

Penerapan Algoritma *Naive Bayes Classifier* dalam Memprediksi Status Keberlanjutan Polis Nasabah Asuransi PT.X

Ajeng Mega Pratiwi*, Aceng Komarudin Mutaqin

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

*pratiwiajengmega2104@gmail.com , Aceng.k.mutaqin@gmail.com

Abstract. This article discusses the classification in predicting the sustainability status of the health insurance customer policy of PT. X uses the Naive Bayes Classifier Algorithm. In predicting the Naive Bayes Classifier Algorithm, it uses the concepts and theories of data mining in the literature related to insurance by calculating the probability of each class of variables using the Bayes theorem in describing the performance of a model or algorithm specifically using the Confusion Matrix. To be able to predict the decisions of health insurance customers in the policy sustainability status, a method of data analysis of registered insurance customers is needed. The data used is data obtained from the insurance company PT. X. The data contains customer information data in the form of 9 variables (Policy Number, Smoking Status, Gender, Age, Marital Status, Dependents, Monthly Premiums, Current Status / whether or not premium payments and insurance policy renewal status). The results of the application of the Naive Bayes Classifier Algorithm show that the algorithm is quite good in predicting the status of the policy extension of the insured health insurance PT. X, with an average accuracy of 85.82%, an average precision of 96.10% and an average recall of 93.55.

Keywords: *Classification, Naive Bayes, Prediction of Health Insurance Policy Renewal.*

Abstrak. Artikel ini membahas tentang klasifikasi dalam memprediksi status keberlanjutan polis nasabah asuransi kesehatan PT. X menggunakan Algoritma Nave Bayes Classifier. Dalam memprediksi Algoritma Naive Bayes Classifier menggunakan konsep dan teori data mining dalam literatur yang berhubungan dengan asuransi dengan menghitung probabilitas setiap kelas variabel menggunakan teorema Bayes dalam menggambarkan kinerja suatu model atau algoritma secara khusus menggunakan Confusion Matrix . Untuk dapat memprediksi keputusan nasabah asuransi kesehatan dalam status kesinambungan polis, diperlukan suatu metode analisis data nasabah asuransi yang terdaftar. Data yang digunakan adalah data yang diperoleh dari perusahaan asuransi PT. X. Data tersebut berisi data informasi nasabah berupa 9 variabel (Nomor Polis, Status Merokok, Jenis Kelamin, Usia, Status Perkawinan, Tanggungan, Premi Bulanan, Status Lancar/tidaknya pembayaran premi dan status perpanjangan polis asuransi). Hasil penerapan Algoritma Naive Bayes Classifier menunjukkan bahwa algoritma tersebut cukup baik dalam memprediksi status perpanjangan polis dari tertanggung asuransi kesehatan PT. X, dengan rata-rata akurasi 85,82%, presisi rata-rata 96,10% dan rata-rata recall 93,55.

Kata Kunci: *Klasifikasi, Nave Bayes, Prediksi Perpanjangan Polis Jaminan Kesehatan.*

A. Pendahuluan

Analisis regresi adalah suatu kajian analisis pada hubungan antara variabel yang biasa disebut dengan variabel bebas (variabel prediktor) terhadap variabel tak bebas (variabel respon) (Gujarati, 2006). Salah satu bentuk paling umum dari bentuk *Count Regression* ialah Regresi Poisson. Regresi poisson merupakan suatu metode regresi yang biasa digunakan menganalisis data dengan variabel responnya berupa data diskrit. Dalam regresi poisson terdapat asumsi yang harus dip Indonesia Setiap manusia akan selalu dihadapkan pada peristiwa yang tidak menentu dalam hidupnya. Peristiwa tersebut seringkali disebut dengan risiko. Hal ini termasuk dalam masalah keuntungan atau kerugian yang dihadapi. Contohnya kecelakaan, kebakaran, hilang, rusak, sakit, meninggal dan lain sebagainya yang mungkin bisa saja terjadi. Untuk mengganti atau menyeimbangkan kembali, asuransi merupakan sarana perlindungan dari kerugian finansial.

Asuransi adalah perjanjian antara dua pihak, pihak pertama disebut penanggung atau perusahaan asuransi dan pihak kedua disebut tertanggung sebagai pemegang polis. Asuransi juga dapat diartikan bentuk jaminan resiko yang mungkin saja terjadi di masa yang akan datang dengan nominal biaya yang tidak sedikit (Ignatius, 2018). Oleh sebab itu dibutuhkan asuransi dalam menjamin sesuatu yang berharga agar mendapatkan langkah preventif sebagai jaminan jika terjadi suatu yang tidak diinginkan. Untuk memperoleh jaminan asuransi atau risiko yang mungkin terjadi, maka pemegang polis atau tertanggung diwajibkan membayar sejumlah uang yang sering disebut premi dalam jangka waktu tertentu yang sudah ditentukan. Polis asuransi merupakan sebuah bukti perjanjian tertulis yang dilakukan oleh pihak perusahaan asuransi (penanggung) dengan nasabah pengguna layanan asuransi (tertanggung). Polis asuransi merupakan hal yang sangat penting di dalam layanan asuransi itu sendiri, karena polis asuransi akan melindungi setiap hak dan kewajiban nasabah dan juga pihak perusahaan asuransi dalam jangka waktu tertentu. Sehingga nasabah harus pintar dalam menentukan perusahaan asuransi yang dipilih.

