

# Pemodelan Prediksi Status Keberlanjutan Polis Asuransi Kendaraan dengan Teknik Pemilihan Mayoritas Menggunakan Algoritma-Algoritma Klasifikasi Data Mining

Dyah Retno Utari<sup>1)</sup>, Arief Wibowo<sup>2)</sup>

<sup>1,2)</sup>Fakultas Teknologi Informasi Universitas Budi Luhur  
Jalan Cileduk Raya, Jakarta Selatan DKI Jakarta

<sup>1)</sup>dyah.retnoutari@budiluhur.ac.id, <sup>2)</sup>arief.wibowo@budiluhur.ac.id

**Abstrak** – Asuransi kendaraan bermotor merupakan jenis usaha pertanggungjawaban terhadap kerugian atau risiko kerusakan yang dapat timbul dari berbagai macam potensi kejadian yang menimpa kendaraan. Persaingan dalam bisnis asuransi khususnya untuk kendaraan bermotor menuntut inovasi dan strategi agar keberlangsungan bisnis tetap terjamin. Salah satu upaya yang dapat dilakukan perusahaan adalah memprediksi status keberlanjutan polis asuransi kendaraan dengan menganalisis data-data profil dan transaksi nasabah. Prediksi terhadap keputusan pemegang polis menjadi sangat penting bagi perusahaan, karena dapat menentukan strategi pemasaran yang mempengaruhi keputusan pelanggan untuk pembaharuan polis asuransi. Penelitian ini telah mengusulkan suatu model prediksi status keberlanjutan polis asuransi kendaraan dengan teknik pemilihan mayoritas dari hasil klasifikasi menggunakan algoritma- algoritma data mining seperti Naive Bayes, Support Vector Machine dan Decision Tree. Hasil pengujian menggunakan confusion matrix menunjukkan nilai akurasi terbaik diperoleh sebesar 93,57%, apapun untuk nilai precision mencapai 97,20%, dan nilai recall sebesar 95,20% serta nilai F-Measure sebesar 95,30%. Nilai evaluasi model terbaik dihasilkan menggunakan pendekatan pemilihan mayoritas (majority voting), mengungguli kinerja model prediksi berbasis pengklasifikasi tunggal.

**Kata kunci:** data mining 1, prediksi keberlanjutan polis asuransi kendaraan 2, majority voting 3, algoritma klasifikasi 4.

**Abstract** – Motor vehicle insurance is a type of business that covers loss or risk of damage that can arise from various potential events that happen to vehicles. Competition in the insurance business, especially for motorized vehicles, demands innovation and strategies to guarantee business continuity. One of the efforts that companies can make is to predict vehicle insurance policies' sustainability status by analyzing customer profile and transaction data. Prediction of the policyholder's decision is essential for the company because it can determine the marketing strategy that influences its decision to renew the insurance policy. This study has proposed a prediction model for vehicle insurance policies' sustainability status with the majority selection technique from the classification results using data mining algorithms such as Naive Bayes, Support Vector Machine, and Decision Tree. The test results using the confusion matrix show that the best accuracy value is obtained at 93.57%, whatever for the precision value reaches 97.20%, and the recall value is 95.20%, and the F-Measure value is 95.30%. The best model evaluation scores are generated using the majority voting approach, outperforming a single classifier-based prediction model's performance.

**Keyword:** data mining, prediction of the sustainability of vehicle insurance policies, majority voting, classification algorithm

## 1 PENDAHULUAN

Asuransi kendaraan bermotor merupakan suatu proteksi yang diberikan perusahaan untuk ditawarkan kepada pemilik kendaraan, dalam penanganan risiko atas kepemilikan kendaraan. Banyak perusahaan asuransi berlomba menawarkan produk proteksi aset tersebut kepada masyarakat baik sebagai pembeli

kendaraan baru maupun untuk perpanjangan polis yang telah dimiliki.

