Penerapan Algoritma C4.5 Untuk Klasifikasi Keterlambatan Pembayaran Premi Asuransi

Jefry Antonius Karlia[1], Wawan Nurmansyah*[2]
[1,2]Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Katolik Musi Charitas, Jl. Bangau No.60 Palembang 30113, Telp. (0711) 321801
e-mail: [1] jefryantonius87@gmail.com, *[2] w_nurmansyah@ukmc.ac.id

(artikel diterima: 28-11-2020, artikel disetujui: 18-04-2021)

Abstrak

Permasalahan yang sering timbul pada perusahaan asuransi adalah banyaknya nasabah yang tidak lancar dalam membayar premi. Prosedur yang berlaku di asuransi pada masa tenggang pembayaran yaitu 30 hari. Nasabah tertanggung wajib mengikuti prosedure pembayaran premi, apabila nasabah tidak membayar premi maka polis asuransi akan dibatalkan, hal tersebut bagian dari kerugian perusahaan. Sebuah perusahaan asuransi mempunyai data yang banyak dan data tersebut dapat diolah untuk menghasilkan sebuah informasi bagaimana mengetahui potensi adanya keterlambatan nasabah dari pola yang dibentuk menggunakan metode C4.5. Penelitian ini dilakukan dengan menerapkan algoritma C4.5 menggunakan data nasabah asuransi. Hasil dari penelitian ini berupa sistem klasifikasi keterlambatan pembayaran premi asuransi yang dapat mengklasifikasi status pembayaran premi nasabah asuransi sebagai pertimbangan untuk menerima nasabah asuransi. Hasil pengujian sistem menunjukkan bahwa sistem dapat mengklasifikasikan status pembayaran premi nasabah asuransi dengan tingkat akurasi klasifikasi sebesar 88%.

***Kata kunci*:** *Algoritma C 4.5, Asuransi, Klasifikasi, Premi*

Abstract

The problem that often arises in insurance companies is the number of customers who do not smoothly pay premiums. The procedure that applies to the insurance during the grace period is 30 days. The insured customer must follow the premium payment procedure, if the customer does not pay the premium, the insurance policy will be canceled, this is part of the company's loss. An insurance company has a lot of data and this data can be processed to produce information on how to find out potential customer delays from a pattern formed using the C4.5 method. This research was conducted by applying the C4.5 algorithm using insurance customer data. The results of this study are a classification system for late payment of insurance premiums that can classify insurance customer premium payment status as a consideration for accepting insurance customers. The system test results show that the system can classify the status of insurance customer premium payments with a classification accuracy of 88%.

***Keywords*:** *Algorithm C 4.5, Insurance, Classification, Premium*

DOI: https://doi.org/10.31001/tekinfo.v9i1.1090
E-ISSN: 2303-1867 | P-ISSN: 2303-1476

1. PENDAHULUAN

Dengan kemajuan teknologi informasi, kebutuhan akan informasi yang akurat sangat dibutuhkan dalam kehidupan sehari-hari, sehingga informasi akan menjadi suatu elemen penting dalam perkembangan masyarakat saat ini dan juga waktu yang akan datang (Novilla, Goejantoro and Amijaya, 2019). Asuransi atau pertanggungan adalah perjanjian antara dua pihak atau lebih dimana pihak penanggung terikat kepada tertanggung dengan menerima premi asuransi, untuk memberikan penggantian kepada tertanggung karena kerugian, kerusakan atau kehilangan keuntungan yang diharapkan, atau tanggung jawab hukum kepada pihak ketiga yang mungkin akan diderita oleh tertanggung, yang timbul dari suatu peristiwa yang tidak pasti. Premi merupakan pendapatan bagi perusahaan asuransi yang jumlahnya ditentukan dalam suatu persentase atau tarif tertentu dari jumlah yang dipertanggungkan. Premi merupakan tarif yang harus dibayar oleh tertanggung agar pihak asuransi dapat menjamin/menanggung kerugian yang dialami oleh tertanggung. Pendapatan premi untuk perusahaan asuransi ditentukan oleh jumlah premi yang dibayar oleh nasabah/tertanggung. Permasalahan yang sering dialami oleh asuransi adalah banyaknya nasabah yang tidak lancar dalam membayar premi (Betrisandi, 2017). Pada prosedur yang berlaku di PT Asuransi Etiqa Internasional Indonesia, terdapat masa tenggang pembayaran yaitu 30 hari dimana dalam masa tersebut nasabah/tertanggung harus membayarkan sejumlah premi yang sudah ditentukan dan apabila nasabah/tertanggung tidak membayar premi, maka polis asuransi akan dibatalkan sehingga keuntungan asuransi akan berkurang dan akan merugikan pihak asuransi.

Sebuah perusahaan asuransi mempunyai data yang besar. Masih banyak yang belum menyadari bahwa dari pengolahan data-data tersebut dapat memberikan informasi yang bermanfaat berupa klasifikasi data nasabah (Bustami, 2013). Dengan proses Data *Mining* dari data nasabah, dapat ditemukan pola-pola ataupun hubungan tertentu antara data untuk menjadi informasi yang bermanfaat (Avegad and Wibowo, 2019). Berdasarkan permasalahan tersebut, maka dilakukanlah penelitian lebih lanjut dengan membuat implementasi aplikasi Data *Mining* agar pihak asuransi dapat mengetahui nasabah yang akan terlambat membayar premi sehingga pihak asuransi dapat mencegah keterlambatan tersebut sedini mungkin.