Penawaran asuransi tidak terbatas pada orang-orang yang terpelajar dan memiliki pendapatan besar, tetapi semua golongan masyarakat dapat turut serta didalamnya. Produk asuransi yang ditawarkan perusahaan asuransi juga semakin bermacam-macam seiring berkembangnya peradaban masyarakat dan resiko hidup yang semakin tinggi. Mulai dari produk asuransi jiwa, asuransi pendidikan, dan asuransi kesehatan. Semakin banyaknya persaingan dalam bisnis asuransi membuat perusahaan asuransi dituntut memiliki inovasi dalam mempertahankan bahkan meningkatkan jumlah nasabahnya. Disisi lain, keputusan nasabah asuransi kesehatan terhadap keberlanjutan polis asuransi yang telah dimiliki tidaklah mudah untuk diprediksi. Untuk dapat memprediksi keputusan nasabah asuransi kesehatan dalam memperpanjang polis yang dimiliki maka diperlukan suatu metode analisis pada data nasabah asuransi yang telah terdaftar. Dalam dunia asuransi akan selalu dihadapkan dengan permasalahan yang sering timbul yaitu banyaknya nasabah yang tidak lancar dalam membayar premi dan memutuskan untuk tidak memperpanjang polis asuransinya. Bahkan ada nasabah yang pindah dari satu perusahaan asuransi ke perusahaan asuransi lain karena produk di perusahaan asuransi lain lebih menarik dan menguntungkan.

Algoritma *Naïve Bayes Classifier* merupakan cara untuk memprediksi status keberlanjutan polis nasabah asuransi dengan klasifikasi data nasabah. Teknik klasifikasi suatu proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan kelas dari suatu objek (Natalius, 2011). Pemanfaatan algoritma *Naïve Bayes* sebagai salah satu algoritma terbaik dalam data mining klasifikasi untuk memprediksi keputusan nasabah asuransi terhadap polis asuranssi yang telah dimiliki. Tujuan penelitian ini adalah menerapkan metode algoritma *Naive Bayes Classifier* dalam memprediksi status keberlanjutan polis nasabah asuransi kesehatan PT. X. Dalam proses evaluasi Algoritma *Naïve Bayes Classifier* yang dihasilkan, digunakan alat analisis yaitu *Confusion Matrix* untuk menggambarkan performa dari sebuah model atau algoritma secara spesifik (Kuncahyo, 2019). Setiap kolom dari matrix tersebut, mempresentasikan kelas aktual dari data, dan setiap kolom mempresentasikan kelas prediksi dari data dengan informasi akurasi, presisi dan recall dari hasil Algoritma.

B. Metodologi Penelitian

Metode Pengumpulan Data

Pada penelitian ini menggunakan Data yang digunakan pada penelitian ini adalah sebanyak 92 pengamatan berisikan data informasi tertanggung polis asuransi kesehatan PT. X berupa 9 variabel (Nomor Tertanggung, Status Merokok, Jenis Kelamin, Usia, Status Pernikahan, Tanggungan, Premi perbulan, Status Lancar/ tidaknya pembayaran premi dan Status Perpanjangan Polis). Data tersebut akan dibagi menjadi 2 yaitu data training dan data testing, dimana pengamatan yang masuk kedalam data training dan data testing dilakukan secara acak. Proporsi untuk data training yaitu sebesar 70% dari data keseluruhan atau sebanyak 64 pengamatan. Sedangkan proporsi untuk data testing yaitu sebesar 30% dari data keseluruhan atau sebanyak 28 pengamatan.

Tahapan Analisis Data

Membagi proporsi data nasabah menjadi 2 bagian kelompok data, yaitu sebagai data latih (Proses *Training*) dan sebagai data uji (Proses *Testing*):

Analisis Data *Training*

1. Membaca data nasabah asuransi yang akan digunakan sebagai data *training*.
2. Memeriksa bentuk dari setiap kriteria nasabah asuransi, baik yang bertipe numerik/kontinyu maupun yang berbentuk kategorik.
3. Jika data berbentuk numerik, maka hitung mean dan standar deviasi tiap atribut.

Jika data bersifat kategorik, maka langkah selanjutnya yaitu menghitung probabilitas prior tiap kelasnya.

