertanggungkan.

Premi merupakan pendapatan bagi perusahaan asuransi yang jumlahnya ditentukan dalam suatu persentase atau tarif tertentu dari jumlah yang dipertanggungkan. Bagi tertanggung premi merupakan beban karena membayar premi merupakan beban tertanggung. Pendapatan premi untuk perusahaan asuransi ditentukan oleh jumlah premi yang dibayar oleh nasabah.

Permasalahan yang sering timbul adalah banyaknya nasabah yang tidak lancar dalam membayar premi. Lancar tidaknya pembayaran premi nasabah berpengaruh terhadap kinerja serta eksistensinya dalam kehidupan sehari – hari. Diperlukan suatu cara agar dapat mengetahui bagaimana pola nasabah yang akan dikatakan lancar maupun tidak lancar agar dapat membantu perusahaan asuransi dalam mengklasifikasikan nasabahnya sehingga pihak asuransi bisa mengatasi sejak dini permasalahan tersebut.

Untuk mengklasifikasi nasabah asuransi yang lancar dan tidak lancar membayar premi penelitian ini menggunakan algoritma *Naive Bayes* berbasis *Backward Elimination.* Algoritma ini dipilih karena bisa mengolah data dalam jumlah yang besar dan menghasilkan klasifikasi terbaik dengan *Backward Elimination*. Dimana *profile* data nasabah sebagai parametrik untuk pengukuran, agar dapat membantu perusahaan asuransi dalam mengambil keputusan menerima atau menolak calon nasabah asuransi nantinya.

2. Metode

Data nasabah asuransi tersebut terdiri dari variabel umur, pekerjaan, penghasilan, jenis asuransi, jangka waktu, cara bayar, mata uang, premi, jumlah hari dan persentase kelancaran sebagai label.

Jumlah data pada kasus ini ada 1066 record tetapi data tersebut masih mengandung duplikasi dan anomali atau inkonsisten data. Untuk meningkatkan keakurasian data, maka dilakukan *reduction* dan *cleaning* data. Setelah *reduction* dan *cleaning* data sudah dilakukan maka jumlah data berkurang menjadi 1013 record.

Untuk mempermudah dan mempercepat proses perhitungan analisis data serta untuk meningkatkan keakurasian hasil perhitungan maka atribut yang bertipe integer dan numerik akan di *discretize*-kan ke dalam sejumlah kecil dari rentang yang berbeda, data yang numerik dinominalkan untuk mempermudah dalam perhitungan probabilitas *Naive Bayes*.

Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukan oleh ilmuwan Inggris *Thomas Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai *Teorema Bayes*. Teorema tersebut dikombinasikan dengan *Naive* dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi *Naive Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya.[1]

ISSN print 2087-1716
ISSN online 2548-7779

Persamaan dari teorema Bayes adalah [4]:

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

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*

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, β *: Parameter regresi,*
x : Atribut bebas, ε *: Nilai kesalahan / standar error estimasi*

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

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

$F_k*=MSR(X_k) / MSE(X_k)$ (5)

Dalam penelitian ini dipilih alat ukur evaluasi berupa *Confusion Matrix* dengan tujuan untuk mempermudah dalam menganilisis 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 [5], *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**		
OBERVED 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)*

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 ttrue.
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 :

ISSN print 2087-1716
ISSN online 2548-7779

$$Accuracy = \left(\frac{a+d}{a+b+c+d}\right) = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