Hasil penelitian sebelumnya, Hasil penelitian menunjukkan bahwa nilai K dan nilai E yang optimal untuk klasifikasi status pembayaran premi di PT. Bumiputera Kota Samarinda menggunakan NWKNN sebesar K=3 dan E=6 dengan nilai akurasi sebesar 75% (Grassella, Purnamasari and Amijaya, 2019). Hasil penelitian menunjukkan bahwa proses bisnis yang ada sekarang ini dibagi berdasarkan *intermediary* (*agent*, *dealer*, dan *customer* langsung). Proses bisnis terbaik saat ini adalah proses bisnis dengan *intermediary agent* karena menghasilkan piutang paling rendah (Kusumawati, Wibisono and Aritonang, 2014). Hasil kesalahan akurasi dengan menggunakan nilai APER (*Apparent Rate Error*) menunjukkan bahwa metode *Naive Bayes* memiliki tingkat akurasi yang lebih tinggi sebesar 5,38% daripada analisis diskriminan *Fisher* sebesar 46,15% dalam menganalisis status pembayaran premi nasabah asuransi (Ainurrochmah, Hayati and Satriya, 2019). Hasil perbandingan perhitungan akurasi dari kedua analisis menunjukkan bahwa jaringan saraf tiruan memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan metode *naïve bayes*. Hasil akurasi klasifikasi Naïve Bayes 82,76% dan jaringan syaraf tiruan 86,21%

DOI: https://doi.org/10.31001/tekinfo.v9i1.1090
E-ISSN: 2303-1867 | P-ISSN: 2303-1476

(Ardyanti, Goejantoro and Amijaya, 2020). Hasilnya, setelah dilakukan pengujian dengan menggunakan parameter biodata nasabah dengan jumlah nasabah sebanyak 1312 ternyata menghasilkan akurasi sebesar secara keseluruhan nilai hasil validasi adalah *accuracy* = 91,06%, *precision* = 100,00% dan *recall* = 78,00%, artinya akurasi pengujian dengan menggunakan algoritma C4.5 baik (Sucipto, 2015).

Hasil pengujian dengan algoritma klasifikasi *Random Forest* mampu menganalisis kredit yang bermasalah dan yang debitur yang tidak bermasalah dengan nilai akurasi sebesar 87,88%. Di samping itu, model pohon keputusan ternyata mampu meningkatkan akurasi dalam menganalisis kelayakan kredit yang diajukan calon debitur (Hanun and Zailani, 2020). Dari penelitian yang dilakukan, akurasi data yang diperoleh dalam penelitian ini adalah sebesar 92,5% dengan error sebesar 7,5% dari 160 data yang digunakan untuk *training* dan 40 data untuk *testing* (Kurniawan and Kriestanto, 2016). Dari penelitian yang dilakukan, diketahui nilai *precision* terbesar dicapai oleh algoritma C4.5 dengan partisi data 90%:10% dengan nilai sebesar 78,08 %. Nilai *recall* terbesar partisi data 80%:20% dengan nilai sebesar 96,4 %. Dari hasil data latih yang sama, ID3 menghasilkan *precision* sebesar 71,51% dan *recall* sebesar 92,09%. Hasil akhir dari penelitian ini membuktikan bahwa pada kasus ini algoritma C4.5 memiliki tingkat akurasi yang tinggi dan lebih baik dari ID3 (Amin, Indwiarti and Sibaroni, 2015). Hasil klasifikasi menggunakan Algoritma C4.5 menunjukan bahwa diperoleh akurasi 97,5%, berdasarkan hasil yang diperoleh menunjukan bahwa algoritma C4.5 cocok digunakan untuk menentukan kelayakan pemberian kredit nasabah pada KOPERIA (Santoso and Sekardiana, 2019). Hasil analisis menunjukkan *Area Under Curve* yang optimis sebesar 0.971 ini menunjukkan hasil klasifikasi berada pada kategori sangat baik (Setiawan, 2020).

Beberapa referensi penelitian yang berbasis pengolahan data dengan menggunakan metode C4.5 dan metode lainnya untuk mendapatkan pola data. Pola data dapat berupa klasifikasi data dan penggunaan hasil pola ini dapat menjadi rule base sebagai penentu bila mana data baru masuk maka dengan rule base tersebut dapat mengklasifikasi data baru tersebut masuk kedalam bagian klasifikasi sehingga hasil dari klasifikasi dapat menjadi pendukung dalam membuat suatu keputusan atau kebijakan yang diambil oleh pimpinan perusahaan.

2. METODE PENELITIAN

2.1 Metode Pengembangan Sistem

Pada penelitian "Penerapan Algoritma C4.5 untuk Klasifikasi Keterlambatan Pembayaran Premi Asuransi" ini, peneliti menggunakan model FDD (*Feature Driven Development*). *Feature Driven Development* (FDD) merupakan proses yang didesain dan dilaksanakan untuk menyajikan hasil kerja secara berulang-ulang dalam waktu tertentu dan dapat diukur (Hariono *et al.*, 2014). FDD adalah pendekatan langsung untuk menghasilkan sistem yang menggunakan metode sederhana yang mudah dipahami dan mudah diimplementasikan, teknik pemecahan masalah, dan pedoman pelaporan yang menyediakan informasi yang dibutuhkan setiap *stakeholders* untuk membuat keputusan yang tepat waktu (Palmer and Felsing, 2002).

