

PREDIKSI STATUS POLIS NASABAH ASURANSI JIWA MENGUNAKAN METODE NAIVE BAYES

Zulianda Saputra¹, Arief Wibowo ^{*2}

^{1,2}Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ¹zuliandasaputra1107@gmail.com, ²arief.wibowo@budiluhur.ac.id

(* : corresponding author)

Abstrak- Masalah yang sering kali dijumpai oleh para perusahaan asuransi yaitu banyaknya nasabah asuransi yang mengalami kemacetan dalam memperpanjang polis asuransi mereka. Memprediksi para nasabah asuransi yang dikategorikan sebagai kelompok yang macet polis dan lancar polis merupakan suatu hal yang diperlukan bagi suatu perusahaan asuransi dalam menentukan nasabahnya. Perusahaan-perusahaan yang mengambil bidang asuransi masih sangat kesulitan dalam memprediksi calon nasabahnya termasuk kategori lancar polis atau macet polis. Oleh karena itu, dibutuhkan suatu sistem prediksi yang dikhususkan untuk memprediksi para calon nasabah asuransi. Sistem prediksi ini berbasis dekstop dengan memanfaatkan algoritme Naive Bayes dalam memprediksi calon nasabah asuransi. Dalam pengujian yang telah dilakukan menggunakan Confusion matrix didapatkan hasil akhir dengan nilai akurasi sebesar 81%, recall sebesar 0,99, dan Precision sebesar 0,81 pada sistem prediksi yang dibuat dengan jumlah data sebanyak 2000 data nasabah. Penelitian ini membuat sistem prediksi status polis data nasabah asuransi dengan harapan dapat membantu para perusahaan dalam mengklasifikasikan para calon nasabah asuransi.

Kata Kunci: Prediksi, Naive Bayes, Status Polis

PREDICTION OF POLICY STATUS CUSTOMERS USING THE NAIVE BAYES METHOD

Abstract- The problem that is often encountered by insurance companies is that many insurance customers experience bottlenecks in extending their insurance policies. Predicting insurance customers who are categorized as groups with bad policies and smooth policies is something that is needed for an insurance company in determining its customers. Companies that take up insurance are still having a hard time predicting whether their prospective customers are in the current or non-performing policy category. Therefore, a prediction system is needed that is specifically designed to predict prospective insurance customers. This prediction system is desktop-based by utilizing the Naive Bayes algorithm to predict potential insurance customers. In the tests that have been carried out by researchers using the Confusion matrix, the final results are obtained with an accuracy value of 81%, recall of 0.99, and Precision of 0.81 on a prediction system made with a total of 2,000 customer data. In this study, researchers created a policy status prediction system for insurance customer data in the hope of helping companies determine the attributes of customer data they are looking for to offer insurance programs.

Keywords: Predictions, Naive Bayes, Policy Status

1. PENDAHULUAN

Asuransi merupakan sarana finansial dalam tata kehidupan rumah tangga, baik dalam menghadapi risiko yang mendasar seperti risiko kematian, atau dalam menghadapi risiko atas harta benda yang dimiliki [1]. Menurut Asosiasi Asuransi Jiwa Indonesia (AAJI), produk asuransi jiwa adalah janji yang tertulis di dalam polis asuransi yang dibuat oleh penanggung kepada tertanggung untuk memberikan kompensasi keuangan apabila sesuatu terjadi kepada tertanggung [2].

Masalah yang sering kali dijumpai oleh para perusahaan asuransi yaitu banyaknya nasabah asuransi yang mengalami kemacetan dalam memperpanjang polis asuransi mereka. Hal tersebut biasanya sering terjadi dikarenakan berbagai faktor, diantara seperti masalah ekonomi, atau mereka menemukan asuransi lain yang memberikan keuntungan manfaat produk asuransi yang lebih menguntungkan atau mereka memang tidak memahami manfaat apa saja yang diberikan oleh suatu perusahaan asuransi atau berupa faktor lainnya.

