

PEMILIHAN JENIS ASURANSI BERDASARKAN DEMOGRAFI CALON PEMEGANG POLIS DENGAN METODE *NAÏVE BAYES CLASSIFIER*

Lailatul M. Chaira¹, Nurul Hidayat², Inu L. Wibowo³, Imam Mukhlash⁴

^{1,2,3,4}Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember Surabaya

⁴imamm@matematika.its.ac.id

Abstrak

Asuransi merupakan salah satu cara untuk memproteksi diri di masa depan. Saat ini, perusahaan asuransi berlomba-lomba untuk menawarkan produk asuransi yang menjanjikan. Dalam rangka bersaing dengan kompetitor lainnya dan demi memenuhi kebutuhan nasabahnya, perusahaan asuransi memerlukan strategi bisnis yang bijak dan tepat agar produknya mendapat respon positif dari calon nasabah. Salah satu permasalahan dalam bidang asuransi adalah bagaimana menentukan jenis asuransi yang tepat untuk calon nasabah. Pada paper ini, dibahas tentang bagaimana menentukan jenis asuransi yang tepat menggunakan task dalam data mining untuk menggali informasi yang berkaitan dengan kebutuhan produk asuransi bagi calon nasabah. Metode yang digunakan untuk klasifikasi adalah Naïve Bayes Classifier. Hasil uji coba menunjukkan bahwa metode NBC mampu mengklasifikasi record dengan tingkat kinerja tertinggi sebesar 94.12% ketika proporsi pembagian data latih 90% dan data uji sebesar 10%. Karena kinerja sistem yang dihasilkan dapat dikatakan baik, sistem dianggap kredibel untuk merekomendasikan produk asuransi kepada calon nasabah

Katakunci: *Asuransi, Data Mining, Klasifikasi, Naïve Bayes Classifier*

1 Pendahuluan

Proteksi diri, baik jiwa maupun aset berharga, merupakan hal yang sangat penting untuk kehidupan individu di lingkungan yang serba beresiko saat ini. Asuransi dapat dijadikan tabungan masa depan maupun perlindungan diri ketika mengalami sakit kritis, kecelakaan, atau bahkan kematian. Saat ini perusahaan asuransi berlomba-lomba untuk menawarkan

produk asuransi yang menjanjikan. Dalam rangka bersaing dengan kompetitor lainnya dan demi memenuhi kebutuhan nasabahnya, perusahaan asuransi memerlukan strategi bisnis yang bijak dan tepat agar produknya mendapat respon positif dari calon nasabah. Oleh karena itu, diperlukan teknik data mining untuk menggali informasi yang berkaitan dengan kebutuhan produk asuransi bagi calon nasabah. Informasi yang didapat dapat berupa pola spesifik ciri demografi nasabah yang biasanya mengambil produk asuransi tertentu. Informasi yang demikian akan membantu perusahaan asuransi dalam menawarkan produk asuransi yang sesuai bagi calon nasabahnya secara ramah dan kredibel. Sehingga diharapkan dapat memberikan dampak positif bagi perusahaan asuransi sendiri. Untuk mendapatkan informasi yang berkualitas dibutuhkan teknik data mining, yaitu klasifikasi.