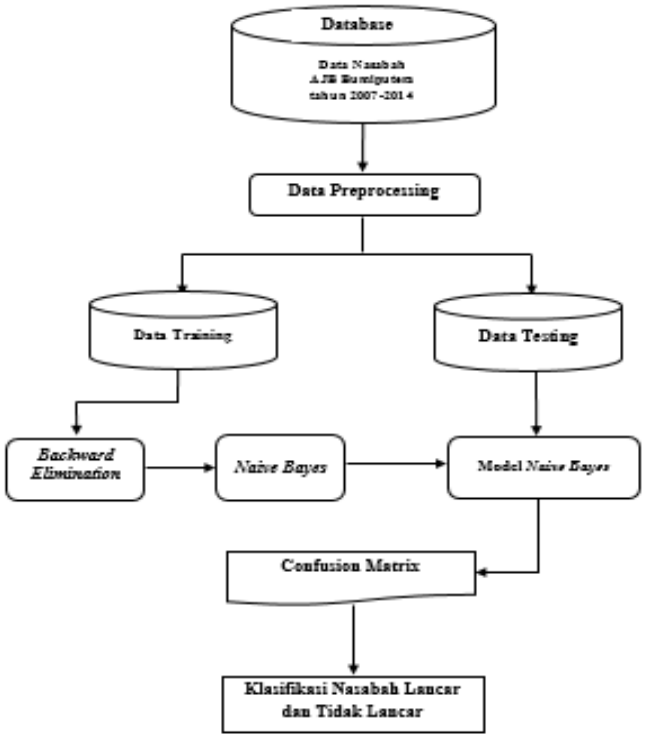
$$Recall = \frac{TP}{TP+FN} \quad (8)$$

Untuk mengukur ketelitian dari suatu model,kita dapat mengukur area di bawah kurva ROC. Akurasi AUC dikatakan sempurna apabila nilai AUC mencapai 1.000, dikatakan baik jika diatas 0.500 dan akurasinya buruk jika nilai AUC dibawah 0.500. [6]

2.1. Tahapan dalam eksperimen

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

- Praprosessing : 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 kalikan semua hasil variabel lancar dan tidak lancar dan tahap keempat membandingkan hasil *class* lancar dan tidak lancar.
- Evaluasi : Proses eksperimen yang dilakukan oleh peneliti yaitu dengan menggunakan tools Rapid Miner dengan tahapan-tahapan seperti gambar berikut :



Gambar 1 Desain eksperimen

3. Hasil dan Pembahasan

3.1 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 umur, pekerjaan, penghasilan, jenis asuransi, jangka waktu, cara bayar, mata uang, premi, jumlah hari.

Dari hasil eksperimen dengan menggunakan algoritma *Naive Bayes* berbasis *Backward Elimination* diperoleh hasil atribut weight yang berpengaruh seperti dalam tabel dibawah ini :

Tabel 2 Attribut Weight

ISSN print 2087-1716
ISSN online 2548-7779

Attribute	Weight
Umur	1
Jangka Waktu	1
Cara Bayar	1
Premi	1
Jumlah Hari	1
Pekerjaan	1
Penghasilan	1
Jenis Asuransi	0
MU	1

Berdasarkan tabel 3.1 pada tahap *Backward Elimination* dari 9 atribut telah terpilih 8 atribut yang dianggap mempengaruhi tingkat akurasi, yaitu umur, jangka waktu, cara bayar, premi, jumlah hari, pekerjaan, penghasilan dan mata uang. Dilihat dari nilai pada tabel yang bernilai 1 hal ini menjelaskan bahwa dalam perhitungan klasifikasi nasabah asuransi jiwa secara optimal melalui 8 atribut yang dipilih melalui *Backward Elimination.* Pengolahan dengan menggunakan atribut tersebut dapat memudahkan pemrosesan dalam menentukan klasifikasi nasabah asuransi jiwa karena ada 1 atribut sebagai label yang menentukan lancar atau tidak lancar. Berikut contoh hasil data setelah seleksi atribut dengan menggunakan *Backward Elimination* :