Berdasarkan hasil survei Populix, ada 67% responden di Indonesia yang telah memiliki asuransi. Responden dari jumlah itu, 83% di antaranya memiliki asuransi dari pemerintah berupa BPJS Kesehatan. Sedangkan, sisanya sebesar 38% memilih asuransi swasta. Mereka lebih banyak menggunakan asuransi kesehatan, asuransi jiwa, dana pensiun, serta asuransi pendidikan. Sementara, 33% responden mengaku tak memiliki asuransi.

Mayoritas atau 57% di antaranya tak memiliki uang untuk membayar premi. Sebanyak 29% responden yang tak punya asuransi karena menganggap premi terlalu mahal. Kemudian, 25% responden yang tak memiliki asuransi karena tak tahu apa saja manfaat yang bisa didapatkan. Sebagai catatan, Populix melakukan survei tersebut terhadap 1.041 responden di Indonesia [3].

Memprediksi para nasabah asuransi yang dikategorikan sebagai kelompok yang macet polis dan lancar polis merupakan suatu hal yang diperlukan bagi suatu perusahaan asuransi dalam menentukan nasabahnya. Mengetahui nasabah tersebut merupakan kelompok macet ataupun lancar dengan melihat kriteria nasabah tersebut bisa dilakukan dengan menggunakan algoritme Naive Bayes.

Pengklasifikasi bayesian adalah pengklasifikasi statistik dan didasarkan pada teorema bayes. Teori keputusan bayes adalah pendekatan statistik yang fundamental dalam pengenalan pola (*Pattern Recognition*), penggunaan algoritme ini dalam hal klasifikasi harus mempunyai masalah yang bisa dilihat statistiknya. *Naive Bayes Classifier* mengestimasi peluang kelas bersyarat dengan mengasumsikan bahwa atribut adalah independen secara bersyarat yang diberikan dengan label kelas [4].

Tujuan penelitian ini diharapkan dapat memanfaatkan salah satu algoritme data mining klasifikasi yaitu Naive Bayes untuk dapat memprediksi suatu data nasabah apakah termasuk pemegang polis macet atau polis lancar pada perusahaan - perusahaan asuransi di Indonesia.

2. METODE PENELITIAN

2.1 Import Data

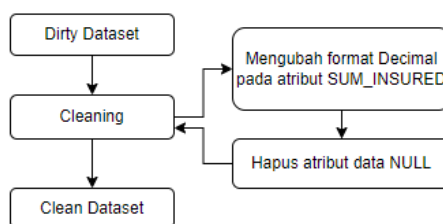
Pada tahap ini, data dirty dataset atau data mentah yang belum melalui proses preprocessing dengan format excell kemudian di import ke dalam tabel pada database. Dirty dataset ini yang kemudian akan dilakukan pelabelan terhadap kategori status polis.

2.2 Labeling

Dataset yang sudah di import ke tabel dalam database kemudian dilakukan proses labeling yang akan diberikan dua kategori label yakni “Lancar” dan “Macet” untuk kebutuhan data training nantinya dan sebagai acuan dalam perhitungan untuk menentukan akurasi dari klasifikasi yang dilakukan.

2.3 Preprocessing

Pada tahapan ini, dilakukan proses *preprocessing* untuk mendapatkan *clean* dataset, sehingga saat proses pengujian dilakukan hasil prediksi status polis menjadi lebih akurat dan tepat.

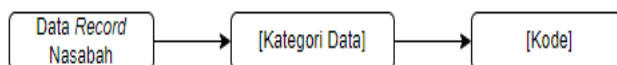


Gambar 1. Proses *Preprocessing*

Pada Gambar 1 berikut ini merupakan ilustrasi dari proses *preprocessing* dimulai dari *dirty* dataset kemudian di lakukan proses *cleaning* sehingga mendapat *clean* dataset.

