

Penerapan Algoritma C4.5 Untuk Klasifikasi Keberhasilan Pengiriman Barang

Ghofar Taufik¹, Dona Jatmika²

Universitas Bina Sarana Informatika, Jl. Kramat Raya No. 98, Senen, Jakarta Pusat, Indonesia¹

STMIK Nusa Mandiri, Jl. Jatiwaringin Raya No.2, Jakarta Timur, Indonesia²

Email: ghofar.gft@bsi.ac.id¹, dona.bobop@gmail.com²

Abstrack – The success rate of shipping goods is a very important factor in the world of ecommerce business with logistics delivery services being the last link in the chain with customers. This greatly affects the satisfaction of customers who expect the goods they buy can be sent on time and in good condition when received. One technique of data mining that can be used to help predictions is to use classification techniques. Classification can be done with a decision tree that is with the C4.5 algorithm. The purpose of this study is to make a classification and apply data mining classification based on the decision tree and the rules generated. Furthermore, the results of the data classification are evaluated using a confusion matrix and ROC curve in the rapid miner application to determine the level of accuracy results. In this study produced an accuracy value of 93% and produced an AUC (Area Under Curve) value of 0.73 with a sufficient classification accuracy (Fair classification). From the results of this study can be a recommendation for the distribution section in choosing logistics services in shipping goods based on categories of goods, the destination address so that the risk of failure in shipping can be reduced.

Keywords - C4.5 algorithm, decision tree, classification.

Intisari – Tingkat keberhasilan pengiriman barang merupakan faktor yang sangat penting dalam dunia bisnis *ecommerce* dengan jasa logistik pengiriman menjadi rantai penghubung terakhir dengan *customer*. Hal ini sangat mempengaruhi kepuasan *customer* yang berharap barang yang mereka beli dapat terkirim tepat waktu dan dalam keadaan baik pada saat diterima. Salah satu teknik dari *data mining* yang dapat digunakan untuk membantu prediksi adalah menggunakan teknik klasifikasi. Klasifikasi dapat dilakukan dengan *decision tree* yaitu dengan algoritma C4.5. Tujuan dari penelitian ini adalah membuat klasifikasi serta menerapkan klasifikasi *data mining* berdasarkan pohon keputusan dan *rules* yang di hasilkan. Selanjutnya hasil klasifikasi data dievaluasi dengan menggunakan *confusion matrix* dan *kurva ROC* pada aplikasi rapid miner untuk mengetahui tingkat hasil akurasi. Pada penelitian ini menghasilkan nilai akurasi sebesar 93% dan menghasilkan nilai *AUC (Area Under Curve)* sebesar 0.739 dengan nilai akurasi klasifikasi cukup (*Fair classification*). Dari hasil penelitian ini dapat menjadi rekomendasi bagian distribusi dalam memilih jasa logistik dalam pengiriman barang berdasarkan kategori barang, alamat tujuan agar resiko kegagalan dalam pengiriman dapat dikurangi.

Kata Kunci - algoritma C4.5, decision tree, klasifikasi.

I. PENDAHULUAN

Setiap perusahaan selalu mengutamakan kepuasan konsumennya apalagi di perusahaan yang bergerak di bidang E-Commerce. Keberhasilan pengiriman merupakan faktor penting dalam dunia bisnis e-commerce dimana logistik pengiriman menjadi penghubung dengan para customer secara langsung dan sangat mempengaruhi kepuasan customer yang berharap barang yang mereka beli dapat sampai dan dengan tepat waktu.

PT. Global Digital Niaga dengan nama brandnya Bilibli.com adalah perusahaan yang bergerak di bidang E-commerce, dalam kegiatan bisnisnya menjual berbagai macam jenis

barang dan membuka pengiriman ke seluruh Indonesia dengan gratis ongkos pengiriman. Di Indonesia sendiri yang merupakan negara kepulauan cukuplah sulit untuk melakukan pengiriman dengan cepat dikarenakan jika dilakukan via jalur laut pengiriman akan sangat lama dan budget pengiriman juga akan sangat mahal, sehingga pengiriman difokuskan ke jalur udara, namun dalam pengiriman barang via udara kategori barang dibagi menjadi 2 kategori yaitu kategori barang berbahaya (*dangerous goods*) dan bukan barang berbahaya (*non dangerous goods*), untuk kategori barang berbahaya sendiri adalah jenis barang yang mengandung cairan, baterai, magnet dan aerosol yang mudah meledak dan dapat membahayakan penerbangan sehingga membutuhkan pengemasan dan penanganan khusus.

Blibli.com sendiri dalam bisnisnya menggunakan partner jasa logistik pengiriman JNE, POS Indonesia dan First Logistik. Logistik yang handal dipasaran sangatlah dibutuhkan dalam bisnis *e-commerce* ini dimana logistik dapat menangani barang berbahaya dan hafal wilayah pengiriman dimana kita berada di wilayah negara kepulauan, karena banyak kasus terjadi dimana ketika pengiriman gagal maka waktu pengiriman akan menjadi lama karena paket akan di retur kembali dan ketika kita berusaha untuk meminta customer menunggu karena kita akan mengganti jasa logistik lain namun banyak *customer* yang memilih membatalkan pembelian dan meminta dananya kembali karena pengiriman yang lama dan gagal ini.

Tujuan dari penelitian ini adalah penerapan *data mining* dengan metode algoritma C4.5 yang dapat menghasilkan informasi berupa klasifikasi keberhasilan pengiriman barang dengan menggunakan data training sebagai *data testing* yang dapat dibangun sebuah *decision tree* yang dapat menghasilkan *rules* yang berlaku. Sedangkan manfaat dari penelitian ini adalah dapat menjadi rekomendasi bagi pendistribusian barang dalam memilih jasa logistik untuk pengiriman barang berdasarkan alamat pengiriman, dan kategori barang untuk meminimalisir resiko kegagalan pengiriman.

II. SIGNIFIKANSI STUDI

A. Data Mining

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [1]. Data mining dapat didefinisikan sebagai proses seleksi, eksplorasi dan pemodelan dari sejumlah besar data untuk dapat menemukan suatu pola ataupun kecenderungan yang biasanya tidak disadari keberadaannya [2]. Data mining melakukan proses pengambilan pengetahuan dengan menggunakan data dalam jumlah besar yang disimpan dalam basis data [3].