Rumus nilai probabilitas *Naïve Bayes Classifier* dapat ditulis dalam bentuk persamaan berikut:

$$P(C|F) = \frac{P(F|C) P(C)}{P(F)} \quad (1)$$

dimana,

F = Data dengan kelas yang belum diketahui

C = Hipotesis data F merupakan suatu kelas spesifik

$P(C|F)$ = Probabilitas hipotesis C berdasarkan kondisi F

$P(F|C)$ = Probabilitas F berdasarkan hipotesis C

$P(C)$ = Probabilitas awal (prior) hipotesis C

$P(F)$ = Probabilitas awal (prior) dari F

Pada Algoritma *Naïve Bayes Classifier* data dengan tipe numerik (kontinyu) dapat dihitung dengan cara mengasumsikan bentuk tertentu dari distribusi probabilitas untuk fitur kontinu dan memperkirakan parameter distribusi dengan data training. Distribusi normal dikarakteristikan dengan dua parameter, yaitu nilai rata-rata (μ) dan nilai varians (σ^2). Untuk setiap kelas y_j , probabilitas bersyarat kelas y_j untuk data x_i adalah:

$$P(X_i = x_i | Y_j = y_j) = \frac{1}{\sigma_{ij}\sqrt{2\pi}} e^{-\frac{(x_{ij}-\mu_{ij})^2}{2\sigma_{ij}^2}} \quad (2)$$

Keterangan:

P = Probabilitas

X_i = Atribut ke i

x_i = Nilai atribut ke i

Y_j = Kelas yang dicari

y_j = Sub kelas Y yang dicari

μ_{ij} = Nilai rata-rata dari seluruh atribut

σ_{ij} = Varians dari seluruh atribut

dimana,

$$\mu_{ij} = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j} \quad (3)$$

$$\sigma^2_{ij} = \frac{\sum_{i=1}^{n_j} (x_{ij} - \mu_{ij})^2}{n_j} \quad (4)$$

Selanjutnya, dalam mengukur performa Algoritma *Naïve Bayes Classifier* menggunakan *confusion matrix* dapat diperoleh data-data lain yang sangat berguna untuk mengukur performa sebuah algoritma, diantaranya:

Akurasi merupakan persentase ketepatan prediksi data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil.

$$\frac{TP+TN}{Total} \quad (5)$$

Presisi merupakan Rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif atau ketika model memprediksi positif, seberapa sering prediksi itu benar.

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

Recall merupakan Rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif atau ketika kelas aktualnya positif, seberapa sering model memprediksi positif.

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

Analisis Data *Testing*

1. Menganalisa data uji dengan menghitung probabilitas bersyarat setiap kriteria.
2. Mencari nilai peluang bersyarat tiap atribut.
3. Untuk data tipe numerik (kontinyu) klasifikasi dapat dilakukan dengan rumus *Densitas Gaussian*.
4. Kemudian, hitung dua parameter pada Distribusi *Gaussian* yakni *mean* (μ) dengan menggunakan Persamaan (2.8) dan *varian* (σ^2) dengan menggunakan.
5. Menemukan model Teorema Algoritma *Naïve Bayes Classifier*.
6. Setelah model prediksi telah dibangun pada data training tahap selanjutnya adalah mengklasifikasi data yang belum diketahui label kelasnya.
7. Menentukan peluang kelas bersyarat $P(X_i = x_i | \text{Perpanjang Polis})$, $P(X_i = x_i | \text{Tidak Perpanjang Polis})$ yang dihasilkan dari proses *training*.
8. Menghitung probabilitas posterior pada data uji, menggunakan.
9. Menentukan label kelas.
10. Membuat *Confusion Matrix*
11. Menghitung keakurasian performa dari sebuah model atau algoritma secara spesifik
12. Menghitung presisi atau membandingkan rasio positif dengan keseluruhan hasil yang diprediksi positif
13. Menghitung Recall/sensitifitas untuk membandingkan rasio benar positif dengan keseluruhan data yang benar positif
14. Penarikan kesimpulan.

C. Hasil Penelitian dan Pembahasan

Menghitung probabilitas prior dari data tertanggung polis asuransi kesehatan PT.X yang termasuk dalam data training. Gambar 1. menyajikan sintaks perangkat lunak R Studio serta outputnya untuk menghitung nilai probabilitas prior.

```
naiveBayes.default(x = X, y = Y, laplace = laplace)

A-priori probabilities:
```

Gambar 1 sintaks dan output perangkat lunak R Studio untuk nilai probabilitas prior.

Berdasarkan Gambar 1 di atas, dapat dilihat bahwa nilai probabilitas prior bahwa tertanggung akan memperpanjang polis yaitu sebesar $47/64 = 0,7344$. Sedangkan nilai probabilitas prior bahwa tertanggung tidak akan memperpanjang polis yaitu sebesar $17/64 = 0,2656$.