2.4 Transformation

Transformasi data dilakukan pada beberapa atribut. Transformasi data bertujuan untuk memudahkan dalam perhitungan prediksi status polis dan pengubahan data menjadi kode yang lebih ringkas bertujuan agar bisa dapat melindungi data nasabah yang merupakan data bersifat sensitif.



Gambar 2. Skenario Proses Tranformasi Data

Terlihat pada Gambar 2 bahwa proses tranformasi data dimulai dengan mengkategorikan data kemudian mengubah data menjadi kode yang lebih ringkas

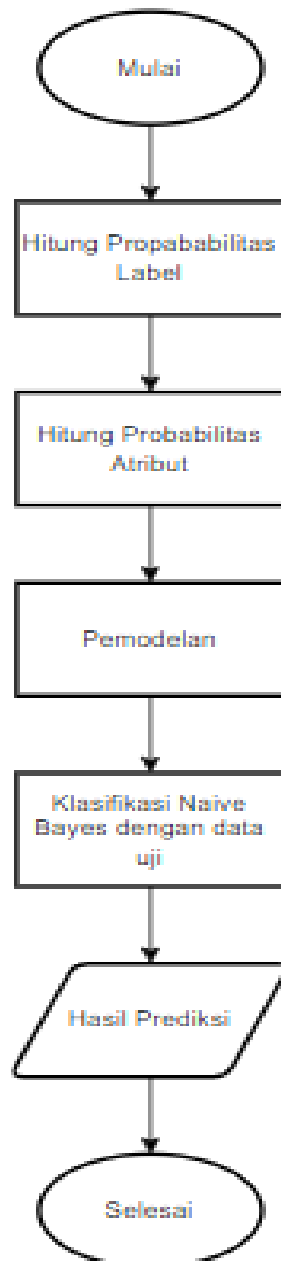
2.5 Classification Naive Bayes

Pada tahap ini, proses perhitungan menggunakan metode *Naive Bayes* terhadap data uji dilakukan dengan melalui sub proses. Proses ini hanya dapat dilakukan apabila tahapan *preprocessing* dan *labeling* sudah dilakukan sebelumnya.

Tahap klasifikasi ini memiliki dua langkah dalam prosesnya sebagai berikut:

- Membangun model dengan menganalisis data *training*.
- Kemudian melakukan klasifikasi, dimana model yang telah dihasilkan digunakan untuk melakukan proses klasifikasi terhadap data yang belum diketahui labelnya.

Pada langkah pertama disebut juga sebagai tahap pembelajaran (*learning*) atau data *training*, yang kemudian diproses menggunakan algoritme klasifikasi hingga dibangun sebuah model klasifikasi dengan cara menganalisis data *training* sebelumnya untuk dilakukan proses kedua yang biasanya juga disebut data *testing*.



Gambar 3. Alur Klasifikasi Naive Bayes

Terlihat pada Gambar 3 merupakan alur klasifikasi *Naive Bayes* untuk memprediksi data status polis asuransi, proses membangun model dimulai dari menghitung probabilitas label dan atribut sehingga ditemukan suatu pemodelan, setelah itu dilakukannya klasifikasi sehingga didapatkannya hasil prediksi.

3. HASIL DAN PEMBAHASAN

3.1 Pengujian

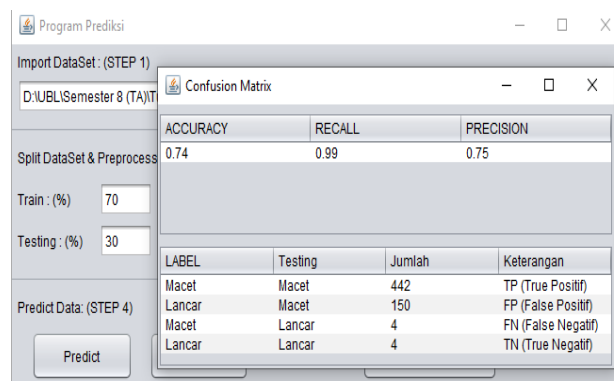
Pengujian pada penelitian ini dilakukan untuk mengetahui seberapa akurat dan sesuai dari program yang dibuat. Adapun penelitian yang dilakukan diantaranya pengujian *Confusion matrix*, pengujian perhitungan klasifikasi *naive bayes* dan pengujian *black box testing*.