Data mining merupakan teknik penggabungan metode-metode analisis data secara terus menerus dengan menggunakan beberapa algoritma untuk memproses jumlah data dengan skala yang besar [4]. Data mining merupakan suatu proses untuk mencari pola atau informasi menarik di dalam data yang terpilih dengan menggunakan metode tertentu [5]. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data [6].

B. Klasifikasi

Klasifikasi adalah proses dari mencari suatu himpunan model (fungsi) yang dapat mendeskripsikan dan membedakan kelas-kelas data atau konsep-konsep, dengan tujuan dapat menggunakan model tersebut untuk memprediksi kelas dari suatu objek yang mana kelasnya belum diketahui [1]. Proses klasifikasi didasarkan pada 4 komponen [7] :

1. Kelas

Merupakan variabel dependen yang berupa kategorikal yang merepresentasikan 'label' yang terdapat pada objek.

2. *Predictor*

Merupakan variabel independen yang direpresentasikan oleh karakteristik (atribut) data.

3. *Training dataset*

Merupakan satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor*.

4. *Testing dataset*

Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi.

C. *Metodologi Penelitian*

Dalam penelitian ini tahapan yang dilakukan adalah meliputi:

1. *Pengumpulan Data*

Karena banyaknya data pengiriman yang akan digunakan sebagai data testing, maka jumlah data yang dikumpulkan adalah kurun waktu satu bulan sebanyak 50.000 data pengiriman barang. Maka sampel yang akan digunakan untuk data testing diambil dengan menggunakan rumus yang didasarkan pada perhitungan *slovin*. Dalam perhitungan menggunakan rumus *slovin* digunakan taraf kepercayaan. Tingkat signifikan diterjemahkan sebagai taraf kepercayaan yang berarti presentase kebenaran bukan secara kebetulan. Dalam penelitian ini akan digunakan taraf signifikan 0,1 atau taraf kepercayaan 90%. Rumus *slovin* dituliskan sebagai berikut [8] :

$$n = \frac{N}{1+Ne^2} \quad (1)$$

n : besaran sampel

N : besaran populasi

e : nilai kritis (batas penelitian) yang diinginkan (persen kelonggaran ketidaktelitian karena kesalahan penarikan sampel) yaitu sebesar 10% dengan tingkat kepercayaan 90%.

Berikut adalah perhitungan dalam mencari jumlah sampel yang akan digunakan di penelitian ini berdasarkan jumlah populasi yang ada yaitu data pengiriman dalam waktu satu bulan yaitu:

$$n = \frac{50,000}{1 + 50,000(0.1)^2} = 99.80 = 100$$

Berdasarkan hasil perhitungan di atas maka di peroleh ukuran sampel yang akan digunakan yaitu sebesar 100 data pengiriman barang.

2. *Pembentukan Model Algoritma C4.5*

Algoritma C4.5 Salah satu algoritma yang dapat digunakan untuk membuat pohon keputusan (*decission tree*) adalah algoritma C4.5 [9]. Algoritma C4.5 merupakan model *tree* mirip sebuah pohon dimana terdapat node internal (bukan daun) yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas [10]. Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan record. Atribut menyatakan sebagai suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon keputusan [11].

Dalam algoritma C4.5 dapat melakukan pemilihan dalam pemecahan kasus-kasus yang terbaik dengan melakukan penghitungan serta membandingkan gain ratio sampai node-node

yang terbentuk pada tahap level berikutnya [12]. Algoritma C4.5 mempunyai input berupa *training samples*. *Training samples* berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya [13]. Tahapan dalam membuat sebuah pohon keputusan dengan algoritma C4.5, yaitu [14] :

- 1) Mempersiapkan data training, dapat diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu.
- 2) Menentukan akar dari pohon dengan menghitung nilai gain yang tertinggi dari masing-masing atribut atau berdasarkan nilai index entropy terendah. Dalam data mining entropy merupakan keberbedaan atau keragaman yang didefinisikan sebagai suatu parameter yang digunakan untuk mengukur tingkat heterogenitas (keberagaman) pada suatu himpunan data [15]. Semakin heterogen dalam suatu himpunan data maka akan semakin besar nilai entropy pada himpunan data tersebut. Sebelumnya dihitung terlebih dahulu nilai index entropy, dengan rumus:

$$Entropy(i) = \sum_{j=1}^m f(i,j). \log_2 f[(i,j)] \quad (2)$$

i : himpunan kasus
 m : jumlah partisi i
 f(i,j) : proposi j terhadap i

- 3) Hitung nilai gain dengan rumus:

$$gain(S,A) = Entropy(s) - \sum_i^n \frac{|s_i|}{|s|} * Entropy(s_i) \quad (3)$$

S : himpunan kasus
 A : fitur
 N : jumlah partisi atribut A
 |Si| : proporsi Si terhadap S
 |S| : jumlah kasus dalam S

- 4) Ulangi langkah ke-2 hingga semua record terpartisi. Proses partisi pohon keputusan akan berhenti disaat:
 - a. Semua tupel dalam record dalam simpul *m* mendapat kelas yang sama
 - b. Tidak ada atribut dalam record yang dipartisi lagi
 - c. Tidak ada record didalam cabang yang kosong.

3. Pengujian

Setelah data testing yang telah dikumpulkan yaitu sebanyak 100, kemudian dilakukan pengujian terhadap model algoritma C4.5 dengan menggunakan *tools RapidMiner*. Yang hasilnya berupa klasifikasi terhadap keberhasilan pengiriman barang.

Penelitian dengan model algoritma C4.5 ini memberikan manfaat bagi perusahaan untuk dapat mengetahui tingkat keberhasilan dari pengiriman yang dilakukan oleh bagian distribusi barang melalui jasa logistik pengiriman barang yang telah dipilih berdasarkan klasifikasi keberhasilannya. Dan dapat dijadikan sebagai prediksi tingkat keberhasilan pengiriman barang kepada konsumen dimasa yang akan datang.