Langkah selanjutnya menghitung probabilitas bersyarat untuk atribut kategorik. Dengan dibantu perangkat lunak Excel, Tabel 1 sampai dengan Tabel 5 menyajikan nilai probabilitas bersyarat untuk atribut status merokok, jenis kelamin, status menikah, tanggungan dan status pembayaran premi.

Tabel 1 Probabilitas Tertanggung dengan Atribut Status Merokok

PROBABILITAS TERTANGGUNG		
P (Merokok Tidak Merokok)	Perpanjang Polis	Tidak perpanjang Polis
Ya	0,0213	0,1176
Tidak	0,9787	0,8824
	1	1

Berdasarkan Tabel 1 di atas, dapat dilihat bahwa nilai probabilitas tertanggung statusnya merokok dengan syarat tertanggung tersebut memperpanjang polis di tahun 2020 yaitu sebesar $1/47 = 0,0213$. Nilai probabilitas bahwa tertanggung statusnya merokok dengan syarat tertanggung tersebut tidak memperpanjang polis di tahun 2020 yaitu sebesar $2/17 = 0,1176$. Nilai probabilitas bahwa tertanggung statusnya tidak merokok dengan syarat tertanggung tersebut memperpanjang polis di tahun 2020 yaitu sebesar $46/47 = 0,9787$. Nilai probabilitas bahwa tertanggung statusnya tidak merokok dengan syarat tertanggung tersebut tidak memperpanjang polis di tahun 2020 yaitu sebesar $15/17 = 0,8824$.

Tabel 2 Probabilitas Tertanggung dengan Atribut Jenis Kelamin

FREKUENSI TERTANGGUNG		
P(Laki-laki Perempuan)	Perpanjang Polis	Tidak perpanjang Polis
Laki-laki	31	7
Perempuan	16	10
	47	17

Berdasarkan Tabel 2 di atas, dapat dilihat bahwa tertanggung berjenis kelamin laki-laki dengan syarat tertanggung tersebut memperpanjang polis di tahun 2020 yaitu sebesar $31/47 = 0,6596$. Nilai probabilitas bahwa tertanggung berjenis kelamin laki-laki dengan syarat tertanggung tersebut tidak memperpanjang polis di tahun 2020 yaitu sebesar $7/17 = 0,4118$. Nilai probabilitas bahwa tertanggung berjenis kelamin perempuan dengan syarat tertanggung tersebut memperpanjang polis di tahun 2020 yaitu sebesar $16/47 = 0,3404$. Nilai probabilitas bahwa tertanggung berjenis kelamin perempuan dengan syarat tertanggung tersebut tidak memperpanjang polis di tahun 2020 yaitu sebesar $10/17 = 0,5882$.

Tabel 3 Probabilitas Tertanggung dengan Atribut Status Menikah

PROBABILITAS TERTANGGUNG		
P (Menikah Belum Menikah)	Perpanjang Polis	Tidak perpanjang Polis

Menikah	0,4681	0,4706
Belum Menikah	0,5319	0,5294
	1	1

Berdasarkan Tabel 3 di atas, dapat dilihat bahwa tertanggung statusnya menikah dengan syarat tertanggung tersebut memperpanjang polis di tahun 2020 yaitu sebesar $22/47 = 0,4681$. Nilai probabilitas bahwa tertanggung statusnya menikah dengan syarat tertanggung tersebut tidak memperpanjang polis di tahun 2020 yaitu sebesar $8/17 = 0,4706$. Nilai probabilitas bahwa tertanggung statusnya belum menikah dengan syarat tertanggung tersebut memperpanjang polis di tahun 2020 yaitu sebesar $25/47 = 0,5319$. Nilai probabilitas bahwa tertanggung statusnya belum menikah dengan syarat tertanggung tersebut tidak memperpanjang polis di tahun 2020 yaitu sebesar $9/17 = 0,5294$.

Tabel 4 Probabilitas Tertanggung dengan Atribut Tanggungan

PROBABILITAS TERTANGGUNG		
P (Ada Tanggungan Tidak Ada Tanggungan)	Perpanjang Polis	Tidak perpanjang Polis
Ada	0,4255	0,5882
Tidak	0,5745	0,4118
	1	1

Berdasarkan Tabel 4 di atas, dapat dilihat bahwa nilai probabilitas tertanggung yang memiliki tanggungan dengan syarat tertanggung tersebut memperpanjang polis di tahun 2020 yaitu sebesar $20/47 = 0,4255$. Nilai probabilitas bahwa tertanggung yang memiliki tanggungan dengan syarat tertanggung tersebut tidak memperpanjang polis di tahun 2020 yaitu sebesar $10/17 = 0,5882$. Nilai probabilitas bahwa tertanggung yang tidak memiliki tanggungan dengan syarat tertanggung tersebut memperpanjang polis di tahun 2020 yaitu sebesar $27/47 = 0,5745$. Nilai probabilitas bahwa tertanggung yang tidak memiliki tanggungan dengan syarat tertanggung tersebut tidak memperpanjang polis di tahun 2020 yaitu sebesar $7/17 = 0,4118$.