Bagi masyarakat, memiliki asuransi kendaraan bermotor memberikan rasa tenang dalam penggunaan kendaraan tersebut, baik dalam cakupan risiko rusak/hilang sebagian atau keseluruhan (*total lost only*). Namun demikian tentu saja ada anggaran biaya yang harus disediakan untuk membayar premi yang dikenakan perusahaan asuransi.

Di tengah kompetitifnya bisnis asuransi kendaraan, berbagai perusahaan asuransi harus memiliki strategi manajemen pemasaran yang baik, karena bagi pemilik kendaraan baru akan tersedia berbagai produk asuransi yang bisa dipilih sesuai dengan biaya dan jaminan proteksi yang diberikan. Persoalan yang menjadi tantangan bagi perusahaan asuransi adalah bagaimana strategi untuk memastikan pelanggan yang telah memiliki premi untuk mau memperpanjang masa asuransi dari kendaraan yang dimiliki.

Salah satu strategi bisnis untuk menjaga loyalitas pelanggan perusahaan asuransi kendaraan adalah menawarkan perpanjangan polis kepada pelanggan dengan berbagai insentif yang kompetitif namun tetap mampu memberikan profit bagi perusahaan. Dalam menawarkan perpanjangan asuransi tidak jarang pelanggan merasa sudah tidak memerlukan asuransi tersebut karena masa angsuran yang wajib diproteksi dengan jaminan asuransi telah berakhir. Kesulitan ini yang menjadi persoalan bagi perusahaan asuransi kendaraan bermotor.

Salah satu strategi yang bisa diterapkan perusahaan untuk bisa mempertahankan atau perpanjangan polis asuransi pelanggan adalah dengan melakukan *profiling* pelanggan selama mereka menjadi pemegang polis. *Profiling* ini akan menjadi dasar kuat dalam melakukan analisis terhadap kemungkinan pelanggan untuk memperpanjang polis yang dimiliki. Pendekatan data mining dapat dijadikan salah satu analisis, khususnya menggunakan metode prediksi dengan algoritma-algoritma klasifikasi seperti *Naive Bayes*, *Support Vector Machine* (SVM) maupun *Decision Tree*.

Berbagai studi terdahulu telah dilakukan dengan data yang dianalisis tentang asuransi kendaraan bermotor. Studi yang pernah dilakukan adalah memprediksi status pembayaran nasabah asuransi menggunakan algoritma *Naive Bayes*. Algoritma tersebut melakukan klasifikasi pada status pembayaran lancar atau tidak lancar. Hasil analisis digunakan memprediksi nasabah baru yang akan bergabung untuk diprediksi apakah pengajuannya dinyatakan ditolak atau diterima [1].

Studi lain juga mempelajari algoritma *Naive Bayes* dengan kombinasi metode *Backward Elimination* untuk mengetahui status nasabah yang tidak lancar dalam membayar premi. Algoritma *Naive Bayes* dipilih karena kemampuannya mengolah data dalam jumlah yang besar. Metode *Backward Elimination* dipilih sebagai pengembangan metode karena mampu menghasilkan klasifikasi terbaik [2].

Analisis juga dilakukan pada perusahaan asuransi, untuk mengetahui status nasabah ke dalam kelas tidak lancar, kurang lancar, dan lancar dalam kemampuan melakukan pembayaran premi. Studi ini menggunakan dua metode untuk melakukan klasifikasi yaitu *Naive Bayes* dan *Regresi Logistik Ordinal*. Hasil penelitian

menunjukkan bahwa dengan metode *Regresi Logistik Ordinal* diperoleh akurasi sebesar 50,90% sementara dengan metode *Naive Bayes* diperoleh nilai akurasi sebesar 55,41% [3].

Studi lain yang memprediksi keberlanjutan polis asuransi dilakukan untuk memprediksi keputusan melakukan pembaharuan (*renewal*) polis pelanggan asuransi kendaraan roda empat. Studi ini menggunakan algoritma *decision tree* atau *C4.5*. Proses evaluasi menggunakan *cross validation*  $k=10$  dan didapat hasil penelitian yang menunjukkan nilai akurasi sebesar 82,14%. Studi ini juga mengimplementasikan pembuatan Prototipe Sistem Informasi Prediksi berdasarkan aturan dari metode *C4.5*. Keakuratan aplikasi prototipe dalam melakukan prediksi keberlanjutan polis asuransi, mampu mencapai 86% dengan *error rate* sebesar 14% [4].