3.1.1 Confusion Matrix

Confusion matrix adalah tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah [5]. Proses evaluasi Algoritme *Naive Bayes Classifier* yang dihasilkan, digunakan alat analisis yaitu *Confusion matrix* untuk menggambarkan performa dari sebuah model atau algoritme secara spesifik [6].

Split Validation merupakan teknik validasi untuk membagi data *training* dan data *testing*. Membagi dengan melakukan percobaan *training* berdasarkan *Split Ratio* yang sudah ditentukan sebelumnya dengan menggunakan *Split Validation*, kemudian sisa dari *Split Ratio* data *testing* adalah data yang akan menjadi data pengujian kebenaran atau keakuratan hasil pembelajaran dan belum pernah digunakan dalam pembelajaran [7].

Pada pengujian pertama dengan pembagian 70% untuk Train dan 30% untuk Testing, dari jumlah data sebanyak 2000 data didapat Training sebanyak 1400 data dan Testing 600 data. Pengujian pertama menghasilkan nilai Accuracy yaitu 0,74 dengan Recall 0,99 dan Precision 0,75. Gambar 4 merupakan hasil dari prediksi menggunakan program pada pengujian pertama.

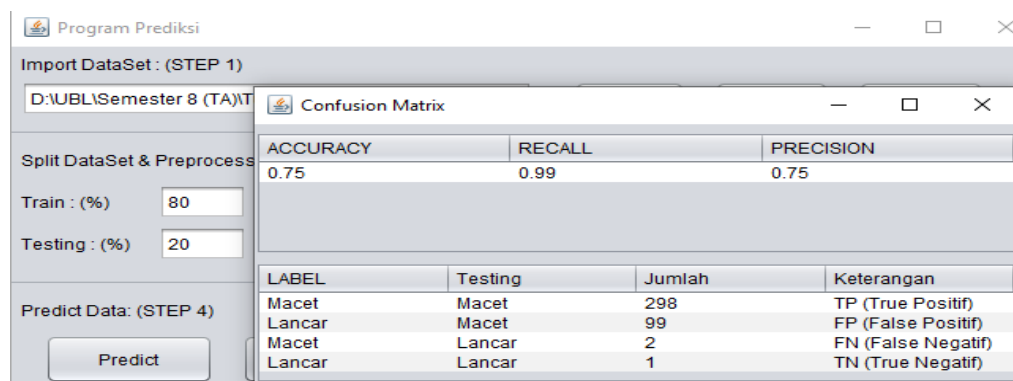


ACCURACY	RECALL	PRECISION
0.74	0.99	0.75

LABEL	Testing	Jumlah	Keterangan
Macet	Macet	442	TP (True Positif)
Lancar	Macet	150	FP (False Positif)
Macet	Lancar	4	FN (False Negatif)
Lancar	Lancar	4	TN (True Negatif)

Gambar 4. Pengujian 1

Pada pengujian kedua dengan pembagian 80% untuk Train dan 20% untuk Testing, dari jumlah data sebanyak 2000 data didapat Training sebanyak 1600 data dan Testing 400 data. Pengujian kedua menghasilkan nilai Accuracy yaitu 0,75 dengan Recall 0,99 dan Precision 0,75. Gambar 5 merupakan hasil dari prediksi menggunakan program pada pengujian kedua.