DOI: https://doi.org/10.31001/tekinfo.v9i1.1090
E-ISSN: 2303-1867 | P-ISSN: 2303-1476

2.2 Landasan Teori

2.2.1 Knowledge Discovery In Database (KDD)

Knowledge Discovery In Database (KDD) merupakan metode untuk memperoleh pengetahuan dari *database* yang ada. Dalam *database* terdapat tabel - tabel yang saling berhubungan / berelasi. Hasil pengetahuan yang diperoleh dalam proses tersebut dapat digunakan sebagai basis pengetahuan (*knowledge base*) untuk keperluan dalam mengambil sebuah keputusan. Istilah *Knowledge Discovery in Database* (KDD) dan data *mining* seringkali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain, dan salah satu tahapan dalam keseluruhan proses KDD adalah data *mining*. Proses KDD secara garis besar dapat dijelaskan sebagai berikut (Mardi, 2017):

a. *Data Selection* : Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *Knowledge Discovery in Database (KDD)* dimulai. Data hasil seleksi yang akan digunakan untuk proses data *mining* disimpan dalam suatu berkas terpisah dari basis data operasional.
b. *Pre-processing / Cleaning :* Sebelum proses data *mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus dari *Knowledge Discovery in Database (KDD).* Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak. Juga dilakukan proses *enrichment*, yaitu proses "memperkaya" data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk *Knowledge Discovery in Database (KDD)*, seperti data atau informasi eksternal lainnya yang diperlukan.
c. *Transformation Coding* adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses *coding* dalam *Knowledge Discovery in Database (KDD)* merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.
d. *Data Mining* adalah proses mencari pola atau informasi yang menarik dalam data dengan menggunakan teknik atau metode tertentu. Teknik-teknik, metode-metode, atau algoritma dalam data *mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database (KDD)* secara keseluruhan.
e. *Interpretation / Evaluation :* Pola informasi yang dihasilkan dari proses data *mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses *Knowledge Discovery in Database (KDD)* yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.2.2 Data Mining

Data *mining* merupakan bagian dari tahapan proses *Knowledge Discovery in Database (KDD)* (Mardi, 2017). Data *mining* merupakan teknologi yang menggabungkan metode analisis tradisional dengan algoritma yang canggih untuk

DOI: https://doi.org/10.31001/tekinfo.v9i1.1090
E-ISSN: 2303-1867 | P-ISSN: 2303-1476

memproses data dengan jumlah besar. Data *mining* adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam *database*. Data *mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan yang bermanfaat yang tersimpan di dalam *database* besar (Novilla, Goejantoro and Amijaya, 2019). Data *mining* bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan data mining adalah kenyataan bahwa data *mining* mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu. Berawal dari beberapa disiplin ilmu, data mining bertujuan untuk memperbaiki teknik tradisional sehingga bisa menangani (Kamagi and Hansun, 2014):

a. Jumlah data yang sangat besar
b. Dimensi data yang tinggi
c. Data yang heterogen dan berbeda sifat

Data *mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Mardi, 2017):

a. *Description* (Deskripsi)
b. *Estimation* (Estimasi)
c. *Prediction* (Prediksi)
d. *Classification* (Klasifikasi)
e. *Clustering* (Pengklusteran)
f. *Association* (Asosiasi)

2.2.3 Klasifikasi

Metode-metode / model-model yang telah dikembangkan oleh periset untuk menyelesaikan kasus klasifikasi antara lain adalah (Mardi, 2017):

a. Pohon keputusan (*Decision Tree*)
b. Pengklasifikasi Bayes / Naive Bayes
c. Jaringan saraf tiruan
d. Analisis statistik
e. Algoritma genetik
f. Rough sets
g. Pengklasifikasi k-nearest neighbour
h. Metode berbasis aturan
i. *Memory* based reasoning
j. *Support vector machine*

2.2.4 Decision Tree

Decision tree merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (*tree*) di mana setiap *node* merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas. *Node* yang paling atas dari *decision tree* disebut sebagai *root*. Pada *decision tree* terdapat 3 jenis *node*, yaitu (Andriani, 2012) :

a. *Root Node*, merupakan *node* paling atas, pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.
b. *Internal Node*, merupakan *node* percabangan, pada *node* ini hanya terdapat satu *input* dan mempunyai *output* minimal dua.

DOI: https://doi.org/10.31001/tekinfo.v9i1.1090
E-ISSN: 2303-1867 | P-ISSN: 2303-1476

c. *Leaf node atau terminal node*, merupakan *node* akhir, pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output*.

Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, CART, dan C4.5. Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan *record.* Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon. *Decision tree* tergantung pada aturan *if-then*, tetapi tidak membutuhkan parameter dan metrik. Strukturnya yang sederhana dan dapat ditafsirkan memungkinkan *decision tree* untuk memecahkan masalah atribut *multi-type. Decision tree* juga dapat mengelola nilai-nilai yang hilang atau data *noise* (Andriani, 2012).