Tabel 3 Contoh Hasil Data

Umur	Jangka Waktu	Cara Bayar	Premi	Jumlah Hari	Pekerjaan	Penghasilan	MU	Pesrsentase Kelancaran
range2 [35 - ∞]	range1 [-∞ - 15]	range1 [-∞ - 2]	range1 [-∞ - 266526]	range2 [-6 - ∞]	PNS	2,5 - 5 juta	1	Lancar
range2 [35 - ∞]	range1 [-∞ - 15]	range1 [-∞ - 2]	range1 [-∞ - 266526]	range2 [-6 - ∞]	PNS	2,5 - 5 juta	1	Lancar
range2 [35 - ∞]	range1 [-∞ - 15]	range1 [-∞ - 2]	range1 [-∞ - 266526]	range2 [-6 - ∞]	PNS	2,5 - 5 juta	1	tidak lancar
range2 [35 - ∞]	range1 [-∞ - 15]	range1 [-∞ - 2]	range1 [-∞ - 266526]	range1 [-∞ - -6]	PNS	2,5 - 5 juta	1	Lancar

Hasil akurasi yang telah didapatkan pada dataset nasabah asuransi jiwa dengan menggunakan algoritma *Naive Bayes* berbasis *Backward Elimination* ditampilkan pada tabel 3.3 yang menunjukkan tingkat akurasi sebesar 85,89%.

Tabel 4 Hasil Confusion Matrix dengan Menggunakan Algoritma Naive Bayes berbasis Backward Elimination

	true lancar	true tidak lancar	class precision
Pred. Lancar	868	141	86,03%
Pred. Tidak Lancar	2	2	50,00%
Class recall	99,77%	1,40%	

Eksperimen ini menggunakan data sebanyak 1013 record. Berdasarkan confusion matrix terlihat bahwa 868 record diprediksi lancar sebagai kelompok data lancar dan sebanyak 141 record diprediksi lancar sebagai kelompok data tidak lancar. Selanjutnya terlihat bahwa 2 record diprediksi tidak lancar sebagai kelompok data lancar dan sebanyak 2 record di diprediksi tidak lancar sebagai kelompok data tidak lancar.

Precision adalah proporsi data yang benar-benar kelas lancar diantara data yang diklasifikasikan sebagai class lancar :

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

$$Precision = \frac{TP}{TP+FP} = \frac{868}{868+141}\ x\ 100\% = 86{,}03\%$$

Recall : Proporsi sampel bernilai true yang diprediksi secara benar.

ISSN print 2087-1716
ISSN online 2548-7779

$$Recall = \frac{TP}{TP+FN} = \frac{868}{868+2}\ x\ 100 = 99{,}77\ \%$$

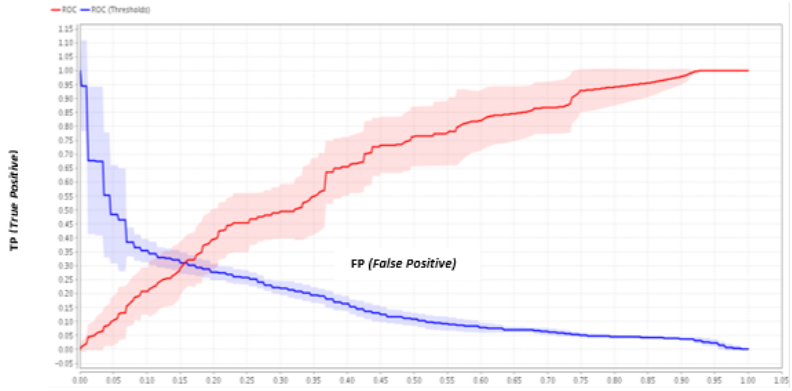
Dari hasil ini, dapat dihitung nilai akurasinya.

$$Akurasi = \frac{banyaknya\ prediksi\ yang\ benar}{total\ banyaknya\ prediksi}\ x\ 100\%$$

$$= \frac{868+2}{868+141+2+2}\ x\ 100\% = 85{,}89\%$$

Dapat disimpulkan bahwa perhitungan persentase tingkat akurasi pada *confusion matrix* mencapai nilai persentase sebesar 85,89%. Sehingga data nasabah asuransi jiwa tersebut dinyatakan akurat menggunakan metode *Naive Bayes* berbasiis *Backward elimination.*