Studi lain menggunakan pendekatan *ensemble*, dengan penerapan metode *Weighted Vote (WV)* dalam data yang mengandung *imbalanced class*. Studi tersebut juga menggunakan pendekatan *Min-Max Normalization* untuk menangani permasalahan *smoothing noisy data*, dengan capaian nilai akurasi sebesar 85,21%, yang menunjukkan bahwa metode yang diusulkan ternyata memiliki kinerja yang lebih baik dari beberapa algoritma-algoritma pembandingan, seperti *Artificial Neural Network*, *C4.5*, *Support Vector Machine*, *k-Nearest Neighbor* serta *Naive Bayes* [5].

Berdasarkan studi terdahulu, maka pada penelitian saat ini diusulkan suatu metode pengambilan keputusan dalam prediksi perpanjangan polis asuransi kendaraan bermotor yang akan menggunakan tiga algoritma klasifikasi yaitu seperti *Naive Bayes*, *Decision Tree* dan *Support Vector Machine*, namun bukan sekedar membandingkan kinerja masing-masing pengklasifikasi (*classifier*) namun juga menentukan keputusan prediksi berdasarkan teknik pemilihan mayoritas (*majority voting*), dari ketiga algoritma tersebut. Hal ini menjadi penting untuk dilakukan agar pengambilan keputusan prediksi menjadi lebih kuat berdasarkan keputusan mayoritas.

## 2 LANDASAN TEORI

Data Mining adalah serangkaian proses untuk ekstraksi informasi menggunakan teknik matematika dan kecerdasan buatan, dikombinasikan dengan statistik, serta *machine learning* sehingga mampu mengidentifikasi informasi yang lebih bermanfaat dan menjadi pengetahuan baru yang lebih berharga. Data Mining merupakan terminologi yang sering kali digunakan untuk menjelaskan proses penemuan *knowledge* atau pengetahuan dari basis data (*knowledge discovery from database*) [6]. Istilah data mining

sebagai cabang disiplin ilmu berisi rangkaian kegiatan menggali dan menemukan pengetahuan dari tumpukan data atau informasi yang tersedia sebelumnya [7].

Pada bidang klasifikasi data mining, algoritme Naïve Bayes dikenal sebagai salah satu algoritme terbaik. Algoritme ini mampu melakukan proses klasifikasi dengan metode probabilitas dan statistik berdasarkan teorema Thomas Bayes, yang memprediksi peluang di masa mendatang berdasarkan data histori yang tersedia sebelumnya. Teorema Bayes dikombinasikan dengan metode Naïve, dan mengasumsikan kondisi antar atribut yang saling tidak terikat atau independen. Dengan kemampuan tersebut maka klasifikasi Naïve Bayes mampu mengidentifikasi ada atau tidak ada ciri tertentu dari sebuah kelas serta ada tidaknya hubungannya dengan ciri dari kelas lainnya [1].

Algoritma klasifikasi lain yang dikembangkan adalah algoritma decision tree atau dikenal dengan istilah ID3 (Iterative Dichotomiser). Ciri dari algoritma C4.5 adalah bentuk pohon keputusan dan menjadi metode pengembangan yang lebih baik dari algoritma ID3. Algoritma ini lebih mudah dimengerti, bersifat fleksibel dan interaktif karena dapat divisualisasikan dalam bentuk aturan terstruktur berupa pohon keputusan.

Struktur pohon dengan mode simpul pada Algoritma C4.5 mendeskripsikan setiap atribut, kemudian setiap cabang akan merepresentasikan hasil pengujian atribut dan setiap daun akan menggambarkan kelas atau label. Algoritma C4.5 bekerja dengan teknik rekursif dengan cara menganalisis setiap simpul keputusan, dan memilih pembagian yang optimal sampai tidak bisa dibagi lagi. Pada prinsipnya, konsep *entropy reduction* atau *information gain* menjadi pola kerja algoritma untuk memilih pembagian yang lebih optimal.