III. HASIL DAN PEMBAHASAN

A. Data Training

Data Training adalah data yang digunakan untuk pembelajaran pada proses data mining atau proses pembentukan pohon keputusan. Pada penelitian ini *data training* diambil berdasarkan sampel yang telah ditentukan menggunakan rumus *slovin* yang menghasilkan 100 data pengiriman. Berikut ini merupakan hasil total *data training*:

TABEL I
DATA TRAINING PENGIRIMAN BARANG

No	Logistik	Alamat Tujuan	Kategori Barang	Kondisi	Status Pengiriman
1	First Logistik	Indonesia Timur	Dangerous Goods	Baik	Tidak
2	JNE	Indonesia Timur	Not Dangerous Goods	Baik	Ya
3	JNE	Jabodetabek	Not Dangerous Goods	Rusak/Hilang	Tidak
4	JNE	Indonesia Timur	Not Dangerous Goods	Baik	Ya
5	POS	Indonesia Timur	Dangerous Goods	Baik	Ya
6	POS	Indonesia Timur	Not Dangerous Goods	Baik	Ya
7	POS	Jawa, Sumatera, Bali	Dangerous Goods	Baik	Ya
8	POS	Jawa, Sumatera, Bali	Dangerous Goods	Baik	Ya
9	POS	Jawa, Sumatera, Bali	Dangerous Goods	Baik	Ya
10	JNE	Jabodetabek	Dangerous Goods	Rusak/Hilang	Tidak
11	JNE	Jabodetabek	Dangerous Goods	Baik	Ya
12	JNE	Jabodetabek	Dangerous Goods	Baik	Ya
13	First Logistik	Jabodetabek	Dangerous Goods	Baik	Ya
14	JNE	Jawa, Sumatera, Bali	Dangerous Goods	Baik	Tidak
15	First Logistik	Jabodetabek	Dangerous Goods	Baik	Ya
16	First Logistik	Jabodetabek	Dangerous Goods	Baik	Ya
17	POS	Indonesia Timur	Not Dangerous Goods	Baik	Ya
18	POS	Jabodetabek	Not Dangerous Goods	Baik	Ya
19	First Logistik	Jabodetabek	Not Dangerous Goods	Baik	Ya
20	JNE	Indonesia Timur	Not Dangerous Goods	Baik	Ya
21	POS	Jabodetabek	Dangerous Goods	Baik	Ya
22	POS	Jabodetabek	Dangerous Goods	Baik	Ya
23	JNE	Jabodetabek	Dangerous Goods	Baik	Ya
24	JNE	Indonesia Timur	Not Dangerous Goods	Baik	Ya
25	JNE	Indonesia Timur	Not Dangerous Goods	Baik	Ya
26	JNE	Indonesia Timur	Dangerous Goods	Baik	Tidak
27	POS	Indonesia Timur	Dangerous Goods	Baik	Ya
28	JNE	Indonesia Timur	Not Dangerous Goods	Baik	Ya
29	JNE	Indonesia Timur	Not Dangerous Goods	Baik	Ya
30	POS	Indonesia Timur	Dangerous Goods	Baik	Ya
31	First Logistik	Indonesia Timur	Not Dangerous Goods	Baik	Ya
32	First Logistik	Indonesia Timur	Not Dangerous Goods	Baik	Ya
33	POS	Indonesia Timur	Not Dangerous Goods	Baik	Ya
34	JNE	Indonesia Timur	Not Dangerous Goods	Rusak/Hilang	Tidak
35	POS	Indonesia Timur	Not Dangerous Goods	Baik	Ya
36	First Logistik	Indonesia Timur	Dangerous Goods	Baik	Tidak
37	First Logistik	Indonesia Timur	Not Dangerous Goods	Baik	Ya
38	First Logistik	Indonesia Timur	Not Dangerous Goods	Baik	Ya
39	First Logistik	Indonesia Timur	Not Dangerous Goods	Baik	Ya
40	First Logistik	Jawa, Sumatera, Bali	Dangerous Goods	Baik	Tidak
41	First Logistik	Jawa, Sumatera, Bali	Dangerous Goods	Baik	Ya
42	First Logistik	Jawa, Sumatera, Bali	Dangerous Goods	Baik	Ya
43	POS	Jawa, Sumatera, Bali	Dangerous Goods	Baik	Ya
44	POS	Jawa, Sumatera, Bali	Dangerous Goods	Baik	Ya
45	POS	Jawa, Sumatera, Bali	Dangerous Goods	Baik	Ya
46	POS	Indonesia Timur	Dangerous Goods	Baik	Tidak
47	JNE	Indonesia Timur	Not Dangerous Goods	Baik	Ya