Dalam penelitian ini, metode klasifikasi yang digunakan untuk membangun sebuah model yang akan dipakai untuk menentukan solusi produk asuransi terbaik bagi calon nasabah berdasarkan ciri demografinya adalah Naïve Bayes Classifier (NBC). Metode klasifikasi ini diturunkan dari penerapan teorema Bayes dengan asumsi saling bebas satu sama lain berdasarkan atribut kelas. Naïve Bayes Classifier dinilai berpotensi baik dalam mengklasifikasi data dibandingkan beberapa metode pengklasifikasian yang lain dalam hal akurasi dan efisiensi komputasi dikarenakan sifat keindependensian fitur-fiturnya [3]. Metode ini akan menghasilkan perhitungan peluang bersyarat setiap atribut terhadap kelas klasifikasi serta peluang prior setiap kelas klasifikasi. Perhitungan yang sudah didapat tersebut akan disimpan dalam sebuah tabel yang dinamakan tabel probabilitas. Tabel probabilitas ini akan digunakan untuk menghitung peluang posterior jika ada data nasabah yang belum diketahui kelas klasifikasinya. Peluang posterior terbesar akan dijadikan kelas klasifikasi bagi nasabah tersebut. Sehingga, tabel probabilitas yang dihasilkan dapat digunakan perusahaan asuransi sebagai bantuan untuk merekomendasikan produk asuransi yang sesuai bagi calon nasabah di masa mendatang. Hal ini tentu akan membawa keuntungan yang besar bagi kedua pihak, baik dari pihak perusahaan asuransi maupun bagi calon nasabahnya.

Data demografi nasabah asuransi yang digunakan berasal dari salah satu perusahaan asuransi yang berada di Indonesia selama kurun waktu 2 tahun, dataset ini berisi 518 record polis asuransi dengan 19 atribut. Record ini mengandung informasi ciri demografi nasabah asuransi yang meliputi nomor polis, nama, jenis kelamin, alamat, agama, kewarganegaraan, tanggal lahir, usia masuk asuransi, pekerjaan, bidang pekerjaan, gaji pemegang rekening, status pernikahan, status merokok, jumlah anak, jumlah tanggungan, jumlah premi, jumlah uang pertanggungan, riwayat penyakit pre-existing, serta produk asuransi. Banyaknya jumlah atribut yang digunakan dalam proses klasifikasi akan menyebabkan waktu klasifikasi menjadi lama, sehingga perlu adanya reduksi atribut yang tidak diperlukan agar proses klasifikasi menjadi lebih efisien. Maka dari itu, peneliti akhirnya memilih 9 atribut akan dijadikan predictor variable, antara lain jenis kelamin,

usia masuk asuransi, pekerjaan, bidang pekerjaan, gaji pemegang rekening, status pernikahan, jumlah anak, status anak, status penyakit pre-existing dan 1 atribut sebagai predicted variable yakni atribut produk asuransi.

2 Preliminaries

2.1 Penelitian Terkait

Penulis memulai penelitian ini dengan terlebih dahulu melakukan studi kepustakaan dari penelitian-penelitian dan sumber-sumber lain. Penelitian tersebut membahas tentang topik yang terkait dengan penelitian penulis, antara lain mengenai masalah yang diangkat serta mengenai metode yang digunakan penulis.

Azuralia Abu Bakar dkk (2010) melakukan penelitian tentang metode dalam data mining untuk mengklasifikasikan data nasabah ke dalam 5 kelas produk asuransi dengan metode asosiatif rule. Sistem yang dihasilkan memiliki kinerja sebesar 58% [1]. Sedangkan Bi Xing dan Huang Feng-wen (2009) melakukan penelitian yang bertujuan untuk merekomendasikan produk asuransi bagi calon nasabah dengan pendekatan Algoritma Backpropagation. Sistem rekomendasi kredibel dihasilkan dari penelitian ini, namun calon nasabah hanya diberikan 1 rekomendasi produk asuransi [2]. Sementara itu, beberapa penelitian terkait pengimplementasian Metode Naïve Bayes Classifier diantaranya yakni penelitian yang dilakukan Bustami (2013) menerapkan Algoritma Naïve Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. Penelitian ini melibatkan 20 dataset nasabah asuransi untuk memprediksi kelancaran pembayaran premi di masa mendatang [7]. Sedangkan penelitian yang dilakukan oleh Murat Karatabak (2015) mengimplementasikan algoritma NBC untuk mengklasifikasikan pasien yang terdiagnosa kanker payudara ke dalam 2 stadium, yaitu tergolong kanker jinak atau kanker ganas. Hasil penelitiannya menunjukkan bahwa keakurasian klasifikasi dengan menggunakan metode Naïve Bayes Classifier yang diperoleh mencapai 96,17% [3].