Hasil pengujian ini telah menghasilkan nilai *accuracy*, *precission* 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 3.1 garis berwarna merah merupakan kurva *ROC* dengan sebesar 0,667 termasuk *klasifikasi baik*, sedangkan garis berwarna biru merupakan kurva ambang (*thresholds).*

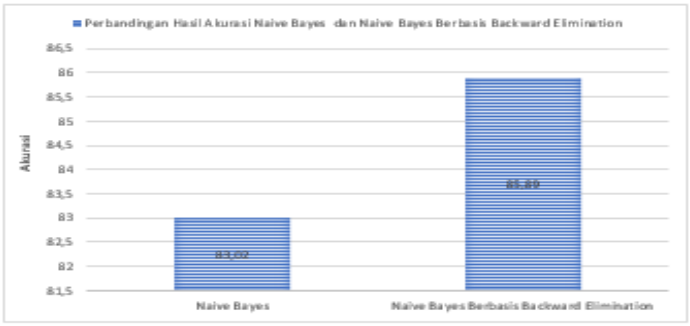


Gambar 2 Kurva ROC

Berdasarkan hasil dengan *Confusion Matrix* dan kurva *ROC* menunjukkan *rule* hasil klasifikasi untuk memprediksi keputusan klien termasuk *klasifikasi baik* sehingga dapat digunakan untuk klasifikasi nasabah asuransi jiwa.

3.2. 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 83,02% sedangkan *Naive Bayes* berbasis *Backward Elimination 85*,89%, 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 3.2.



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

ISSN print 2087-1716
ISSN online 2548-7779

4. Kesimpulan Dan Saran

4.1 Kesimpulan

1. Dari sembilan atribut yang digunakan yaitu umur, pekerjaan, penghasilan, jenis asuransi, jangka waktu, cara bayar, mata uang, premi, dan jumlah hari dengan menggunakan algoritma *Naive Bayes* berbasis *Backward Elimination* didapatkan delapan atribut weight yaitu umur, jangka waktu, cara bayar, premi, jumlah hari, pekerjaan, penghasilan dan mata uang dalam mengklasifikasi nasabah yang lancar dan tidak lancar dalam pembayaran premi.
2. Secara mandiri tingkat akurasi yang dihasilkan algoritma *Naive Bayes* adalah 83,02 % dengan menambahkan seleksi fitur *Backward Elimination* menghasilkan akurasi 85,89 % dalam klasifikasi nasabah lancar dan tidak lancar.

4.2 Saran

1. Untuk meningkatkan akurasi klasifikasi nasabah lancar dan tidak lancar dapat dilakukan dengan penggabungan beberapa algoritma dan dapat juga menggunakan seleksi fitur yang lain.
2. Menggunakan dataset yang lain sebagai perbandingan tingkat akurasi yang dihasilkan algoritma Naive Bayes

Daftar Pustaka

[1] Bustami. 2013. Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. TECHSI.
[2] Sunjana. 2010. Klasifikasi Nasabah Sebuah Asuransi Menggunakan Algoritma C4.5. Yogyakarta.
[3] Jurek A *et al*. 2008. *Improving Naive Bayes Models of Insurance Risk by Unsupervised Classification.* Computer Science and Information Technology pp.137-144.
[4] Larose D. T. 2006. *Data Mining Methods And Models.* Canada: John Wiley & Sons. Inc.19.
[5] Gorunescu, F. 2010. *Data Mining: Concept, Models and Techniques.* Romania:Springer.
[6] Nuraeni F. 2013. Algoritma C4.5 Untuk Klasifikasi Pola Pembayaran Kredit Motor Pada Perusahaan Pembiayaan (*Leasing*). Seminar Nasional Informatika.
[7] Santosa, B. 2007. Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis. Jogyakarta: Graha Ilmu.
[8] Kusrini & Luthfi. E. Taufiq. 2009. *Algoritma Data Mining*. Yogyakarta: Andi Offset.
[9] Republik Indonesia. 1992. Undang – Undang No. 2 Tahun 1992 Tentang Usaha Perasuransian. Sekertariat Negara. Jakarta.