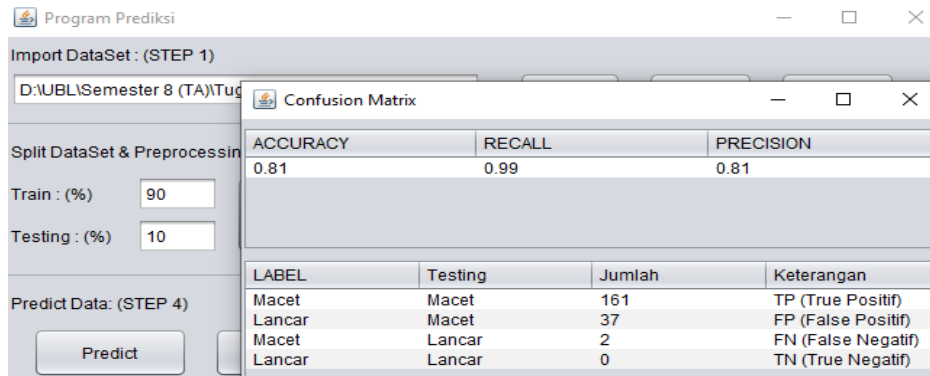


ACCURACY	RECALL	PRECISION
0.75	0.99	0.75

LABEL	Testing	Jumlah	Keterangan
Macet	Macet	298	TP (True Positif)
Lancar	Macet	99	FP (False Positif)
Macet	Lancar	2	FN (False Negatif)
Lancar	Lancar	1	TN (True Negatif)

Gambar 5. Pengujian 2

Pada pengujian ketiga dengan pembagian 90% untuk Train dan 10% untuk *Testing*, dari jumlah data sebanyak 2000 data didapat *Training* sebanyak 1800 data dan *Testing* 200 data. Pengujian ketiga menghasilkan nilai *Accuracy* yaitu 0,81 dengan *Recall* 0,99 dan *Precision* 0,75. Gambar 6 merupakan hasil dari prediksi menggunakan program pada pengujian ketiga.



Gambar 6. Pengujian 3

Tabel 1 merupakan rincian pengujian, terlihat bahwa hasil akhir yang diperoleh dari pengujian yang telah dilakukan didapatkan nilai akurasi terbesar yaitu 0,81 pada pengujian ketiga serta penambahan data training menjadi salah satu faktor nilai akurasi naik dan turun.