Algoritma klasifikasi lain seperti *Support Vector Machine* (SVM) berusaha mendapatkan *hyperplane* terbaik yang memiliki fungsi sebagai pemisah dua buah kelas. Pola dalam lingkung kelas  $-1$  misalnya, disimbolkan dengan warna tertentu, dan pattern pada kelas  $+1$ , disimbolkan dengan warna yang berbeda. Dengan demikian maka pemecahan masalah klasifikasi dapat diselesaikan dengan apabila ditemukan garis (*hyperplane*) yang menjadi pemisah di antara kedua kelompok tersebut.

Untuk menentukan *hyperplane* terbaik diperoleh dengan cara mengukur margin *hyperplane* tersebut dan mencari titik maksimalnya. Jarak antara *hyperplane* dengan *pattern* terdekat dari masing-masing kelas akan menjadi margin, dan selanjutnya *pattern* terdekat dinamakan dengan *support vector*. Pada pembelajaran berbasis Support Vector Machine, metode utamanya adalah dalam menentukan lokasi *hyperplane* pemisah [8].

Dari berbagai algoritma klasifikasi data mining, telah banyak studi terdahulu yang membandingkan kinerja algoritma klasifikasi. Salah satu studi bertujuan untuk mengetahui kinerja algoritma-algoritma klasifikasi seperti Decision Tree (DT) C4.5, Neural Network (NN), Bayesian Network (BN), K- Nearest Neighbors (KNN), dan SVM (Support Vector Machine) dan Naïve Bayes (NB).

Hasil studi tersebut telah menunjukkan fakta bahwa metode Bayesian Network memiliki kinerja terbaik dengan nilai akurasi sebesar 80,61%. Nilai presisi dan recall sebesar 0,78 dan 0,80, serta nilai AUC (*Area Under Curve*) sebesar 0,83 yang termasuk dalam kategori baik. Algoritme Decision Tree atau C4.5 memiliki kinerja paling tidak optimal karena hanya mampu mencapai nilai akurasi sebesar 76,6% [9].

*Multiple Classifiers System* (MCS) adalah salah satu kategori *Hybrid Intelligent System* (HIS). *Multiple Classifiers System* merupakan kombinasi beberapa classifier baik berbasis Homogeneous (sejenis) maupun Heterogeneous (berbeda jenis) untuk memberikan hasil keputusan. *Multiple Classifiers System* dapat juga diartikan kumpulan single classifier dengan prediksinya masing-masing yang digabungkan dengan suatu cara untuk menentukan klasifikasi objek yang baru. *Multiple Classifiers* dikenal pula dengan beberapa istilah diantaranya *Ensemble Method*, *Ensemble Classifier*, *Committee* atau *Classifier Fusion*.

*Hybrid Intelligent System* (HIS) adalah kombinasi bebas dari teknik-teknik komputasi untuk memecahkan masalah, meliputi seluruh fase komputasi dari normalisasi data hingga akhir pengambilan keputusan. Sistem ini memadukan pandangan yang berbeda menjadi suatu sistem yang efektif.

Pendekatan yang ditawarkan HIS adalah menggunakan kekuatan dari masing-masing classifier untuk meningkatkan kinerja dengan menggabungkan classifier-classifier tersebut. Berdasarkan pernyataan Wolpert tersebut maka tidak ada sebuah pendekatan dengan model *single classifier* yang optimal untuk seluruh masalah- masalah pengenalan pola. Oleh sebab itu untuk pemecahan masalah klasifikasi digunakan *Multiple Classifiers System* yang mengeksplorasi kekuatan dari masing-masing model single classifier untuk menghasilkan kualitas yang tinggi yang dapat melebihi kinerja single classifier.