2.2.5 Algoritma C4.5

Di akhir tahun 1970 hingga di awal tahun 1980-an, J. Ross Quinlan seorang peneliti di bidang mesin pembelajaran mengembangkan sebuah model pohon keputusan yang dinamakan ID3 (*Iterative Dichotomiser*), walaupun sebenarnya proyek ini telah dibuat sebelumnya oleh E.B. Hunt, J. Marin, dan P.T. Stone. Kemudian Quinlan membuat algoritma dari pengembangan ID3 yang dinamakan C4.5 yang berbasis *supervised learning*. Serangkaian perbaikan yang dilakukan pada ID3 mencapai puncaknya dengan menghasilkan sebuah sistem praktis dan berpengaruh untuk *decision tree* yaitu C4.5. Perbaikan ini meliputi metode untuk menangani *numeric attributes, missing values, noisy* data, dan aturan yang menghasilkan *rules* dari *trees.* Ada beberapa tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5, yaitu (Andriani, 2012):

a. Mempersiapkan data *training*. Data *training* biasanya diambil dari data histori yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.
b. Menghitung akar dari pohon. Akar akan diambil dari atribut yang akan terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan rumus :

$$Entropy\ (S) = \sum_{i=1}^{n} - pi\ log_2\ pi \quad \text{.....(1)}$$

Keterangan :
S = Himpunan kasus
n = Jumlah partisi S
Pi = Proporsi Si terhadap S

Kemudian hitung nilai gain menggunakan rumus :

$$Gain\ (S, A) = Entropy\ (S) - \sum_{i=1}^{n} \frac{|Si|}{S} * Entropy\ (Si) \quad \text{.....(2)}$$

DOI: https://doi.org/10.31001/tekinfo.v9i1.1090
E-ISSN: 2303-1867 | P-ISSN: 2303-1476

Keterangan :
S = Himpunan Kasus
A = Fitur
n = Jumlah partisi atribut A
|Si| = Proporsi Si terhadap S
|S| = Jumlah kasus dalam S

c. Ulangi langkah ke 2 dan langkah ke 3 hingga semua *record* terpartisi.
d. Proses partisi pohon keputusan akan berhenti saat :
1) Semua *record* dalam simpul N mendapat kelas yang sama.
2) Tidak ada atribut di dalam *record* yang dipartisi lagi.
3) Tidak ada *record* di dalam cabang yang kosong.

2.2.6 Pengujian Akurasi Klasifikasi

Sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua data dengan benar, tetapi tidak dipungkiri bahwa kinerja suatu sistem tidak bisa 100% benar sehingga sebuah sistem klasifikasi harus diukur tingkat akurasi klasifikasinya.Untuk menghitung akurasi klasifikasi digunakan rumus (Putranto, Wuryandari and Sudarno, 2015):

$$Akurasi = \frac{Jumlah\ data\ yang\ diklasifikasikan\ secara\ benar}{Jumlah\ klasifikasi\ yang\ dilakukan}\ x\ 100\% \quad .....(3)$$

3. HASIL DAN PEMBAHASAN

3.1 Perhitungan Algoritma C4.5

Berikut hasil perhitungan algoritma C4.5 pada 3.949 data yang digunakan pada proses *training* data. Terdapat 3 kategori pada atribut Premi yaitu 1>5.000.000, 5.000.001-10.000.000, >10.000.000. Terdapat 4 kategori pada atribut Jenis Asuransi yaitu *Engineering*, Perkapalan, *Motor Vehicle* dan *Property*. Terdapat 2 kategori pada atribut Kepemilikan yaitu Pribadi dan Kelompok. Terdapat 12 kategori pada atribut Bulan yaitu 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 dan 12. Terdapat 4 kategori pada atribut Sumber yaitu *Agent*, *Broker*, *Direct*, *Leasing*. Lalu, akan dihitung jumlah data, jumlah lancar, jumlah terlambat, *entropy* (1) dan *gain* (2) dari setiap kategori yang ada. Hasil perhitungan dapat dilihat pada Tabel 1.

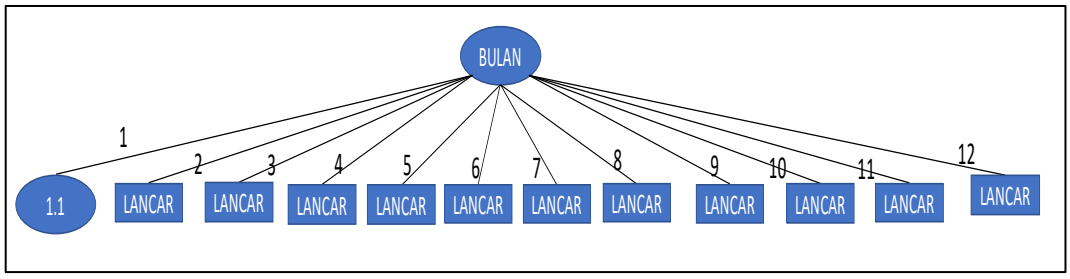
Tabel 1 Perhitungan C4.5 *Node* 1

Atribut	Kategori	Jumlah	Lancar	Terlambat	Entropy	Gain
Total		3.949	3.506	443	0,506457572	
Premi						0,007674359
	1-5.000.000	2.873	2.548	325	0,509262133	
	5.000.001-10.000.000	820	731	89	0,495486017	
	>1.000.000	161	136	25	0,622896063	

DOI: https://doi.org/10.31001/tekinfo.v9i1.1090
E-ISSN: 2303-1867 | P-ISSN: 2303-1476

Atribut	Kategori	Jumlah	Lancar	Terlambat	Entropy	Gain
Jenis Asuransi						0,001751514
	Engineering	85	67	18	0,744842397	
	Perkapalan	71	59	12	0,655444445	
	Motor vehicle	2.381	2.126	255	0,491099679	
	Property	1.412	1.254	158	0,505614539	
Kepemilikan						1,82094E-06
	Pribadi	2.549	2.264	285	0,505347112	
	Kelompok	1.400	1.242	158	0,508474265	
Bulan						0,1174183
	1	263	148	115	0,988613081	
	2	382	347	35	0,441864225	
	3	279	269	10	0,222893152	
	4	299	277	22	0,379138545	
	5	253	232	21	0,412676354	
	6	319	317	2	0,054893891	
	7	473	468	5	0,084554028	
	8	381	377	4	0,084081285	
	9	289	276	13	0,264688775	
	10	306	292	14	0,268068431	
	11	335	277	58	0,664824815	
	12	370	226	144	0,96427431	
Sumber						0,00150369
	Agent	3.462	3.082	380	0,499201111	
	Broker	2.97	267	30	0,472189385	
	Direct	115	98	17	0,604373242	
	Leasing	75	59	16	0,747806158	

Atribut yang mempunyai nilai gain terbesar adalah atribut bulan dengan gain 0,1174183 sehingga atribut bulan yang dimasukkan ke dalam *node* pada *decision tree* yang divisualisasikan pada gambar 2.