Tabel 1. Rincian Pengujian

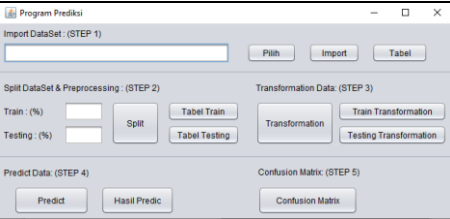
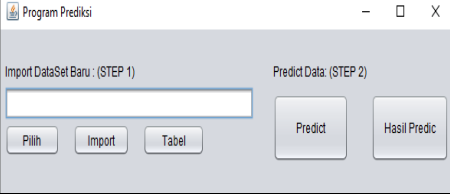
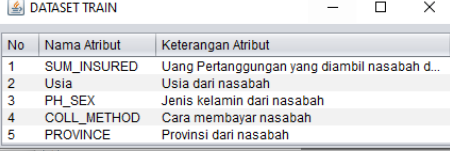
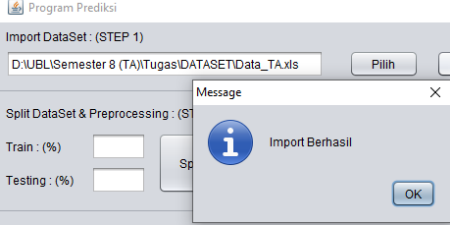
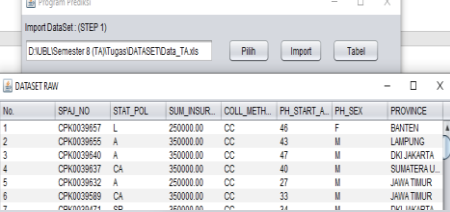
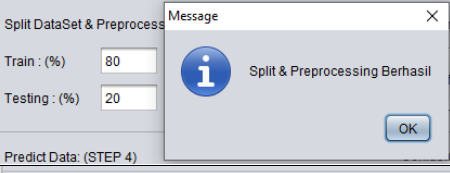
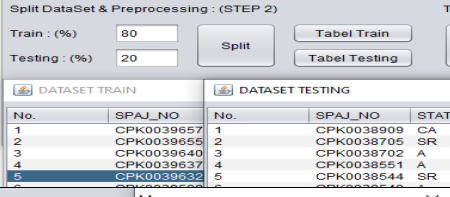
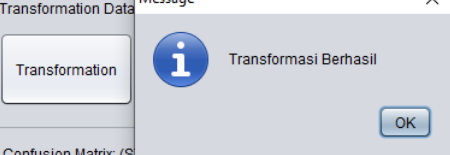
No	Split Validation	Hasil	Referensi
1.	Train 70%, Test 30% Jumlah Data: 2000 Train: 1400 Test: 600	Accuracy: 0,74 Recall: 0,99 Precision: 0,75	Gambar 4.15
2.	Train 80%, Test 20% Jumlah Data: 2000 Train: 1600 Test: 400	Accuracy: 0,75 Recall: 0,99 Precision: 0,75	Gambar 4.16
3.	Train 90%, Test 10% Jumlah Data: 2000 Train: 1800 Test: 200	Accuracy: 0,81 Recall: 0,99 Precision: 0,81	Gambar 4.17

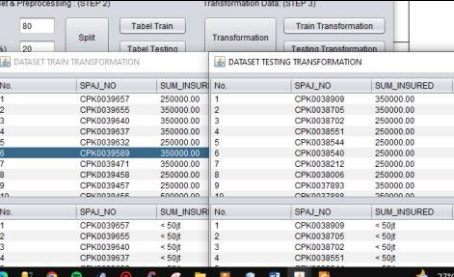
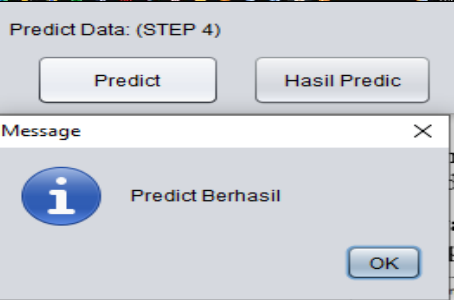
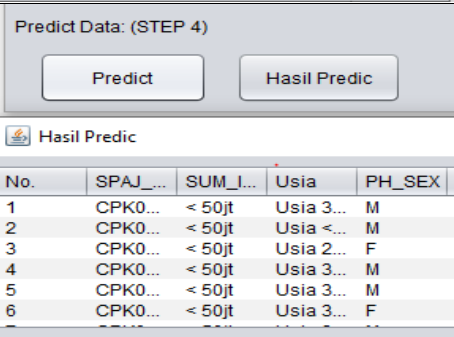
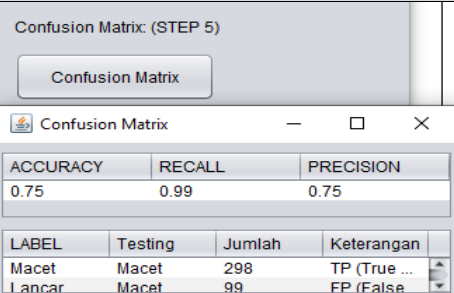
3.1.2 Black Box Testing

Pada pengujian ini bertujuan untuk dapat menguji program aplikasi sesuai dengan fungsi yang telah dibuat, serta memastikan setiap fungsi bekerja sesuai fungsinya.