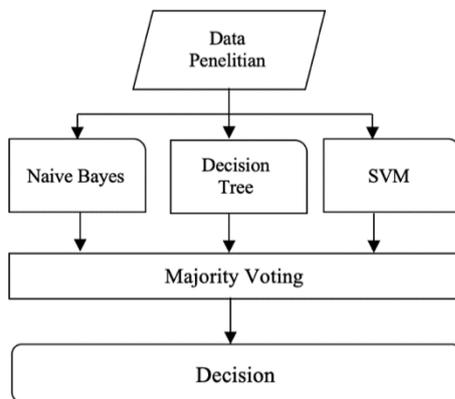
Objek Baru atau *Ensemble Design* mengarah ke bagaimana karakteristik sebuah *classifier* komplementer untuk mencapai tingkat akurasi dan keanekaragaman yang tinggi. Desain gabungan (*fuser*) bisa dikembangkan dengan mengacu pada Class Label Fusion (Mayoritas Voting). Dua pendekatan yang digunakan untuk membangun *Multiple Classifiers* adalah *Homogeneous Multiple Classifiers*, menggunakan pengklasifikasi yang sama untuk diproses dengan input data yang berbeda dan

*Heterogen Multiple Classifiers*, menggunakan berbagai pengklasifikasi untuk memproses input data yang sama. [10].

Berdasarkan teori-teori yang dijelaskan, maka studi saat ini tidak hanya bertujuan untuk membandingkan kinerja algoritma-algoritma klasifikasi, namun juga akan mengeksplorasi proses pemodelan yang akan menggunakan metode *Heterogen Multiple Classifiers* (HMC) dalam pengambilan keputusan akhir menggunakan teknik *majority voting* atau pemilihan mayoritas. Diharapkan pengembangan metode ini menghasilkan pemodelan dengan kinerja klasifikasi yang lebih baik.

### 3 METODOLOGI PENELITIAN

Rancangan penelitian Model prediksi Penentuan Pembaharuan Polis Asuransi Berbasis Heterogeneous Multiple Classifiers ini, dapat dilihat pada gambar berikut:



Gambar 1 Kerangka Penelitian

Terlihat pada Gambar 1 bahwa *Multiple Classifiers System* (MCS) terdiri dari tiga algoritma klasifikasi yaitu Naive Bayes, Decision Tree dan Support Vector Machine. *Multiple Classifiers System* ini akan memilih teknik pengambilan keputusan suara mayoritas dari tiga pengklasifikasi yang ada, dengan demikian keputusan yang dihasilkan akan memiliki dominasi.

Pada studi kali ini, data penelitian dibagi menjadi dua, sebesar 70% sebagai data latih dan sisanya sebanyak 30% sebagai data uji, dengan pembagian menggunakan metode *stratified sampling*.

Data penelitian merupakan dataset dari profil nasabah asuransi yang memiliki polis, baik yang melanjutkan perpanjangan premi maupun yang tidak. Adapun struktur dataset terlihat pada Tabel 1.

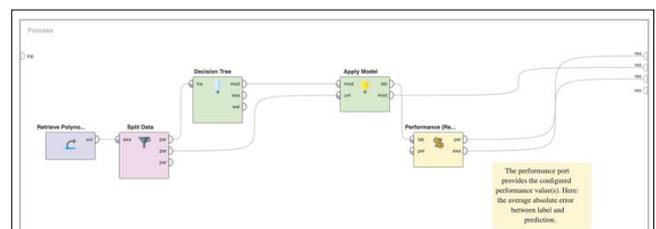
Tabel 1 Dataset Penelitian

No.	Atribut	Nilai
1.	Year_of_Manufacturing	<=2010
		>2010
2.	Policy_period (year)	1
		2
		3
		4
		5
3.	Customer_Type_Description	Personal
		Company
		Truck
		Sedan
		Multy Purpose Vehicle
4.	Object_Description	Pick up
		Minibus
		Partial Loss Accident
		Partial Loss Stolen
		No Claim
5.	Claim_Type	<=300 juta
		>300juta - 1M
		Besar
6.	Sum_Insured	Kecil
		No Claim
		Tinggi
7.	Loss_Ratio	Rendah
		Tinggi
		Rendah
8.	Claim_Ratio	1
		>1
		No Claim
9.	Claim_Frequency	Ya
		Tidak
10.	Renewal_decision	
11.		