Gambar 2 *Decision Tree Node* 1

Karena data pada atribut bulan dengan kategori 1 masih kurang tepat untuk mengklasifikasi data dimana tidak terdapat data yang dominan pada salah satu kategori pada atribut Terlambat yang merupakan atribut target, maka dibuat *node* 1.1 untuk dianalisis lebih lanjut. Lalu, akan dihitung jumlah data, jumlah lancar, jumlah

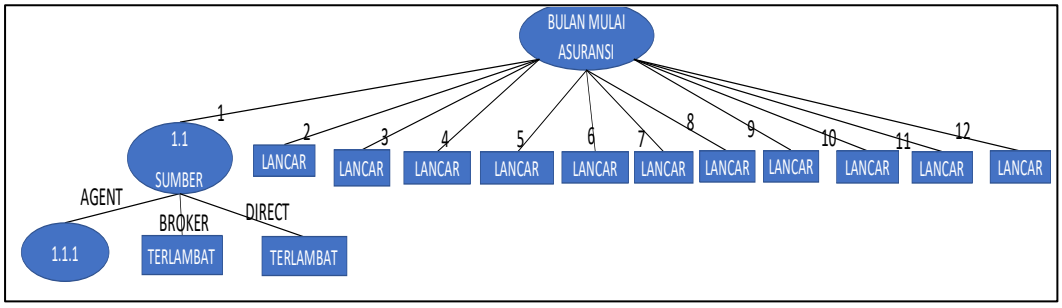
DOI: https://doi.org/10.31001/tekinfo.v9i1.1090
E-ISSN: 2303-1867 | P-ISSN: 2303-1476

terlambat, *entropy* dan *gain* dari setiap kategori yang ada dimana data tersebut mempunyai atribut bulan dengan kategori 1. Karena atribut bulan sudah digunakan pada *decision tree*, maka atribut bulan tidak digunakan lagi. Hasil perhitungan dapat dilihat pada Tabel 2.

Tabel 2 Perhitungan C4.5 *Node* 1.1

Atribut	Kategori	Jumlah	Lancar	Terlambat	Entropy	Gain
Total		263	148	115	0,988613081	
Premi						0,017025842
	1-5.000.000	189	107	82	0,987341726	
	5.000.001-10.000.000	55	31	24	0,988283611	
	>1.000.000	15	6	9	0,970950594	
Jenis asuransi						0,01764637
	Engineering	2	0	2	0	
	Perkapalan	7	5	2	0,863120569	
	Motor vehicle	177	94	83	0,997212189	
	Property	77	49	28	0,945660305	
Kepemilikan						0,001777984
	Pribadi	176	96	80	0,994030211	
	Kelompok	87	52	35	0,972279462	
Sumber						0,029708154
	Agent	254	147	107	0,982035859	
	Broker	6	0	6	0	
	Direct	3	1	2	0,918295834	
	Leasing	0	0	0	0	

Atribut yang mempunyai nilai gain terbesar adalah atribut sumber dengan gain 0,029708154 sehingga atribut jenis asuransi yang dimasukkan ke dalam *node* pada *decision tree* yang divisualisasikan pada gambar 3.



Gambar 3 *Decision Tree Node* 1.1

Karena data pada atribut sumber dengan kategori *agent* masih kurang tepat untuk mengklasifikasi data dimana tidak terdapat data yang dominan pada salah satu kategori pada atribut Terlambat yang merupakan atribut target, maka dibuat *node* 1.1.1 untuk dianalisis lebih lanjut. Lalu, akan dihitung jumlah data, jumlah lancar, jumlah terlambat, *entropy* dan *gain* dari setiap kategori yang ada dimana data tersebut

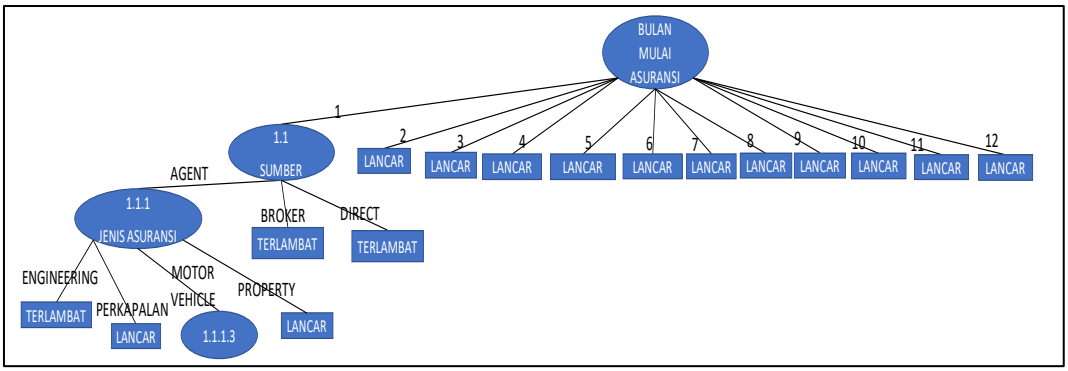
DOI: https://doi.org/10.31001/tekinfo.v9i1.1090
E-ISSN: 2303-1867 | P-ISSN: 2303-1476

mempunyai atribut bulan dengan kategori 1 dan atribut sumber dengan kategori *agent*. Karena atribut bulan dan sumber sudah digunakan pada *decision tree*, maka atribut bulan dan sumber tidak digunakan lagi. Hasilnya dapat dilihat pada Tabel 3.