No	Logistik	Alamat Tujuan	Kategori Barang	Kondisi	Status Pengiriman
48	POS	Indonesia Timur	Not Dangerous Goods	Baik	Ya
49	First Logistik	Jawa, Sumatera, Bali	Dangerous Goods	Baik	Ya
50	First Logistik	Jawa, Sumatera, Bali	Dangerous Goods	Baik	Ya
51	First Logistik	Jawa, Sumatera, Bali	Dangerous Goods	Baik	Ya
52	POS	Jawa, Sumatera, Bali	Dangerous Goods	Baik	Ya
53	POS	Jawa, Sumatera, Bali	Dangerous Goods	Baik	Ya
54	JNE	Jabodetabek	Not Dangerous Goods	Baik	Ya
55	JNE	Jabodetabek	Not Dangerous Goods	Baik	Ya
56	JNE	Jabodetabek	Dangerous Goods	Baik	Ya
57	JNE	Jabodetabek	Dangerous Goods	Baik	Ya
58	JNE	Jabodetabek	Dangerous Goods	Baik	Ya
59	JNE	Indonesia Timur	Not Dangerous Goods	Baik	Ya
60	JNE	Indonesia Timur	Not Dangerous Goods	Baik	Ya
61	First Logistik	Jabodetabek	Dangerous Goods	Baik	Ya
62	First Logistik	Jabodetabek	Dangerous Goods	Baik	Ya
63	First Logistik	Jabodetabek	Dangerous Goods	Baik	Ya
64	First Logistik	Jabodetabek	Dangerous Goods	Baik	Ya
65	First Logistik	Jabodetabek	Dangerous Goods	Baik	Ya
66	POS	Jabodetabek	Dangerous Goods	Baik	Ya
67	POS	Jabodetabek	Dangerous Goods	Baik	Ya
68	POS	Jabodetabek	Dangerous Goods	Baik	Ya
69	POS	Jabodetabek	Dangerous Goods	Baik	Ya
70	POS	Jabodetabek	Dangerous Goods	Baik	Ya
71	POS	Jabodetabek	Dangerous Goods	Baik	Ya
72	JNE	Indonesia Timur	Not Dangerous Goods	Baik	Ya
73	JNE	Jabodetabek	Not Dangerous Goods	Baik	Ya
74	JNE	Jabodetabek	Not Dangerous Goods	Baik	Ya
75	JNE	Jabodetabek	Not Dangerous Goods	Baik	Ya
76	JNE	Jabodetabek	Not Dangerous Goods	Baik	Ya
77	JNE	Jabodetabek	Not Dangerous Goods	Baik	Ya
78	JNE	Jabodetabek	Not Dangerous Goods	Baik	Ya
79	First Logistik	Indonesia Timur	Dangerous Goods	Baik	Tidak
80	JNE	Jabodetabek	Not Dangerous Goods	Baik	Ya
81	JNE	Jabodetabek	Not Dangerous Goods	Baik	Ya
82	First Logistik	Jabodetabek	Not Dangerous Goods	Baik	Ya
83	First Logistik	Jabodetabek	Not Dangerous Goods	Baik	Ya
84	First Logistik	Jabodetabek	Not Dangerous Goods	Baik	Ya
85	First Logistik	Jabodetabek	Not Dangerous Goods	Baik	Ya
86	POS	Indonesia Timur	Not Dangerous Goods	Baik	Ya
87	POS	Jabodetabek	Not Dangerous Goods	Baik	Ya
88	JNE	Jawa, Sumatera, Bali	Not Dangerous Goods	Baik	Ya
89	JNE	Jawa, Sumatera, Bali	Not Dangerous Goods	Baik	Ya
90	JNE	Jawa, Sumatera, Bali	Not Dangerous Goods	Baik	Ya
91	First Logistik	Jawa, Sumatera, Bali	Dangerous Goods	Baik	Tidak
92	JNE	Jawa, Sumatera, Bali	Not Dangerous Goods	Baik	Ya
93	First Logistik	Jawa, Sumatera, Bali	Not Dangerous Goods	Baik	Ya
94	First Logistik	Jawa, Sumatera, Bali	Not Dangerous Goods	Baik	Ya
95	First Logistik	Jawa, Sumatera, Bali	Not Dangerous Goods	Baik	Ya
96	First Logistik	Jawa, Sumatera, Bali	Not Dangerous Goods	Baik	Ya
97	POS	Jawa, Sumatera, Bali	Not Dangerous Goods	Baik	Ya
98	POS	Jawa, Sumatera, Bali	Not Dangerous Goods	Baik	Ya
99	POS	Jawa, Sumatera, Bali	Not Dangerous Goods	Baik	Ya
100	JNE	Jawa, Sumatera, Bali	Not Dangerous Goods	Baik	Ya

Atribut yang digunakan dalam penentuan keberhasilan dan kegagalan pengiriman adalah logistik, alamat tujuan, kategori barang, kondisi barang. Data yang telah didapat telah di

analisis bahwa data tersebut memiliki 89 data yang berjumlah “Ya” (berhasil) dan 11 data yang berjumlah “Tidak” (gagal).

B. Pemodelan Menggunakan Algoritma C4.5

Berikut ini merupakan uraian langkah langkah perhitungan dalam algoritma C4.5 dalam penyelesaian kasus keberhasilan pengiriman barang yang akan dibagi menjadi label “Ya” (Berhasil) atau “Tidak” (Gagal).

1. Perhitungan Mencari Entropy

$$Entropy (S) = \left(- \left(\frac{Sum (Ya)}{Total} \right) * \log_2 \left(\frac{Sum (Ya)}{Total} \right) \right) + \left(- \left(\frac{Sum (Tidak)}{Total} \right) * \log_2 \left(\frac{Sum (Tidak)}{Total} \right) \right)$$

$$Entropy (Seluruh Kasus) = \left(- \left(\frac{89}{100} \right) * \log_2 \left(\frac{11}{100} \right) \right) + \left(- \left(\frac{11}{100} \right) * \log_2 \left(\frac{89}{100} \right) \right) = 0,499915958$$

Sesuai dari hasil perhitungan diatas di dapat nilai Entropy dari seluruh kasus adalah 0,499915958 dan untuk selanjutnya hitung pula nilai entropy untuk setiap nilai atribut seperti rumus entropy seluruh kasus tersebut dan di dapat hasil sesuai tabel II untuk entropy seluruh kasus dan setiap nilai atribut.

TABEL II
PERHITUNGAN ENTROPY

Atribut	Nilai Atribut	Sum Total	Sum (Tidak)	Sum (Ya)	Entropy
Total	Total	100	11	89	0,499915958
Logistik	First Logistik	32	5	27	0,625262405
	JNE	37	5	32	0,571354974
	POS	31	1	30	0,205592508
Alamat Tujuan	Indonesia Timur	31	6	25	0,708835673
	Jabodetabek	41	2	39	0,281193796
	Jawa, Sumatera, Bali	28	3	25	0,491237342
Kategori Barang	Dangerous Goods	47	9	38	0,704576713
	Not Dangerous Goods	53	2	51	0,23181305
Kondisi Barang	Baik	97	8	89	0,410837839
	Rusak/ Hilang	3	3	0	0

2. Perhitungan mencari Gain

a. Gain Logistik

$$Gain (S.A) = Entropy (total) - \left(\frac{Sum (total)}{jumlah kasus} \right)$$

$$Gain (logistik) = 0,499916 - \left(\frac{32}{100} \right) * 0,625262 + \left(\frac{37}{100} \right) * 0,571355 + \left(\frac{31}{100} \right) * 0,205593 = 0,024697$$