### 4 HASIL DAN PEMBAHASAN

#### 4.1 Pengujian dan Pemodelan Menggunakan Metode C4.5

Proses pengujian dan seluruh tahap pemodelan pada penelitian ini diselesaikan dengan *tools* RapidMiner. Proses pembentukan model terlihat pada Gambar 2.



Gambar 2 Pemodelan dalam RapidMiner

Terlihat pada Gambar 2, pemodelan akan memisahkan data training dengan data testing untuk selanjutnya dianalisis menggunakan algoritma Decision Tree dan menghasilkan nilai kinerja model.

Pengujian pada model klasifikasi yang dihasilkan oleh metode C4.5, dilakukan dengan membandingkan hasil klasifikasi dari model klasifikasi yang dihasilkan dari pengujian terhadap data testing dengan data training.

Pengukuran nilai akurasi, precision, sensitivitas/recall, dan F-Measure didasarkan pada *Confusion Matrix* yang dihasilkan dari proses pengujian. Berikut adalah program pengujian dan pengukuran kinerja model klasifikasi metode C4.5 atau *Decision Tree*.

**Tabel 2 Hasil pemodelan dengan Decision Tree**

Akurasi Data Training	Akurasi Data Testing	Precision Data Testing	Recall Data Testing	F-Measure Data Testing
96,5%	89,8%	88,1%	88,8%	87,3%

Seperti yang dapat dilihat pada Tabel 2 dengan teknik klasifikasi menggunakan Algoritma C4.5 memiliki nilai akurasi data testing sebesar 89,8% yang artinya level kedekatan nilai prediksi telah mendekati nilai aktual. Nilai *precision* menjelaskan level ketepatan informasi dari jawaban yang diberikan oleh sistem, sebesar 88,1%. Nilai *recall* menjelaskan tingkat kesuksesan sistem dalam proses menemukan kembali sebuah informasi, mencapai sebesar 88,8% dan *F-Measure* yang merupakan bobot *harmonic mean* dari *recall* dan *precision* didapat nilai sebesar 87,3%.

#### 4.2 Pengujian dan Pemodelan Menggunakan Metode Naive Bayes.

Pengujian pada model klasifikasi yang dihasilkan oleh metode Naive Bayes, dilakukan dengan membandingkan hasil klasifikasi dari model klasifikasi yang dihasilkan dari pengujian terhadap data testing dengan data training.

Pengukuran nilai akurasi, *precision*, sensitivitas/*recall*, dan *F-Measure* didasarkan pada *Confusion Matrix* yang dihasilkan dari proses pengujian. Berikut adalah program pengujian dan pengukuran kinerja model klasifikasi metode Naive Bayes.

**Tabel 3 Hasil pemodelan dengan Naive Bayes**

Akurasi Data Training	Akurasi Data Testing	Precision Data Testing	Recall Data Testing	F-Measure Data Testing
97,2%	91,4%	91,5%	90,7%	88,8%

Seperti yang dapat dilihat pada Tabel 3 dengan teknik klasifikasi menggunakan Algoritma Naive Bayes memiliki nilai akurasi data testing sebesar 91,4% yang artinya level kedekatan nilai prediksi telah mendekati nilai aktual. Nilai *precision* menjelaskan level ketepatan informasi dari jawaban yang diberikan oleh sistem, sebesar 91,5%. Nilai *recall* menjelaskan tingkat kesuksesan model dalam menemukan kembali sebuah informasi sebesar 90,7% dan *F-Measure* yang merupakan bobot *harmonic mean* dari *recall* dan *precision* didapat nilai sebesar 88,8%.

#### 4.3 Pengujian dan Pemodelan Menggunakan Metode Support Vector Machine

Pengujian pada model klasifikasi yang dihasilkan oleh metode SVM, dilakukan dengan membandingkan hasil klasifikasi dari model klasifikasi yang dihasilkan dari pengujian terhadap data testing dengan data training.