Tabel 3 Perhitungan C4.5 *Node* 1.1.1

Atribut	Kategori	Jumlah	Lancar	Terlambat	Entropy	Gain
Total		254	147	107	0,982035859	
Premi						0,016257766
	1-5.000.000	183	107	76	0,979200065	
	5.000.001-10.000.000	55	31	24	0,988283611	
	>1.000.000	12	5	7	0,979868757	
Jenis asuransi						0,018377232
	Engineering	1	0	1	0	
	Perkapalan	6	5	1	0,650022422	
	Motor vehicle	174	94	80	0,995325107	
	Property	73	48	25	0,92715874	
Kepemilikan						0,002649509
	Pribadi	172	96	76	0,99022469	
	Kelompok	82	51	31	0,956652272	

Atribut yang mempunyai nilai gain terbesar adalah atribut jenis asuransi dengan gain 0,018377232 sehingga atribut jenis asuransi yang dimasukkan ke dalam *node* pada *decision tree* yang divisualisasikan pada gambar 4.



Gambar 4 *Decision Tree Node* 1.1.1

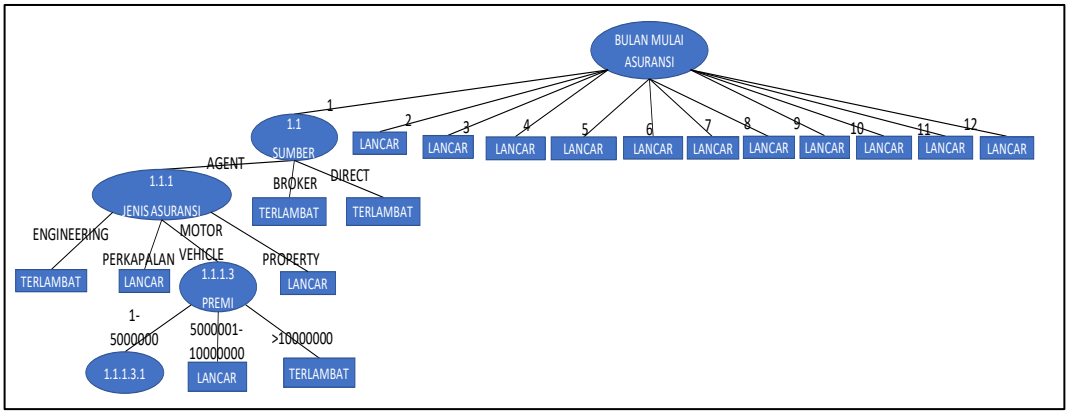
Karena data pada atribut jenis asuransi dengan kategori *motor vehicle* masih kurang tepat untuk mengklasifikasi data dimana tidak terdapat data yang dominan pada salah satu kategori pada atribut Terlambat yang merupakan atribut target, maka dibuat *node* 1.1.1.3 untuk dianalisis lebih lanjut. Lalu, akan dihitung jumlah data, jumlah lancar, jumlah terlambat, *entropy* dan *gain* dari setiap kategori yang ada dimana data tersebut mempunyai atribut bulan dengan kategori 1, atribut sumber dengan kategori *agent* dan atribut jenis asuransi dengan kategori *motor vehicle*. Karena atribut bulan, sumber dan jenis asuransi sudah digunakan pada *decision tree*, maka atribut bulan, sumber dan jenis asuransi tidak digunakan lagi. Hasilnya dapat dilihat pada Tabel 4.

DOI: https://doi.org/10.31001/tekinfo.v9i1.1090
E-ISSN: 2303-1867 | P-ISSN: 2303-1476

Tabel 4 Perhitungan C4.5 *Node* 1.1.1.3

Atribut	Kategori	Jumlah	Lancar	Terlambat	Entropy	Gain
Total		174	94	80	0,995325107	
Premi						0,027256375
	1-5.000.000	116	60	56	0,999142104	
	5.000.001-10.000.000	48	28	20	0,979868757	
	>1.000.000	6	2	4	0,918295834	
Kepemilikan						0,001110955
	Pribadi	125	66	59	0,99773667	
	Kelompok	49	28	21	0,985228136	

Atribut yang mempunyai nilai gain terbesar adalah atribut premi dengan gain 0,027256375 sehingga atribut premi yang dimasukkan ke dalam *node* pada *decision tree* yang divisualisasikan pada gambar 5.