Diketahui entropy seluruh kasus adalah 0,499916 dengan jumlah seluruh kasus 100. Jumlah kasus nilai atribut first logistik 32, untuk entropy first logistik 0,625262. Jumlah kasus nilai atribut JNE 37 dengan entropy JNE 0,571355. Jumlah kasus Nilai Atribut POS 31 dengan entropy POS 0,205593. Dari hasil perhitungan didapat hasil untuk nilai Gain atribut logistik sebesar 0,024697.

b. Gain Alamat

$$Gain (alamat) = 0,499916 - \left(\frac{31}{100} \right) * 0,708836 + \left(\frac{41}{100} \right) * 0,281194 + \left(\frac{28}{100} \right) * 0,491237 = 0,027341$$

Diketahui entropy seluruh kasus 0,499916 dengan jumlah seluruh kasus 100. Jumlah kasus nilai atribut Indonesia Timur 31 dan entropy Indonesia Timur 0,708836. Jumlah kasus nilai atribut Jabodetabek 41, entropy Jabodetabek 0,281194. Jumlah kasus Nilai Atribut Jawa, Sumatera, Bali 28 dan entropy Jawa, Sumatera, Bali 0,491237. Dari hasil perhitungan maka hasil untuk nilai Gain atribut alamat tujuan sebesar 0,027341.

Untuk selanjutnya hitung juga gain untuk atribut kategori barang dan kondisi barang sehingga didapat hasil gain seperti pada tabel III.

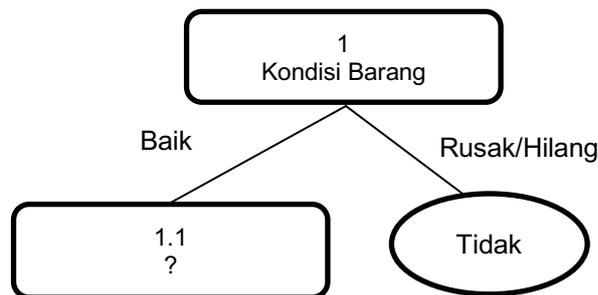
TABEL III
ENTROPY DAN GAIN

Atribut	Nilai Atribut	Sum Total	Sum (Tidak)	Sum (Ya)	Entropy	Gain
Total	Total	100	11	89	0,499916	
Logistik	First Logistik	32	5	27	0,625262	0,024697
	JNE	37	5	32	0,571355	
	POS	31	1	30	0,205593	
Alamat Tujuan	Indonesia Timur	31	6	25	0,708836	0,027341
	Jabodetabek	41	2	39	0,281194	
	Jawa, Sumatera, Bali	28	3	25	0,491237	
Kategori Barang	Dangerous Goods	47	9	38	0,704577	0,045904
	Not Dangerous Goods	53	2	51	0,231813	
Kondisi Barang	Baik	97	8	89	0,410838	0,101403
	Rusak/ Hilang	3	3	0	0,000000	

Setelah semua dihitung maka akan tampak nilai gain untuk masing masing atribut logistik 0,024697, alamat tujuan 0,027341, kategori barang 0,045904 dan kondisi barang 0,101403. Dan didapat kesimpulan atribut kondisi barang mempunyai nilai gain tertinggi yaitu 0,101403.

3. Menentukan Pohon Keputusan

Berdasarkan perhitungan entropy dan gain sebelumnya maka kondisi barang mempunyai gain terbesar yaitu 0,101403 maka atribut kondisi barang menjadi Node akar (root Node). Di gambarkan dalam pohon keputusan berikut ini :



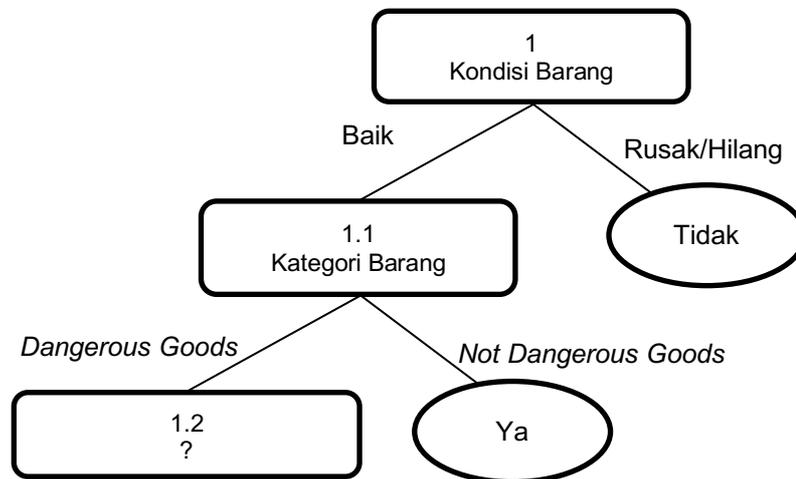
Gambar 1. Pohon Keputusan Node 1

Berdasarkan pohon keputusan Node 1 (root Node) yang telah terbentuk, Node 1.1 akan dilanjutkan ke tahap selanjutnya. Pada tabel data Training difilter dengan mengambil data kondisi barang yang baik saja, kemudian dilakukan kembali perhitungan entropy pada setiap nilai atribut dan hitung pula setiap gain pada setiap atribut. Sehingga didapat hasilnya seperti pada table IV di bawah ini, kemudian tentukan atribut yang mempunyai gain tertinggi untuk membentuk Node pohon keputusan berikutnya.

TABEL IV
HASIL ANALISIS NODE 1.1

Atribut	Nilai Atribut	Sum Total	Sum (Tidak)	Sum (Ya)	Entropy	Gain
Total	Total	97	8	89	0.410838	
Logistik	First Logistik	32	5	27	0.625262	0.02573
	JNE	34	2	32	0.322757	
	POS	31	1	30	0.205593	
Alamat Tujuan	Indonesia Timur	30	5	25	0.650022	0.068000
	Jabodetabek	39	0	39	0	
	Jawa, Sumatera, Bali	28	3	25	0.491237	
Kategori Barang	Dangerous Goods	46	8	38	0.666578	0.094729
	Not Dangerous Goods	51	0	51	0	

Gain tertinggi dimiliki oleh atribut kategori barang yaitu 0,094729 maka kategori barang dijadikan Node sebagai Node akar 1.1. Nilai yang di jadikan leaf (Cabang) adalah Dangerous Goods Dan Not DG maka pohon keputusan akan terlihat seperti gambar 2.