Tabel 2 Black Box Testing

No.	Komponen yang Diuji	Skenario dan Hasil Uji		
		Tampilan	Hasil Diharapkan	Kesimpulan
1.	Menu Utama		Menampilkan halaman utama	Berhasil

2.	Menu Prediksi		Menampilkan halaman menu prediksi	Berhasil
3.	Menu Prediksi Data Nasabah Baru		Menampilkan halaman menu prediksi data nasabah baru	Berhasil
4.	Master Atribut		Menampilkan tabel atribut	Berhasil
5.	Button Import dataset		Mengimport dataset ke database	Berhasil
6.	Button Tabel Data Raw		Menampilkan Dataset Raw	Berhasil
7.	Button Split dan Preprocessing		Melakukan proses <i>split</i> dan <i>preprocessing</i>	Berhasil
8.	Button Tabel Train dan Testing		Menampilkan tabel data <i>Train</i> dan <i>Testing</i>	Berhasil
9.	Button Transformation data		Melakukan proses transformasi data	Berhasil

10.	<i>Button Tabel Transformation</i>		Menampilkan tabel transformasi data	Berhasil																																			
11.	<i>Button Predict</i>		Melakukan proses <i>predict</i>	Berhasil																																			
12.	<i>Button Hasil Predict</i>	 <table border="1" data-bbox="571 974 1027 1182"> <thead> <tr> <th>No.</th> <th>SPAJ_...</th> <th>SUM_I...</th> <th>Usia</th> <th>PH_SEX</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>CPK0...</td> <td>< 50jt</td> <td>Usia 3...</td> <td>M</td> </tr> <tr> <td>2</td> <td>CPK0...</td> <td>< 50jt</td> <td>Usia <...</td> <td>M</td> </tr> <tr> <td>3</td> <td>CPK0...</td> <td>< 50jt</td> <td>Usia 2...</td> <td>F</td> </tr> <tr> <td>4</td> <td>CPK0...</td> <td>< 50jt</td> <td>Usia 3...</td> <td>M</td> </tr> <tr> <td>5</td> <td>CPK0...</td> <td>< 50jt</td> <td>Usia 3...</td> <td>M</td> </tr> <tr> <td>6</td> <td>CPK0...</td> <td>< 50jt</td> <td>Usia 3...</td> <td>F</td> </tr> </tbody> </table>	No.	SPAJ_...	SUM_I...	Usia	PH_SEX	1	CPK0...	< 50jt	Usia 3...	M	2	CPK0...	< 50jt	Usia <...	M	3	CPK0...	< 50jt	Usia 2...	F	4	CPK0...	< 50jt	Usia 3...	M	5	CPK0...	< 50jt	Usia 3...	M	6	CPK0...	< 50jt	Usia 3...	F	Menampilkan tabel Hasil <i>predict</i> serta tabel perhitungan detail	Berhasil
No.	SPAJ_...	SUM_I...	Usia	PH_SEX																																			
1	CPK0...	< 50jt	Usia 3...	M																																			
2	CPK0...	< 50jt	Usia <...	M																																			
3	CPK0...	< 50jt	Usia 2...	F																																			
4	CPK0...	< 50jt	Usia 3...	M																																			
5	CPK0...	< 50jt	Usia 3...	M																																			
6	CPK0...	< 50jt	Usia 3...	F																																			
13.	<i>Button Confusion Matrix</i>	 <table border="1" data-bbox="571 1332 1027 1473"> <thead> <tr> <th>ACCURACY</th> <th>RECALL</th> <th>PRECISION</th> </tr> </thead> <tbody> <tr> <td>0.75</td> <td>0.99</td> <td>0.75</td> </tr> </tbody> </table> <table border="1" data-bbox="571 1411 1027 1473"> <thead> <tr> <th>LABEL</th> <th>Testing</th> <th>Jumlah</th> <th>Keterangan</th> </tr> </thead> <tbody> <tr> <td>Macet</td> <td>Macet</td> <td>298</td> <td>TP (True ...</td> </tr> <tr> <td>Lancar</td> <td>Macet</td> <td>99</td> <td>FP (False ...</td> </tr> </tbody> </table>	ACCURACY	RECALL	PRECISION	0.75	0.99	0.75	LABEL	Testing	Jumlah	Keterangan	Macet	Macet	298	TP (True ...	Lancar	Macet	99	FP (False ...	Menampilkan hasil <i>confusion matrix</i>	Berhasil																	
ACCURACY	RECALL	PRECISION																																					
0.75	0.99	0.75																																					
LABEL	Testing	Jumlah	Keterangan																																				
Macet	Macet	298	TP (True ...																																				
Lancar	Macet	99	FP (False ...																																				