2.2 Naïve Bayes Classifier

Metode Naive Bayes Classifier merupakan salah satu metode yang terdapat pada teknik klasifikasi. Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan 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 [4, 5, 6].

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^q P(X_i|Y)}{P(X)} \quad (1)$$

dengan $P(Y|X)$ adalah probabilitas data dengan vektor X pada kelas Y . $P(Y)$ adalah probabilitas awal kelas Y . $\prod_{i=1}^q P(X_i|Y)$ merupakan probabilitas independen kelas Y dari semua fitur dalam vektor X . Nilai $P(X)$ selalu tetap sehingga dalam perhitungan prediksi nantinya kita tinggal menghitung bagian $P(Y) \prod_{i=1}^q P(X_i|Y)$ dengan memilih yang terbesar menjadi kelas yang dipilih sebagai hasil prediksi.

2.3 Laplacian Smoothing

Teknik Laplacian Smoothing digunakan untuk mengatasi nilai probabilitas kondisional pada Naïve Bayes Classifier yang dapat bernilai 0. Cara yang digunakan pada teknik ini adalah dengan menambahkan angka 1 pada perhitungan Likelihood. Persamaan di bawah ini menunjukkan perhitungan nilai Likelihood untuk algoritma Naïve Bayes Classifier [8]

$$P(X_i|Y) = \frac{1 + n(X_i, Y)}{|W| + n(Y)} \quad (2)$$

dengan

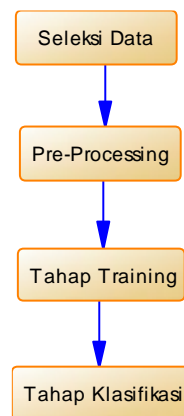
$n(X_i, Y)$ = Jumlah term X_i yang ditemukan di seluruh data pelatihan dengan kategori Y .

$n(Y)$ = Jumlah term di seluruh data pelatihan dengan kategori Y .

$|W|$ = Jumlah seluruh term X_i dari seluruh data pelatihan

3 METODOLOGI PENELITIAN

Secara umum, skema metodologi penelitian ini digambarkan pada Gambar 1 sebagai berikut:



Gambar 1 Diagram Alir Metode Penelitian

3.1 Seleksi Data

Penelitian ini menggunakan metode Naïve Bayes Classifier untuk mengklasifikasikan data nasabah asuransi ke dalam produk asuransi yang sesuai dengan ciri demografinya. Dimana, nasabah akan diklasifikasikan ke dalam salah satu 4 kelas produk asuransi. 4 produk asuransi yang digunakan sebagai kelas klasifikasi yaitu Maestrolink Plus (MLP), Super Protection Plan (SPP), Smart Kidz (SK), dan Maestro Hospital Plan (MHP). Dalam penelitian ini, dataset yang digunakan berjumlah 518 record dan diambil dari salah satu perusahaan asuransi yang ada di Indonesia selama kurun waktu 2 tahun, yaitu mulai Januari 2014 - Januari 2016. Dimana 117 data nasabah yang memilih produk asuransi MLP, 142 data nasabah yang mengambil produk asuransi SPP, 108 data nasabah yang mengambil produk asuransi jenis SK, dan 97 data nasabah yang memilih produksi asuransi MHP.

Data demografi yang digunakan dalam penelitian ini adalah sebagai berikut:

a. Atribut Kategorik

1. Jenis Kelamin: Laki-laki atau Perempuan
2. Pekerjaan: Belum Bekerja, Mahasiswa/Pelajar, Karyawan, Ibu Rumah Tangga, TNI/Polri, PNS, atau Pemilik Usaha
3. Bidang Pekerjaan: Jasa/Keuangan, Pemerintahan, Transportasi, Perdagangan, Konstruksi, Manufaktur, Sumber Daya Alam, atau Lainnya
4. Status Pernikahan: Menikah, Belum Menikah atau Duda/Janda.
5. Status Perokok: Y bagi perokok dan T bagi bukan perokok
6. Kondisi Pre-Existing: Y bagi yang memenuhi kondisi *Pre-Existing* dalam riwayat kesehatannya dan T bagi yang tidak memiliki kondisi *Pre-Existing*.
7. Jumlah Anak: Dimana nasabah akan dibagi kedalam 3 kelas jumlah anak, 'Tidak Ada' jika nasabah tidak memiliki anak, 'Normal' jika nasabah memiliki anak antara 1-3 dan 'Banyak' untuk nasabah yang memiliki anak lebih dari 3.
8. Plan Asuransi: Produk asuransi yang diambil oleh nasabah, yaitu MLP, SPP, SK, MHP.

b. Atribut Numerik

1. Usia: Merupakan usia ketika pertama kali masuk sebagai nasabah asuransi
2. Gaji Pemegang Rekening: Merupakan rata-rata gaji per-bulan pembayar premi.

Profil demografi ini di representasikan dalam sebuah vektor $\vec{x} = \{x_1, x_2, \dots, x_9\}$ yang mengkarakteristikan seorang nasabah asuransi. Pendekatan *Naïve Bayes* dipakai untuk mengestimasi peluang posterior maksimum nasabah ke-*i* terhadap kelas produk asuransi MLP, SPP, SK, atau MHP berdasarkan vektor \vec{x} . Tujuan akhir penelitian ini adalah untuk

menghasilkan sistem rekomendasi produk asuransi yang sesuai bagi calon nasabah asuransi berdasarkan ciri demografinya

3.2 Pemrosesan Awal Data

Pada umumnya data mentah yang diperoleh masih mengandung beberapa data yang tidak bisa digunakan dalam proses *data* mining dikarenakan adanya kesalahan atau anomali pada data. Sehingga diperlukan pemrosesan awal data atau *Pre-processing* sebelum melakukan penggalian informasi terhadap data. Dimana data yang sudah diperoleh akan di normalisasi, dengan kata lain data akan dihilangkan *noise*-nya. Dalam penelitian ini *pre-processing* dimulai dengan penghapusan duplikasi data, apabila terdapat 2 atau lebih data dengan nomor polis yang sama dalam satu dataset maka salah satu data tersebut harus dihilangkan. Selanjutnya adalah melakukan penanganan nilai yang hilang, *missing value* ini bisa terjadi dikarenakan beberapa hal, bisa jadi karena memang nasabah tidak memilikinya, nasabah lupa mengisikan atau kesalahan inputasi pada sistem database perusahaan. Untuk beberapa data yang salah satu atributnya tidak terisi akan langsung diganti nilai rata-ratanya bagi atribut yang bersifat numerik (kontinyu) sedangkan untuk atribut yang tergolong kategorikal, nilai yang hilang akan diganti dengan modus dalam atribut tersebut.

3.3 Tahap Pelatihan

Setelah tidak ada *noise* dalam dataset, tahap yang selanjutnya adalah melakukan proses pelatihan. Tahapan ini terdiri dari beberapa langkah:

1. Mencari Peluang *Prior*.

Untuk mencari nilai peluang prior $P(Y)$ tiap kelas, dapat dicari dengan menghitung pecahan tiap data l yang dimiliki tiap kelas.

2. Mencari nilai *likelihood*

a. *Likelihood* atribut Kategorikal

Pada atribut kategori x_i , peluang bersyarat $P(X_i = x_i | Y = y)$, $i = 1, 2, 3, \dots$ dicari menurut pecahan data *training* pada kelas y yang memuat nilai atribut X_i . Namun ketika $P(X_i = x_i | Y = y) = 0$ maka $P(X_i = x_i | Y = y)$ akan dihitung menggunakan rumus *laplacian smoothing*.