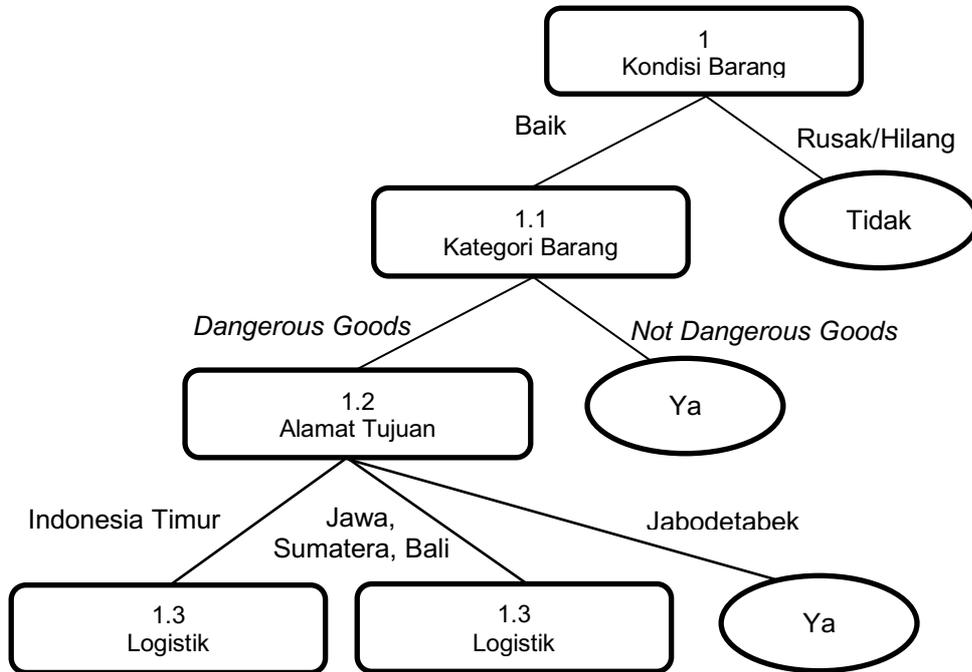
Gambar 2. Pohon Keputusan Node 1.1

Berdasarkan pohon keputusan *node* 1.1 maka akan dilanjutkan ke tahap selanjutnya. Pada tabel data *training* difilter dengan mengambil data kategori barang *dangerous goods*, kemudian dilakukan kembali perhitungan *entropy* pada setiap nilai atribut dan hitung pula setiap *gain* pada setiap atribut. Sehingga didapat hasilnya seperti pada table V di bawah ini, kemudian tentukan atribut yang mempunyai *gain* tertinggi untuk membentuk *node* pohon keputusan berikutnya.

TABEL V
HASIL ANALISIS NODE 1.2

Atribut	Nilai Atribut	Sum Total	Sum (Tidak)	Sum (Ya)	Entropy	Gain
Total	Total	46	8	38	0,66657836	
Logistik	First Logistik	18	5	13	0,85240518	0,067416
	JNE	8	2	6	0,81127812	
	POS	20	1	19	0,28639696	
Alamat Tujuan	Indonesia Timur	8	5	3	0,954434	0,258429
	Jabodetabek	22	0	22	0	
	Jawa, Sumatera, Bali	16	3	13	0,69621226	

Gain tertinggi dimiliki oleh atribut alamat tujuan dengan nilai 0,258429 maka alamat tujuan dijadikan *node* sebagai *node* akar 1.2. nilai yang di jadikan *leaf* (cabang) adalah Indonesia Timur, Jawa, Sumatera, Bali dan Jabodetabek maka pohon keputusan akan terlihat seperti :



Gambar 3. Pohon Keputusan Node 1.2

Berdasarkan pohon keputusan node 1.2 dimana alamat tujuan menjadi hasil node 1.2 maka atribut yang tersisa adalah logistik maka atribut logistik menjadi *node* terakhir dan difilter berdasarkan alamat tujuan Indonesia Timur dan Jawa Sumatera Bali dan Jabodetabek. Berikut hasil filter data training dan atribut terakhir.

TABEL VI
HASIL FILTER ALAMAT TUJUAN INDONESIA TIMUR

Atribut	Nilai Atribut	Sum Total	Sum (Tidak)	Sum (Ya)
Total	Total	8	5	3
Logistik	First Logistik	3	3	0
	JNE	1	1	0
	POS	4	1	3

Maka dapat di ambil keputusan dari masing masing logistik tujuan Indonesia Timur adalah First Logistik = Tidak, JNE = Tidak, POS = Ya.