Tabel 5 Probabilitas Tertanggung dengan Atribut Status Pembayaran Premi

PROBABILITAS TERTANGGUNG		
P (Lancar Tidak Lancar)	Perpanjang Polis	Tidak perpanjang Polis
Lancar	0,8936	0,5294
Tidak Lancar	0,1064	0,4706
	1	1

Berdasarkan Tabel 5 di atas, dapat dilihat bahwa nilai probabilitas tertanggung yang status pembayaran preminya lancar dengan syarat tertanggung tersebut memperpanjang polis di tahun 2020 yaitu sebesar $42/47 = 0,8936$. Nilai probabilitas bahwa tertanggung yang status pembayaran preminya lancar dengan syarat tertanggung tersebut tidak memperpanjang polis di tahun 2020 yaitu sebesar $9/17 = 0,5294$. Nilai probabilitas bahwa tertanggung yang status pembayaran preminya tidak lancar dengan syarat tertanggung tersebut memperpanjang polis di tahun 2020 yaitu sebesar $5/47 = 0,1064$. Nilai probabilitas bahwa tertanggung yang status pembayaran preminya tidak lancar dengan syarat tertanggung tersebut tidak memperpanjang polis di tahun 2020 yaitu sebesar $8/17 = 0,4706$.

Selanjutnya, akan dihitung nilai probabilitas bersyarat untuk atribut numerik. Dalam data tertanggung polis asuransi kesehatan di PT. X tahun 2020 terdapat 2 atribut numerik yaitu usia dan pembayaran premi per bulan. Pada atribut usia nilai rata-rata usia tertanggung yang memperpanjang polis asuransi kesehatan di tahun 2020 yaitu sebesar $\mu_{ij} = 33$,

sedangkan nilai rata-rata usia tertanggung yang tidak memperpanjang polis asuransi kesehatan di tahun 2020 yaitu sebesar $\mu_{ij} = 36,4706$. Dan nilai varians usia tertanggung yang memperpanjang polis asuransi kesehatan di tahun 2020 yaitu sebesar $\sigma^2_{ij} = 383,4348$. Sedangkan, nilai varians usia tertanggung yang tidak memperpanjang polis asuransi kesehatan di tahun 2020 yaitu sebesar $\sigma^2_{ij} = 258,1397$. Dihitung nilai fungsi probabilitas usia untuk

contoh tertanggung usia 23 tahun,

$$f(\text{usia} = 23|\text{perpanjang polis}) = \frac{1}{19,5815\sqrt{2\pi}} e^{-\frac{(23-33)^2}{2(383,4348)}} = 0,0179$$

$$f(\text{usia} = 23|\text{tidak perpanjang polis}) = \frac{1}{16,0667\sqrt{2\pi}} e^{-\frac{(23-36,4706)^2}{2(258,1397)}} = 0,0175$$

Berdasarkan perhitungan di atas dapat dilihat bahwa tertanggung usia 23 tahun dengan syarat perpanjang polis asuransi kesehatan di tahun 2020 memiliki nilai fungsi probabilitas sebesar 0,0179. Sedangkan tertanggung usia 23 tahun dengan syarat tidak perpanjang polis asuransi kesehatan di tahun 2020 memiliki nilai fungsi probabilitas sebesar 0,0175.

Selanjutnya, akan dihitung nilai probabilitas bersyarat untuk atribut pembayaran premi per bulan. Pada atribut pembayaran premi per bulan didapat nilai rata-rata premi per bulan tertanggung yang memperpanjang polis asuransi kesehatan di tahun 2020 yaitu sebesar $\mu_{ij} = 883.500$, sedangkan nilai rata-rata premi per bulan tertanggung yang tidak memperpanjang polis asuransi kesehatan di tahun 2020 yaitu sebesar $\mu_{ij} = 1.014.941$. Dengan nilai varians tertanggung yang memperpanjang polis asuransi kesehatan di tahun 2020 yaitu sebesar $\sigma^2_{ij} = 719.258.750.000$. Sedangkan, nilai varians tertanggung yang tidak memperpanjang polis asuransi kesehatan di tahun 2020 yaitu sebesar $\sigma^2_{ij} = 311.118.058.823$. Dihitung nilai fungsi probabilitas pembayaran premi per bulan untuk contoh tertanggung dengan pembayaran premi per bulan Rp.300.000,

$$f(\text{Pembayaran Premi} = 300.000|\text{Perpanjang Polis}) = \frac{1}{848.091\sqrt{2\pi}} e^{-\frac{(300.000-883.500)^2}{2(719.258.750.000)}} = 3,7135 \times 10^{-7}$$

$$f(\text{Pembayaran Premi} = 300.000|\text{Tidak Perpanjang Polis}) = \frac{1}{557.779\sqrt{2\pi}} e^{-\frac{(300.000-1.014.941)^2}{2(311.118.058.823)}} = 3,1463 \times 10^{-7}$$