Dari pengujian yang telah dilakukan, maka didapat kesimpulan bahwa aplikasi dapat berjalan sesuai harapan, dimana fitur maupun fungsi dari setiap menu maupun objek yang ada berfungsi dengan baik dan sesuai dengan tujuan perancangan.

4. KESIMPULAN

Berdasarkan hasil implementasi dan analisis data yang dilakukan dapat ditarik kesimpulan dari pembuatan sistem prediksi status polis nasabah asuransi menggunakan metode Naive Bayes sebagai berikut:

- Keakuratan prediksi selain dengan jumlah atribut yang digunakan juga dipengaruhi oleh tingkat data model, semakin banyak data model maka prediksi akan semakin baik.
- Hasil pengujian yang telah dilakukan mendapatkan hasil seperti yang tertera pada Tabel 1, dari hasil pengujian didapat hasil semakin banyak data training maka nilai akurasi menjadi semakin baik.
- Sistem harus sering melalui tahapan percobaan agar sistem prediksi ini menjadi semakin baik.
- Proses prediksi membutuhkan waktu sedikit lama dikarenakan mengolah data langsung dari SQL Server.

DAFTAR PUSTAKA

- [1] D. Badruzaman, “Perlindungan Hukum Tertanggung Dalam Pembayaran Klaim Asuransi Jiwa,” *Amwaluna: Jurnal Ekonomi dan Keuangan Syariah*, vol. 3, no. 1, pp. 96-118, 2019.
- [2] S. Sulastri, A. Rifin and B. Sanim, “Strategi Pengembangan Asuransi Jiwa,” *Jurnal Aplikasi Bisnis dan Manajemen*, vol. 4, no. 1, pp. 44-51, 2018.
- [3] dataindonesia.id, “dataindonesia,” 2023. [Online]. Available: <https://dataindonesia.id/varia/detail/survei-mayoritas-warga-indonesia-miliki-asuransi-kesehatan>. [Accessed 2023 May 2023].
- [4] M. H. Rifqo and A. Wijaya, “Implementasi Algoritme Naive Bayes Dalam Penentuan Pemberian Kredit,” *Pseudocode*, vol. 4, no. 2, pp. 120-128, 2017.
- [5] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *Jurnal Sains Komputer & Informatika*, vol. 5, no. 2, pp. 697-711, 2021.
- [6] A. M. Pratiwi and A. K. Mutaqim, “Penerapan Algoritme Naïve Bayes Classifier dalam Memprediksi Status Keberlanjutan Polis Nasabah Asuransi PT.X,” *Jurnal Riset Statistika*, vol. 1, no. 2, pp. 117-126, 2017.
- [7] A. S. Febriarini and E. Z. Astuti, “Penerapan Algoritme C4.5 untuk Prediksi Kepuasan Penumpang Bus Rapid Transit (BRT) Trans Semarang,” *Eksplora Informatika*, vol. 8, no. 2, pp. 95-103, 2019.
- [8] ojk.go.id, “Otoritas Jasa Keuangan,” 2017. [Online]. Available: <https://www.ojk.go.id/id/kanal/iknb/pages/asuransi.aspx>. [Accessed 10 May 2023].