b. *Likelihood* atribut Numerik

Cara untuk mencari peluang bersyarat $P(X_i = x_i | Y = y)$ untuk atribut yang bersifat numerik atau kontinyu, diasumsikan bentuk tertentu distribusi peluang untuk variable kontinyu dan mengestimasi parameter distribusi menggunakan data training. Distribusi Gaussian sering digunakan untuk merepresentasikan peluang bersyarat untuk atribut numerik. Distribusi dikarakterisasi dengan 2 parameter yaitu *mean* (μ) dan *varian* (σ^2). Untuk kelas y_i , peluang bersyarat untuk atribut x_i adalah:

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sigma_{ij}\sqrt{2\pi}} \exp - \frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \quad (3)$$

Parameter μ_{ij} dapat diestimasi berdasarkan sampel *mean* dari $X_i(\rightarrow)$ untuk seluruh data training yang dimiliki kelas y_j . Dengan cara yang sama σ_{ij}^2 dapat diestimasi dari sampel *varian* (s^2) data latih tersebut.

3.4 Tahap Klasifikasi

Tahap ketiga yaitu tahap klasifikasi. Setelah model prediksi telah dibangun pada data training, saatnya mengklasifikasi data yang belum diketahui label kelasnya.

Langkah-langkah pada tahap ini adalah:

1. Penentuan peluang bersyarat $P(X_i = x_i | Y = y)$ sesuai dengan hasil training dimana x_i sesuai dengan *input user*.
2. Mencari $\prod_{i=1}^d P(X_i = x_i | Y = y)$
3. Mencari peluang posterior $P(Y|X) = \frac{P(Y) \prod_{i=1}^d P(X_i|Y)}{P(X)}$
4. Kondisi untuk mengklasifikasi *record*
 - a. Jika $P(MLP|X) > P(SPP|X), P(MLP|X) > P(SK|X), P(MLP|X) > P(MHP|X)$
maka *record* diklasifikasikan masuk kelas MLP
 - b. Jika $P(SPP|X) > P(MLP|X), P(SPP|X) > P(SK|X), P(SPP|X) > P(MHP|X)$
maka *record* akan diklasifikasikan masuk kelas SPP.
 - c. Jika $P(MHP|X) > P(SPP|X), P(MHP|X) > P(SK|X), P(SPP|X) > P(MLP|X)$
maka *record* akan diklasifikasikan masuk kelas MHP.
 - d. Jika $P(SK|X) > P(SPP|X), P(SK|X) > P(MHP|X), P(SK|X) > P(MLP|X)$
maka ia tergolong dalam kelas SK. Namun, apabila peluang posterior maksimum berada pada kelas SK tetapi usia nasabah lebih dari 21 tahun atau sudah bekerja, maka nasabah tersebut tidak memenuhi syarat untuk mengambil produk asuransi SK, sehingga ia akan di rekomendasikan pada kelas peluang posterior maksimum ke dua.

3 Hasil Penelitian

Pengujian pertama yang akan diamati merupakan hasil penerapan tingkat presisi. Pengujian tingkat presisi merupakan perbandingan prosentase antara jumlah instance pada data testing yang mampu diklasifikasikan dengan benar dengan keseluruhan jumlah testing instance.

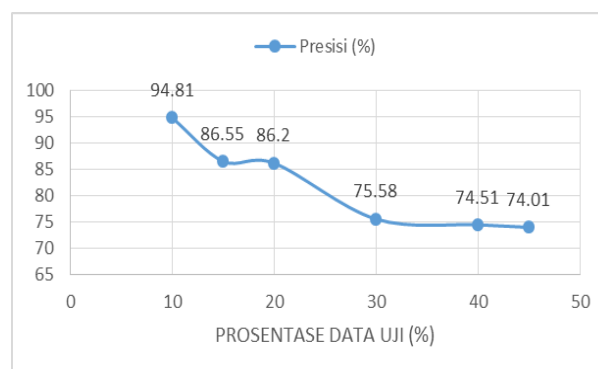
Hasil yang di dapat dengan melakukan 10-20 kali percobaan dengan perbandingan data training dan data testing yang berbeda, didapat rata-rata presisi pengujian tertulis dalam Tabel 1.