Berdasarkan perhitungan di atas dapat dilihat bahwa tertanggung dengan pembayaran premi per bulan Rp.300.000 bersyarat perpanjang polis asuransi kesehatan di tahun 2020 memiliki nilai fungsi probabilitas sebesar $3,7135 \times 10^{-7}$. Sedangkan tertanggung dengan pembayaran premi per bulan Rp.300.000 bersyarat tidak perpanjang polis asuransi kesehatan di tahun 2020 memiliki nilai fungsi probabilitas sebesar $3,1463 \times 10^{-7}$.

Langkah selanjutnya, akan dilakukan analisis data testing. Sebanyak 28 tertanggung yang termasuk dalam data testing akan diprediksi status perpanjang polisnya di tahun 2020. Misalkan akan diprediksi status perpanjang polis untuk tertanggung dengan nomor tertanggung 10, berjenis kelamin laki-laki, usia 23 tahun, status pernikahan belum menikah, tidak punya tanggungan, tidak merokok, premi per bulan sebesar Rp. 300.000 dan status pembayaran preminya lancar.

Dengan menggunakan informasi tersebut, dapat dihitung nilai probabilitas bersama untuk semua atribut data bagi tertanggung 10 bersyarat perpanjang polis dan bersyarat tidak perpanjang polis, masing-masing adalah

$$\begin{aligned} P(\text{atribut tertanggung 10}|\text{perpanjang polis}) &= P(\text{tidak merokok}|\text{perpanjang polis}) \times P(\text{laki - laki}|\text{perpanjang polis}) \\ &\times P(\text{belum menikah}|\text{perpanjang polis}) \times P(\text{tidak ada tanggungan}|\text{perpanjang polis}) \times P(\text{lancar}|\text{perpanjang polis}) \times P(\text{usia} = 23|\text{perpanjang polis}) \times P(\text{premi perbulan} = \text{Rp. 300.000}|\text{perpanjang polis}) \\ &= (0,9787) \times (0,6596) \times (0,5319) \times (0,5745) \times (0,8936) \times (0,0179) \times (3,7135 \times 10^{-7}) \\ &= 1,1717 \times 10^{-9} \end{aligned}$$

$$\begin{aligned} P(\text{atribut tertanggung 10}|\text{tidak perpanjang polis}) &= P(\text{tidak merokok}|\text{tidak perpanjang polis}) \times P(\text{laki - laki}|\text{tidak perpanjang polis}) \times P(\text{belum menikah}|\text{tidak perpanjang polis}) \times P(\text{tidak ada tanggungan}|\text{tidak perpanjang polis}) \times P(\text{lancar}|\text{tidak perpanjang polis}) \times P(\text{usia} = 23|\text{perpanjang polis}) \times P(\text{premi perbulan} = \text{Rp. 300.000}|\text{perpanjang polis}) \\ &= (0,8824) \times (0,4118) \times (0,5294) \times (0,4118) \times (0,5294) \times (0,0175) \times (3,1463 \times 10^{-7}) \\ &= 2,3091 \times 10^{-10} \end{aligned}$$

Dengan demikian didapat probabilitas posterior baik untuk perpanjangan polis maupun untuk tidak perpanjangan polis, yaitu

$$P(\text{perpanjangan polis}|\text{atribut tertanggung 10}) = \frac{P(\text{atribut tertanggung 10}|\text{perpanjangan polis}) \times P(\text{perpanjangan polis})}{P(\text{atribut tertanggung 10}|\text{perpanjangan polis}) \times P(\text{perpanjangan polis}) + P(\text{atribut tertanggung 10}|\text{tidak perpanjangan polis}) \times P(\text{tidak perpanjangan polis})}$$

$$= \frac{1,1717 \times 10^{-9} \times 0,7344}{1,1717 \times 10^{-9} \times 0,7344 + 2,3091 \times 10^{-10} \times 0,2656} = 0,9335$$

$$P(\text{tidak perpanjangan polis}|\text{atribut tertanggung 10}) = \frac{P(\text{atribut tertanggung 10}|\text{tidak perpanjangan polis}) \times P(\text{tidak perpanjangan polis})}{P(\text{atribut tertanggung 10}|\text{perpanjangan polis}) \times P(\text{perpanjangan polis}) + P(\text{atribut tertanggung 10}|\text{tidak perpanjangan polis}) \times P(\text{tidak perpanjangan polis})}$$

$$= \frac{2,3091 \times 10^{-10} \times 0,2656}{(1,1717 \times 10^{-9}) \times 0,7344 + (2,3091 \times 10^{-10}) \times 0,2656} = 0,0665$$

Karena probabilitas posterior untuk perpanjangan polis lebih besar dibandingkan dengan probabilitas posterior untuk tidak perpanjangan polis, maka tertanggung 10 diklasifikasikan dalam kelas perpanjangan polis. Langkah selanjutnya melakukan prediksi pada sisa polis lainnya yang ada di data testing. Tabel 6 menyajikan hasil prediksi status perpanjangan polis dan status perpanjangan polis sebenarnya untuk 28 tertanggung yang ada pada data testing. Hasil prediksinya diperoleh dengan bantuan perangkat lunak R. Studio 1.1.463.