TABEL VII
HASIL FILTER ALAMAT TUJUAN JAWA, SUMATERA, BALI

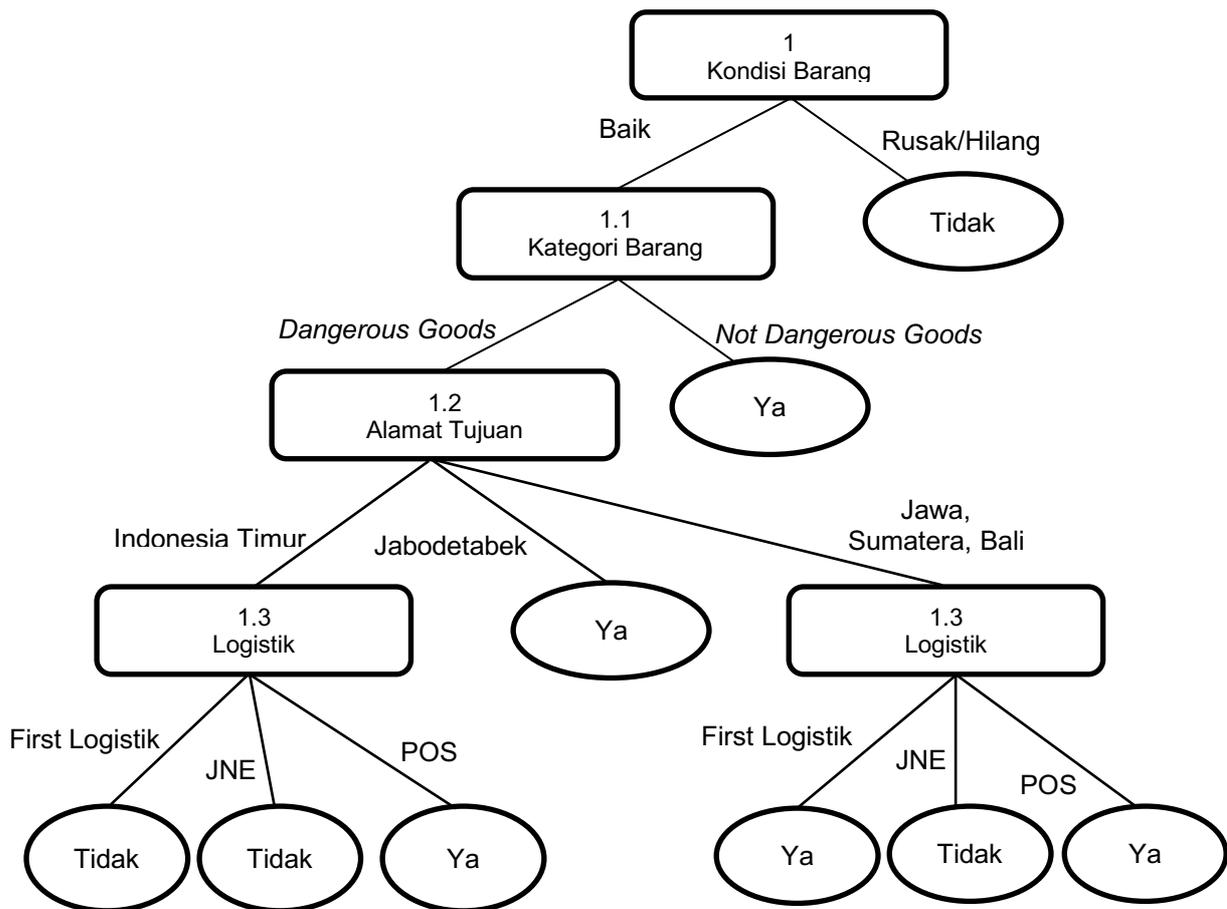
Atribut	Nilai Atribut	Sum Total	Sum (Tidak)	Sum (Ya)
Total	Total	16	3	13
Logistik	First Logistik	7	2	5
	JNE	1	1	0
	POS	8	0	8

Maka dapat di ambil keputusan dari masing masing logistik tujuan Jawa, Sumatera, Bali adalah First Logistik = Ya, JNE = Tidak, POS = Ya.

TABEL VIII
HASIL FILTER ALAMAT TUJUAN JABODETABEK

Atribut	Nilai Atribut	Sum Total	Sum (Tidak)	Sum (Ya)
Total	Total	27	5	22
Logistik	First Logistik	8	0	8
	JNE	6	0	6
	POS	8	0	8

Dapat di ambil keputusan dari masing masing logistik tujuan Jabodetabek adalah Ya. Sehingga dari hasil ini dapat digambarkan dalam pohon keputusan seluruhnya sebagai berikut pada gambar :



Gambar 4. Pohon Keputusan Terakhir

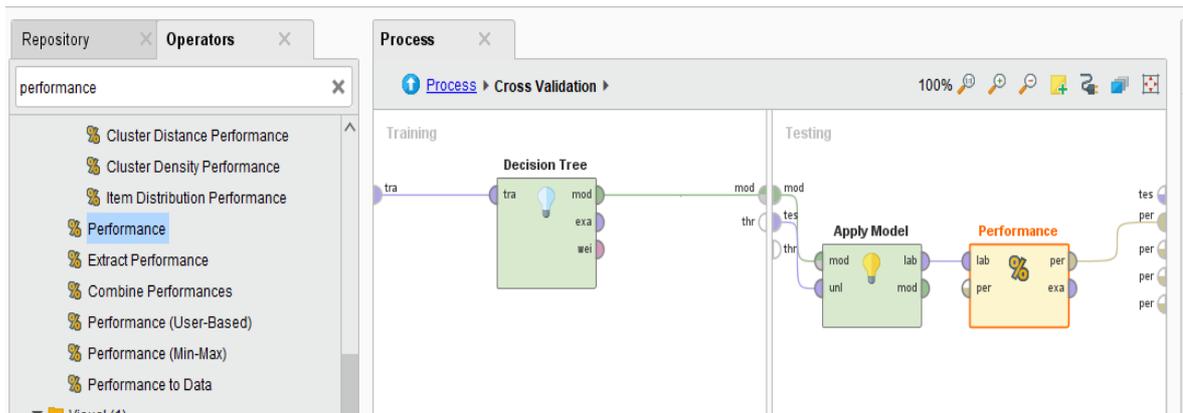
Dari pohon keputusan tersebut didapat Rules sebagai berikut :

1. **IF** (Kondisi Barang = Rusak/Hilang) **THEN** Tidak
2. **IF** (Kondisi barang = Baik **AND** Kategori Barang = Not Dangerous Goods) **THEN** Ya
3. **IF** (Kondisi barang = Baik **AND** Kategori Barang = Dangerous Goods **AND** Alamat Tujuan = Jabodetabek) **THEN** Ya
4. **IF** (Kondisi barang = Baik **AND** Kategori Barang = Dangerous Goods **AND** Alamat Tujuan = Indonesia Timur **AND** Logistik = First Logistik) **THEN** Tidak
5. **IF** (Kondisi barang = Baik **AND** Kategori Barang = Dangerous Goods **AND** Alamat Tujuan = Indonesia Timur **AND** Logistik = JNE) **THEN** Tidak
6. **IF** (Kondisi barang = Baik **AND** Kategori Barang = Dangerous Goods **AND** Alamat Tujuan = Indonesia Timur **AND** Logistik = POS) **THEN** Ya