Berdasarkan Tabel 1 dan Gambar 2 dapat disimpulkan bahwa bahwa nilai presisi klasifikasi dengan Metode Naïve Bayes Classifier akan cenderung menurun seiring dengan bertambahnya data yang digunakan dalam pengujian. Hasil presisi tertinggi didapatkan ketika menggunakan perbandingan 90% data latih dan 10% data uji, yaitu dengan rata-rata presisi sebesar 94.81% dan rata-rata waktu eksekusi selama 21.62 detik.

Dari Tabel 1 juga dapat disimpulkan bahwa semakin banyak data yang digunakan sebagai data uji, maka proses testing juga akan semakin lama dikarenakan sistem harus melakukan perulangan perhitungan sejumlah data yang dijadikan data uji.

Tabel 1.
Hasil Presisi dan Running Time dengan *Naïve Bayes Classifier*

Data Training	Data Testing	Presisi	Running Time
467(90%)	51(10%)	94.81%	21.62 detik
441(85%)	77(15%)	86.55%	30.5 detik
415(80%)	103(20%)	86.2%	40.48 detik
363(70%)	155(30%)	75.58%	51.3 detik
311(60%)	207(40%)	74.51%	57.37 detik
85(55%)	233(45%)	74.01%	66.3 detik



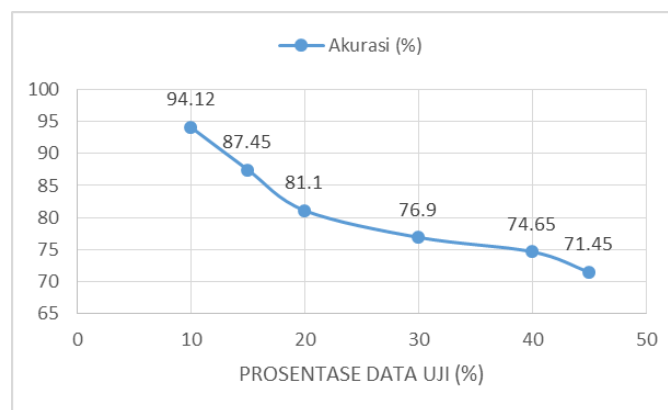
Gambar 2 Grafik Prosentase Data Uji Terhadap Rata-rata Presisi

Pengujian kedua yang akan diamati adalah hasil tingkat akurasi. Pengujian akurasi merupakan perbandingan antara jumlah prosentasi jumlah *instance* yang benar setiap kelas klasifikasi dengan jumlah kelas klasifikasi.

Hasil rata-rata akurasi yang diperoleh setelah melakukan 10-20 kali percobaan dengan perbandingan data training dan data testing. Hasil pengujian selengkapnya ditunjukkan pada Tabel 2.

Tabel 2.
Hasil Total Akurasi Klasifikasi dengan metode *Naïve Bayes*

Data Training	Data Testing	MLP (%)	SPP (%)	SK (%)	MHP (%)	Akurasi
467 (90%)	51 (10%)	91	92.9	100	92.58	94.12%
441 (85%)	77 (15%)	88.9	75.3	100	82.6	87.45%
415 (80%)	103(20%)	85.3	53.39	100	85.71	81.1%
363(70%)	155(30%)	84.6	59.38	88.3	67.48	76.9%
311(60%)	233(45%)	86	60.54	97.2	54.86	74.65%
285(55%)	233(45%)	82.72	54.1	98.4	50.59	71.45%