Tabel 6 Data Prediksi Perpanjangan Polis Asuransi Kesehatan

No	Nama Tertanggung Utama	Prediksi Perpanjangan	Status Perpanjangan Polis tahun 2020
1	Tertanggung 10	Perpanjangan Polis	Perpanjangan Polis
2	Tertanggung 14	Tidak perpanjangan Polis	Tidak perpanjangan Polis
3	Tertanggung 23	Tidak perpanjangan Polis	Perpanjangan Polis
4	Tertanggung 30	Perpanjangan Polis	Tidak perpanjangan Polis
5	Tertanggung 32	Perpanjangan Polis	Perpanjangan Polis
6	Tertanggung 35	Perpanjangan Polis	Perpanjangan Polis
7	Tertanggung 39	Perpanjangan Polis	Perpanjangan Polis
8	Tertanggung 40	Perpanjangan Polis	Perpanjangan Polis
9	Tertanggung 5	Tidak perpanjangan Polis	Tidak perpanjangan Polis
10	Tertanggung 51	Perpanjangan Polis	Perpanjangan Polis
11	Tertanggung 52	Perpanjangan Polis	Perpanjangan Polis
12	Tertanggung 54	Perpanjangan Polis	Perpanjangan Polis
13	Tertanggung 55	Perpanjangan Polis	Perpanjangan Polis
14	Tertanggung 57	Perpanjangan Polis	Perpanjangan Polis
15	Tertanggung 58	Perpanjangan Polis	Perpanjangan Polis
16	Tertanggung 59	Perpanjangan Polis	Perpanjangan Polis
17	Tertanggung 60	Perpanjangan Polis	Perpanjangan Polis
18	Tertanggung 62	Tidak perpanjangan Polis	Tidak perpanjangan Polis
19	Tertanggung 69	Perpanjangan Polis	Perpanjangan Polis
20	Tertanggung 70	Tidak perpanjangan Polis	Perpanjangan Polis
21	Tertanggung 77	Perpanjangan Polis	Perpanjangan Polis
22	Tertanggung 78	Perpanjangan Polis	Perpanjangan Polis
23	Tertanggung 82	Perpanjangan Polis	Perpanjangan Polis
24	Tertanggung 83	Perpanjangan Polis	Perpanjangan Polis
25	Tertanggung 86	Tidak perpanjangan Polis	Tidak perpanjangan Polis
26	Tertanggung 87	Perpanjangan Polis	Perpanjangan Polis
27	Tertanggung 88	Perpanjangan Polis	Perpanjangan Polis
28	Tertanggung 91	Perpanjangan Polis	Perpanjangan Polis

Dibuat *confusion matrix* untuk melihat perbandingan hasil klasifikasi yang dilakukan oleh sistem (model) dengan klasifikasi sebenarnya. Tabel 7 menyajikan hasil *confusion matrix* yang diperoleh dengan bantuan perangkat lunak R. Studio 1.1.463.

Tabel 7 Confusion Matrix

Nilai Prediksi \ Nilai Aktual	Nilai Aktual	
	Perpanjang Polis	Tidak perpanjang Polis
Perpanjang Polis	TP = 21	FP = 1
Tidak perpanjang Polis	FN = 2	TN = 4

Berdasarkan Tabel 7 di atas dapat dilihat bahwa terdapat 21 tertanggung yang termasuk dalam *True Positive* (TP), artinya menurut data aktual ada sebanyak 21 tertanggung akan memperpanjang polis asuransi kesehatan dan dari model yang dibuat memprediksi bahwa 21 tertanggung ini akan memperpanjang polis asuransi kesehatan. Terdapat 4 tertanggung yang termasuk dalam *True Negative* (TN), artinya menurut data aktual ada sebanyak 4 tertanggung tidak akan memperpanjang polis asuransi kesehatan dan model yang dibuat memprediksi bahwa 4 tertanggung ini tidak akan memperpanjang polis asuransi kesehatan. Terdapat 1 tertanggung yang termasuk dalam *False Positive* (FP), artinya 1 tertanggung ini menurut data aktual akan memperpanjang polis asuransi kesehatan tetapi dari model yang dibuat memprediksi bahwa 1 tertanggung ini tidak akan memperpanjang polis asuransi kesehatan. Terdapat 2 tertanggung yang termasuk dalam *False Negative* (FN), artinya 2 tertanggung ini menurut data aktual tidak akan memperpanjang polis asuransi kesehatan tetapi dari model yang dibuat memprediksi bahwa 2 tertanggung ini akan memperpanjang polis asuransi kesehatan.