Pengukuran nilai akurasi, *precision*, sensitivitas/*recall*, dan *F-Measure* didasarkan pada *Confusion Matrix* yang dihasilkan dari proses pengujian. Berikut adalah program pengujian dan pengukuran kinerja model klasifikasi metode Naive Bayes.

**Tabel 3 Hasil pemodelan dengan SVM**

Akurasi Data Training	Akurasi Data Testing	Precision Data Testing	Recall Data Testing	F-Measure Data Testing
96,6%	92,5%	92,1%	93,8%	89,1%

Seperti yang dapat dilihat pada Tabel 3 dengan teknik klasifikasi menggunakan Algoritma Support Vector Machine memiliki nilai akurasi data testing sebesar 92,5% yang artinya level kedekatan nilai prediksi telah mendekati nilai aktual. Nilai *precision* menjelaskan level ketepatan informasi dari jawaban yang diberikan oleh sistem, sebesar 92,1%. Nilai *recall* menjelaskan tingkat kesuksesan model dalam menemukan kembali sebuah informasi sebesar 93,8% dan *F-Measure* yang merupakan bobot *harmonic mean* dari *recall* dan *precision* didapat nilai sebesar 89,1%.

#### 4.4 Pengujian dan Pemodelan Menggunakan Metode Heterogeneous Multiple Classifiers (HMC)

Pengujian pada model klasifikasi yang dihasilkan oleh metode HMC, dilakukan dengan membandingkan hasil klasifikasi dari model klasifikasi yang dihasilkan dari pengujian terhadap data testing dengan data training.

Pengukuran nilai akurasi, *precision*, sensitivitas/*recall*, dan *F-Measure* didasarkan pada *Confusion Matrix* yang dihasilkan dari proses pengujian. Berikut adalah program pengujian dan pengukuran kinerja model klasifikasi metode HMC.

**Tabel 4 Hasil pemodelan dengan HMC**

Akurasi Data Training	Akurasi Data Testing	Precision Data Testing	Recall Data Testing	F-Measure Data Testing
98,1%	95,8%	93,9%	97,2%	93,9%

Seperti yang dapat dilihat pada Tabel 4 dengan teknik klasifikasi menggunakan teknik HMC memiliki nilai akurasi data testing sebesar 95,8% yang artinya level kedekatan nilai prediksi telah mendekati nilai aktual. Nilai *precision* menjelaskan level ketepatan informasi dari jawaban yang diberikan oleh sistem, sebesar 93,9%, nilai *recall* menjelaskan tingkat kesuksesan model dalam menemukan kembali sebuah informasi sebesar 97,2% dan *F-Measure* yang

merupakan bobot harmonic mean dari recall dan precision didapat nilai sebesar 93,9%.

#### 4.5 Perbandingan Kinerja Model Klasifikasi C4.5, Naive Bayes, Support Vector Machine dan HMC

Berdasarkan hasil pengujian terhadap dataset Laporan Aktuaria yang mana berisi info para pemegang polis yang melakukan pembaharuan polis/renewal pada polis mereka. Dengan menggunakan jumlah data latih maupun data uji, maka diperoleh hasil prediksi terbaik dari komparasi perhitungan untuk metode algoritma C4.5, *Naive Bayes*, *Support Vector Machine* dan *Heterogeneous Multiple Classifiers* yang digunakan dan dapat dilihat pada tabel 5. Perbandingan kinerja model klasifikasi dari setiap metode dapat dilihat pada table 5.