Gambar 3 Grafik Prosentase Data Uji Terhadap Rata-rata Akurasi

Berdasarkan Tabel 2 dan Gambar 3 dapat disimpulkan disimpulkan bahwa rata-rata akurasi klasifikasi dengan Metode *Naïve Bayes Classifier* akan cenderung menurun seiring dengan bertambahnya data yang digunakan dalam pengujian. Dari Tabel 2 dapat juga diketahui bahwa ketika melakukan proses *mining* dengan menggunakan proporsi data latih kurang dari 85%, *Naïve Bayes* tidak cukup baik dalam mengklasifikasikan data uji yang terletak pada kelas SPP. Ini dikarenakan karakteristik demografi nasabah yang memilih produk SPP hampir mirip dengan nasabah yang mengambil produk SK dan kebanyakan nasabah yang mengambil SPP yang masih berusia dibawah 21 tahun dan pekerjaannya masih belum bekerja atau seorang pelajar/mahasiswa akan diklasifikasikan sistem ke dalam kelas SK. Sedangkan rata-rata hasil akurasi tertinggi yaitu 94.12%.

Nilai presisi dan akurasi yang di dapat memang tidak mampu mencapai 100%. Hal ini dikarenakan seorang nasabah bisa mengambil lebih dari 1 produk asuransi. Oleh karena itu, dalam implementasi sistem rekomendasi produk asuransi yang dibuat, sistem tidak hanya merekomendasikan satu jenis rekomendasi bagi seorang calon nasabah, melainkan sistem juga akan merekomendasikan semua produk asuransi secara berurutan berdasarkan

peringkat probabilitas posterior tiap kelas produk asuransi dan syarat khusus dari produk yang direkomendasikan harus memenuhi kriteria demografi calon nasabah.

Melihat hasil rata-rata presisi dan akurasi sistem tertinggi dengan menggunakan metode *Naïve Bayes Classifier* yaitu ketika menggunakan perbandingan 90:10 untuk data training dan data uji, maka tabel probabilitas yang didapat dari proses pelatihan tersebut dapat digunakan sebagai acuan untuk merekomendasikan produk asuransi yang dinilai sesuai untuk calon nasabah.

4 Kesimpulan

Dari hasil pengujian memperlihatkan hasil eksperimen dengan berbagai proporsi antara data training dan data uji. Presisi dan total akurasi tertinggi *Naïve Bayes Classifier* diperoleh ketika menggunakan 90% data latih dan 10% data uji dengan rata-rata tingkat presisi sebesar 94.81% dan akurasi sebesar 94.12%. Dari hasil percobaan di atas, dapat disimpulkan bahwa *Naïve Bayes Classifier* dapat mampu membantu merekomendasikan produk asuransi yang tepat kepada calon nasabah berdasarkan ciri demografinya.

5 Daftar Pustaka

- [1] Bakar, A. A., Othman, Z., Yusoff M. S. N. M., & Ismail, R., "Development of Knowledge Model for Insurance Product Decision using the Associative Classification Approach.," in 10th International on Intelligent System Design and Application, Malaysia (2010).
- [2] Xing, B., Jian, L., & Feng-wen, H., "The Application of Improved BP Algorithm in Customer Classification of Life Insurance," in 16th International Conference on Management Science & Engineering, China. (2009).
- [3] Karatabak, M., "A New Classifier for Breast Cancer Detection Based on Naïve Bayes Classifier," *Measurement*, Vol. 72. (2015) 32-36.
- [4] Kusriani & Luthfi, E. T., *Algoritma Data Mining*, Andi Offset, Yogyakarta (2009).
- [5] Kusumadewi, S., "Klasifikasi Status Gizi Menggunakan Naïve Bayes Classification", *Communication and Information Technology Journal.*, Vol: 3 (2009), 6-11.
- [6] Han, J., & Kamber, M., *Data Mining: Concepts and Techniques*, 3rd Edition. Morgan Kaufman Publisher, San Fransisco (2012).
- [7] Bustami, B. Penerapan Algoritma Naïve Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. *TECHSI. Jurnal Penelitian Teknik Informatika*, Vol: 3 No. 2 (2013) 127-146.
- [8] Saptano R., Wiranto, Suryono W. D. Sistem Klasifikasi Keluhan Di UPT TIK UNS Menggunakan Algoritma Naive Bayes Classifier. Seminar Nasional Teknologi Informasi dan Komunikasi. Yogyakarta. (Maret 2016)