Selanjutnya akan dilakukan perhitungan performa Algoritma *Naïve Bayes Classifier*. Ada 3 nilai yang akan dihitung yaitu nilai akurasi, presisi dan recall. Adapun nilai akurasinya adalah

$$\text{Nilai akurasi} = \frac{TP+TN}{TP+FP+FN+TN} = \frac{21+4}{21+1+2+4} = \frac{25}{28} = 0,8929.$$

$$\text{Nilai presisi} = \frac{TP}{TP+FP} = \frac{21}{21+1} = \frac{21}{22} = 0,9545$$

$$\text{Nilai recall} = \frac{TP}{TP+FN} = \frac{21}{21+2} = \frac{21}{23} = 0,9130$$

Berdasarkan perhitungan di atas didapat nilai akurasi sebesar 89,29%, yang artinya model secara akurat mengklasifikasikan status perpanjang polis tertanggung dengan benar sebanyak 89,29% dari seluruh tertanggung, dengan nilai presisi sebesar 95,45% dan nilai recall sebesar 91,30%.

D. Kesimpulan

Pada artikel ini diterapkan *Algoritma Naïve Bayes Classifier* dalam memprediksi status perpanjangan polis tertanggung asuransi kesehatan di PT. X berdasarkan data tertanggung di tahun 2020. Hasil penerapan *Algoritma Naïve Bayes Classifier* menunjukkan bahwa algoritma tersebut cukup baik dalam memprediksi status perpanjangan polis tertanggung asuransi kesehatan PT. X, dengan nilai akurasi sebesar 89,29%, yang artinya model secara akurat mengklasifikasikan status perpanjang polis tertanggung dengan benar sebanyak 89,29% dari seluruh tertanggung, dengan nilai presisi sebesar 95,45% dan nilai recall sebesar 91,30%.

Acknowledge

Penelitian ini dapat terlaksana dengan baik tentunya berkat bantuan dari berbagai pihak. Peneliti mengucapkan terimakasih kepada bapak Dr. Aceng Komarudin Mutaqin, MT., M.Si., para dosen statistika Uniba, rekan-rekan seperjuangan yaitu mahasiswa statistika atas bantuan, doa, dan bimbingannya.

Daftar Pustaka

- [1] Kusriani & Luthfi, E. T., *Algoritma Data Mining*. Andi Offset. Yogyakarta (2009).
- [2] Patil, T.R., Sherekar, M. S. (2013). *Performance Analysis of Naïve Bayes and J48 Classification Algorithm for Data Classification*, International Journal of Computer Science and Application, 6(2), 0974-1011.
- [3] Bustami. (2014). Penerapan Algoritma Naïve Bayes untuk Mengklasifikasi Data Nasabah Asuransi. Jurnal Penelitian Teknik Informatika, Vol: 8 No.1.

- [4] Ignatius, R. Y. S. (2018). *Pengantar Asuransi*. Jakarta: ACA Asuransi.
- [5] Kitab Undang-Undang Hukum Dagang (KUHD), Bab IX Tentang Asuransi atau Pertanggungan pada Umumnya, Pasal 246.
- [6] N. Nuraeni, “Penentuan Kelayakan Kredit Dengan Algoritma Naïve Bayes Classifier: Studi Kasus Bank Maya pada Mitra Usaha Cabang PGC,” J. Tek. Komput. AMIK BSI, vol. 3, no. 1, pp. 9–15, 2017.
- [7] . Hasymi Ali, *Pengantar Asuransi*, Bumi Aksara, Jakarta, 1993.
- [8] Saruni D & Yudo Devianto. (2019). *Utilization of Prediction Data for Prospective Decision Customers Insurance Using the Classification Method of C.45 and Naive Bayes Algorithms. Journal of Phys.: Conf. Ser.* **1179** 012023
- [9] Natalius, Samuel, 2010, *Metode Naive Bayes Classifier dan Penggunaannya Pada Klasifikasi Dokumen*, Program Studi Sistem dan teknologi Informasi, Sekolah Teknik Elektro dan Informatika, Institut Teknologi Bandung.
- [10] Manning, Christopher D. Raghavan, Prabhakar, Schütze, Hinrich (2009) *An Introduction to Information Retrieval* Cambridge University Press, Cambridge, England
- [11] Saleh, A. (2015) *Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga* Citec Journal, Vol. 2, No. 3, ISSN: 2354-5771
- [12] Abbas Salim, *Dasar-dasar Asuransi*, Tarisito, Bandung 1985.
- [13] Kuncahyo S Nugroho (2019), *Confusion Matrix* untuk Evaluasi Model pada *Supervised Learning*.