**Tabel 5** Perbandingan Kinerja Model Klasifikasi Data Testing

Algoritma	Akurasi	Precision	Recall	F-Measure
C4.5	89,8%	88,1%	88,8%	87,3%
NB	91,4%	91,5%	90,7%	88,8%
SVM	92,5%	92,1%	93,8%	89,1%
HMC	93,57%	97,20%	95,20%	95,30%

Berdasarkan hasil pengukuran kinerja masing-masing model klasifikasi dengan metode C4.5, *Naive Bayes*, SVM dan *Heterogeneous Multiple Classifiers* dapat dilihat bahwa nilai akurasi dari data training dan data testing yang dihasilkan oleh model klasifikasi berbasis *Heterogeneous Multiple Classifiers* lebih tinggi dari semua model berbasis *Single Classifier*. Hal ini terjadi karena mayoritas pemilihan telah menghasilkan banyak kesamaan untuk keputusan benar dibandingkan hasil dari satu pengklasifikasi. Hal tersebut dapat dilihat pada nilai *Precision* dan *Recall* pada dua pengklasifikasi yang dominan yaitu *Naive Bayes* dan *Support Vector Machine*.

Hasil penelitian juga turut membuktikan bahwa model klasifikasi berbasis *Heterogeneous Multiple Classifiers* memiliki kinerja yang lebih baik daripada model pengklasifikasi berbasis *Single Classifier*, baik itu C4.5, *Naive Bayes* maupun *Support Vector Machine*.

## 5 SIMPULAN

Dari seluruh tahap penelitian yang telah diselesaikan maka studi ini mendapatkan beberapa kesimpulan, antara lain:

- Model *Heterogeneous Multiple Classifiers* memiliki kinerja yang lebih baik saat dilakukan pemodelan yang terdiri dari tiga jenis pengklasifikasi tunggal yang terdiri dari C4.5, *Naive Bayes* maupun *Support Vector Machine*.
- Nilai akurasi, *precision* dan *recall* dari setiap pengklasifikasi tunggal terbukti tidak lebih tinggi

dari pada kinerja dari pemodelan yang menggunakan *Heterogeneous Multiple Classifier*, namun demikian secara satuan pengklasifikasi, metode SVM terbukti lebih unggul dalam hal kinerja dibanding dua pengklasifikasi lain yaitu C4.5 maupun *Naive Bayes*.

## KEPUSTAKAAN

- [1] Bustami, "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi," *Jurnal Informatika*, vol. 8, no. 1, hal. 884–898, 2014.
- [2] Betrisandi, "Klasifikasi Nasabah Asuransi Jiwa Menggunakan Algoritma Naive Bayes Berbasis Backward Elimination," *ILKOM Jurnal Ilmiah*, vol. 9, no. 1, hal. 96–101, 2017.
- [3] R. Sutitis, S. Suparti, dan D. Ispriyanti, "Klasifikasi Tingkat Kelancaran Nasabah Dalam Membayar Premi Dengan Menggunakan Metode Regresi Logistik Ordinal Dan Naive Bayes," *Jurnal Gaussian*, vol. 4, no. 3, hal. 651–659, 2015.
- [4] W. F. Pattipeilohy, A. Wibowo, dan D. R. Utari, "Fakultas Teknik – Universitas Muria Kudus," in *Prosiding SNATIF ke-4 Tahun 2017*, 2017, hal. 96–101.
- [5] A. Alhamad, A. I. S. Azis, B. Santoso, dan S. Taliki, "Prediksi Penyakit Jantung Menggunakan Metode-Metode Machine Learning Berbasis Ensemble – Weighted Vote," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 5, no. 3, hal. 352, 2019.
- [6] Kusriani, *Algoritma Data Mining*. Jakarta, 2009.
- [7] R. T. Wulandari, *Data Mining Teori dan Aplikasi Rapidminer*. 2017.
- [8] A. S. Nugroho, "Pengantar Support Vector Machine," 2007, hal. 1–6.
- [9] M. Windarti dan A. Suradi, "Perbandingan Kinerja 6 Algoritme Klasifikasi Data Mining untuk Prediksi Masa Studi Mahasiswa," *Telematika*, vol. 12, no. 1, hal. 14, 2019.
- [10] I. G. A. Suwartane, M. Syafrullah, dan K. Adiyarta, "Prediction models of diabetes diseases based on Heterogeneous Multiple Classifiers," *International Journal of Advances in Soft Computing and its Applications*, vol. 10, no. 2, hal. 33–49, 2018.