Gambar 5 *Decision Tree Node* 1.1.1.3

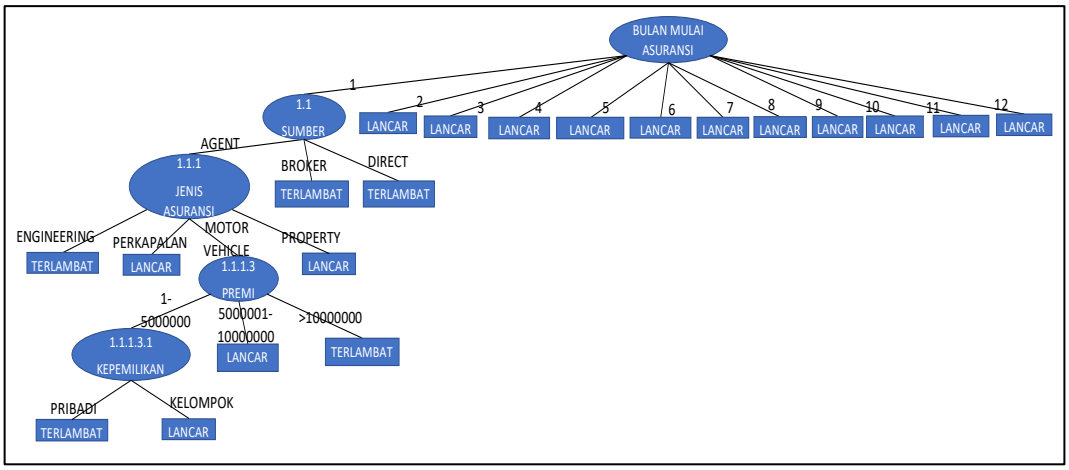
Karena data pada atribut premi dengan kategori 1-5000000 masih kurang tepat untuk mengklasifikasi data dimana tidak terdapat data yang dominan pada salah satu kategori pada atribut Terlambat yang merupakan atribut target, maka dibuat *node* 1.1.1.3.1 untuk dianalisis lebih lanjut. Lalu, akan dihitung jumlah data, jumlah lancar, jumlah terlambat, *entropy* dan *gain* dari setiap kategori yang ada dimana data tersebut mempunyai atribut bulan dengan kategori 1, atribut sumber dengan kategori *agent*, atribut jenis asuransi dengan kategori *motor vehicle* dan atribut premi dengan kategori 1-5000000. Karena atribut bulan, sumber, jenis asuransi dan premi sudah digunakan pada *decision tree*, maka atribut bulan, sumber, jenis asuransi dan premi tidak digunakan lagi. Hasilnya dapat dilihat pada Tabel 5.

Tabel 5 Perhitungan C4.5 *Node* 1.1.1.3.1

Atribut	Kategori	Jumlah	Lancar	Terlambat	Entropy	Gain
Total		116	60	56	0,999142104	
Kepemilikan						0,006064392
	Pribadi	82	40	42	0,999570839	
	Kelompok	34	20	14	0,977417818	

DOI: https://doi.org/10.31001/tekinfo.v9i1.1090
E-ISSN: 2303-1867 | P-ISSN: 2303-1476

Atribut yang mempunyai nilai gain terbesar adalah atribut kepemilikan dengan gain 0,007909951 sehingga atribut premi yang dimasukkan ke dalam *node* pada *decision tree* yang divisualisasikan pada gambar 6.



Gambar 6 *Decision Tree Node* 1.1.1.3.1

Karena semua *node* sudah dianalisis, maka *decision tree* pada kasus ini sudah selesai.

3.2 Pengujian

Pengujian hasil *decision tree* bertujuan untuk menghitung akurasi dari *decision tree* yang telah dibentuk. Semakin tinggi akurasi *decision tree* maka semakin akurat *decision tree* yang telah dibentuk untuk mengklasifikasi data-data nasabah asuransi. Data yang digunakan untuk menguji *decision tree* yang telah dibentuk adalah data pembayaran premi nasabah asuransi yang sudah diketahui statusnya lancar atau terlambat yang dipilih secara acak. Jumlah data yang digunakan untuk pengujian adalah sebanyak 200 data.

Tabel 6 Hasil Pengujian

	Hasil yang Sebenarnya	**Hasil *Testing***
Jumlah Lancar	176	200
Jumlah Terlambat	24	0

Dari hasil pengujian, terdapat 176 data yang dapat diklasifikasikan dengan akurat, sedangkan terdapat 24 data yang tidak dapat diklasifikasikan dengan akurat. Untuk menentukan akurasi dari *decision tree* yang telah dibentuk, digunakan rumus (3).

$$Akurasi = \frac{176}{200}\ x\ 100\%$$

$$= 88\%$$

DOI: https://doi.org/10.31001/tekinfo.v9i1.1090
E-ISSN: 2303-1867 | P-ISSN: 2303-1476

Berdasarkan perhitungan diatas, didapatkan akurasi dari *decision tree* yang telah dibentuk sebesar 88% dengan menggunakan 200 data yang dipilih secara acak.

4. KESIMPULAN

Berdasarkan hasil implementasi dan pengujian sistem yang telah dilakukan, sistem untuk melakukan klasifikasi keterlambatan pembayaran premi nasabah asuransi berhasil dibangun. Sistem dapat mengimplementasikan algoritma C4.5 sehingga dapat menghasilkan pohon keputusan (*decision tree*) dan sistem juga dapat mengklasifikasikan data baru sesuai dengan aturan-aturan klasifikasi yang ada pada pohon keputusan (*decision tree*) dimana tingkat akurasi klasifikasi yang didapatkan sebesar 88%. Terbentuknya pola yang mengklasifikasi data nasabah dengan menggunakan metode C4.5 memberikan visualisasi yang dapat dijadikan opsi dalam menentukan keputusan bagi pimpinan perusahaan.

DAFTAR PUSTAKA

Ainurrochmah, A., Hayati, M. N. and Satriya, A. M. A. (2019) 'Perbandingan Klasifikasi Analisis Diskriminan Fisher dan Metode Naïve Bayes', *Jurnal Aplikasi Statistika & Komputasi Statistik*, 11(2), pp. 37–48.

Amin, R. K., Indwiarti and Sibaroni, Y. (2015) 'Implementasi Klasifikasi Decision Tree dengan Algoritma C4.5 dalam Pengambilan Keputusan Permohonan Kredit oleh Debitur (Studi Kasus: Bank Pasar Daerah Istimewa Yogyakarta)', *e-Proceeding of Engineering*, 2(1), pp. 1768–1778.

Andriani, A. (2012) 'Penerapan Algoritma C4.5 Pada Program Klasifikasi Mahasiswa Dropout', in *Seminar Nasional Matematika*, pp. 139–147.

Ardyanti, H., Goejantoro, R. and Amijaya, F. D. T. (2020) 'Perbandingan Metode Klasifikasi Naïve Bayes Dan Jaringan Saraf Tiruan (Studi Kasus: Pt Asuransi Jiwa Bersama Bumiputera Tahun 2018)', *Jurnal EKSPONENSIAL*, 11(2), pp. 145–152.

Avegad, J. and Wibowo, A. (2019) 'Data Mining Klasifikasi Untuk Memprediksi Status Keberlanjutan Polis Asuransi Kesehatan Dengan Algoritme Naïve Bayes', in *Seminar Nasional Teknologi Informasi dan Komunikasi STI&K (SeNTIK)*. Jakarta: STMIK Jakarta STI&K, pp. 219–223.

Betrisandi (2017) 'Klasifikasi Nasabah Asuransi Jiwa Menggunakan Algoritma Naive Bayes Berbasis Backward Elimination', *ILKOM Jurnal Ilmiah*, 9(1), pp. 96–101.

Bustami (2013) 'Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi', *TECHSI: Jurnal Penelitian Teknik Informatika*, 6(2), pp. 127–146.

Grassella, Purnamasari, I. and Amijaya, F. D. T. (2019) 'Klasifikasi Status Pembayaran Premi Menggunakan Algoritma Neighbor Weighted K-Nearest Neighbor (NWKNN) (Studi Kasus: PT. Bumiputera Kota Samarinda)', *VARIANCE: Journal of Statistics and Its Applications*, 1(2), pp. 56–63.

Hanun, N. L. and Zailani, A. U. (2020) 'Penerapan Algoritma Klasifikasi Random Forest untuk Penentuan Kelayakan Pemberian Kredit di Koperasi Mitra Sejahtera', *Journal Of Technology Information*, 6(1), pp. 7–14.

Hariono, M. F. *et al.* (2014) 'Developing Review Websites Using Feature Driven Development (FDD)', *ULTIMATICS*, 6(2), pp. 100–104.

Kamagi, D. H. and Hansun, S. (2014) 'Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa', *Jurnal ULTIMATICS*, 6(1), pp. 15–20.

Kurniawan, D. A. and Kriestanto, D. (2016) 'Penerapan Naïve Bayes Untuk Prediksi Kelayakan Kredit', *JIKO (Jurnal Informatika dan Komputer)*, 1(1), pp. 19–23.

Kusumawati, A., Wibisono, Y. Y. and Aritonang, K. (2014) 'Perbaikan Proses Bisnis untuk Mengurangi Piutang di PT. Asuransi Astra Buana Cabang Bandung', *Jurnal Rekayasa Sistem Industri*, 3(1), pp. 20–26.

Mardi, Y. (2017) 'Data Mining : Klasifikasi Menggunakan Algoritma C4.5', *Jurnal Edik Informatika*, 2(2), pp. 213–219.

Novilla, D. A., Goejantoro, R. and Amijaya, F. D. T. (2019) 'Klasifikasi Data Nasabah Asuransi Dengan Menggunakan Metode Naive Bayes (Studi Kasus : PT . Prudential Life Jalan Mt . Haryono Samarinda) Classification of Insurance Data Customers Using Naive Bayes Method (Case Study : PT . Prudential Life MT . Haryon', *Jurnal EKSPONENSIAL*, 10(2), pp. 95–102.

Palmer, S. R. and Felsing, J. M. (2002) *A Practical Guide to Feature-Driven Development the Coad Series*. Upper Saddle River, New Jersey: Prentice Hall PTR.

Putranto, R. A., Wuryandari, T. and Sudarno (2015) 'Perbandingan Analisis Klasifikasi antara Decision Tree dan Support Vector Machine Multiclass untuk Penentuan Jurusan pada Siswa SMA', *JURNAL GAUSSIAN*, 4(4), pp. 1007–1016.

Santoso, T. B. and Sekardiana, D. (2019) 'Penerapan Algoritma C4.5 untuk Penentuan Kelayakan Pemberian Kredit (Studi Kasus : Koperia - Koperasi Warga Komplek Gandaria)', *Jurnal Algoritma, Logika dan Komputasi*, 2(1), pp. 130–137.

Setiawan, R. (2020) 'Analisis Kelayakan Pemberian Kredit Nasabah Koperasi Menggunakan Algoritma C4.5', *Techno Xplore : Jurnal Ilmu Komputer dan Teknologi Informasi*, 5(2), pp. 74–78.

Sucipto, A. (2015) 'Prediksi Kredit Macet melalui Perilaku Nasabah pada Koperasi Simpan Pinjam dengan Menggunakan Metode Algoritma Klasifikasi C4.5', *Jurnal DISPROTEK*, 6(1), pp. 75–87.

DOI: https://doi.org/10.31001/tekinfo.v9i1.1090
E-ISSN: 2303-1867 | P-ISSN: 2303-1476