7. **IF** (Kondisi barang = Baik **AND** Kategori Barang = Dangerous Goods **AND** Alamat Tujuan = Jawa Sumatra Bali **AND** Logistik = First Logistik) **THEN** Ya
8. **IF** (Kondisi barang = Baik **AND** Kategori Barang = Dangerous Goods **AND** Alamat Tujuan = Jawa Sumatra Bali **AND** Logistik = JNE) **THEN** Tidak
9. **IF** (Kondisi barang = Baik **AND** Kategori Barang = Dangerous Goods **AND** Alamat Tujuan = Jawa Sumatra Bali **AND** Logistik = POS) **THEN** Ya

C. *Pengujian Dengan Rapid Miner 9.0*

Pada penelitian ini peneliti menggunakan Rapid Miner 9.0 untuk menguji bagaimana proses klasifikasi dan pohon keputusan yang dihasilkan dari data pengiriman barang. berikut ini merupakan uraian langkah langkah bagaimana menentukan pohon keputusan menggunakan Rapid Miner.



Gambar 5. Desain Model Algoritma C4.5

Model algoritma C4.5 dalam gambar 5 diatas dengan proses *cross validation* digunakan untuk menguji *decision tree* dengan *apply model* dan *performance* dengan menggunakan *tools Rapid Miner*.



Gambar 6. Pohon Keputusan Prediksi Opportunity

Setelah bagian proses selesai maka output dapat dilihat di *views result* pada halaman *decision tree* menu *graph* untuk menampilkan pohon keputusan dan menu *description* pada halaman *decision tree* akan menampilkan *rules* yang dihasilkan seperti pada gambar 6 diatas.

Table View
 Plot View

accuracy: 93.00% +/- 10.05% (mikro: 93.00%)

	true Tidak	true Ya	class precision
pred. Tidak	7	3	70.00%
pred. Ya	4	86	95.56%
class recall	63.64%	96.63%	

Gambar 7. Output akurasi Algoritma C 4.5

Hasil tingkat akurasi bisa dilihat pada halaman *performance* seperti gambar 7. Dimana dari hasil pengujian menggunakan aplikasi RapidMiner didapat hasil jumlah *True Positive* (TP) adalah sebanyak 86 *record* diklasifikasikan sebagai keberhasilan pengiriman dan *False Negative* (FN) sebanyak 4 *record* diklasifikasikan kategori Gagal kirim tetapi ada kemungkinan berhasil. Berikutnya untuk jumlah *True Negative* (TN) sebanyak 7 *record* dikategorikan sebagai kegagalan pengiriman dan jumlah *False positive* (FP) dikategorikan sebagai Berhasil tetapi memiliki risiko kegagalan yang berjumlah 3 *record*.

Berdasarkan data diatas maka dapat kita hitung untuk mencari nilai *accuracy*, *sensitivity*, *specifity*, *ppv* dan *npv* sebagai berikut:

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{86}{86 + 4} = 0.95556$$

$$\text{Specificity} = \frac{TN}{TN + FP} = \frac{7}{7 + 3} = 0.7$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{86 + 7}{86 + 7 + 3 + 4} = 0.93$$

$$\text{PPV} = \frac{TP}{TP + FP} = \frac{86}{86 + 3} = 0.96629$$

$$\text{NPV} = \frac{TN}{TN + FN} = \frac{7}{7 + 4} = 0.63636$$

Dari perthitungan di atas nilai dari *accuracy*, *sensitivity*, *specifity*, *ppv* dan *npv* dapat dilihat pada tabel dibawah ini yang sudah di konversi kedalam bentuk persentase.

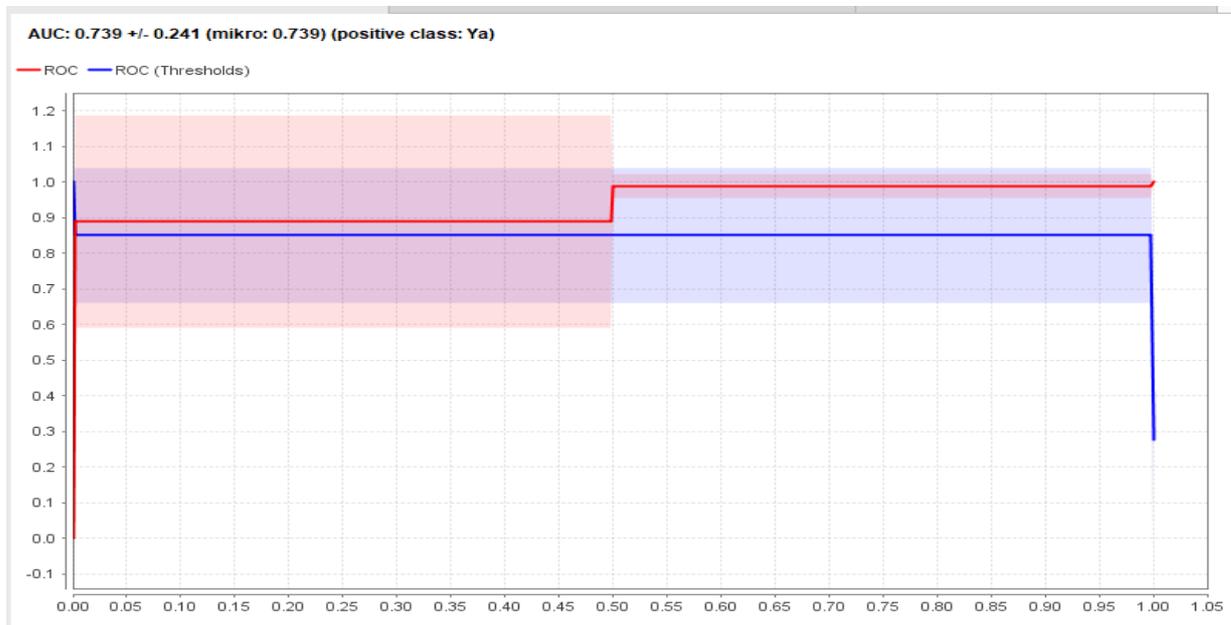
TABEL IX

HASIL PERHITUNGAN ACCURACY, SENSITIVITY, SPECIFITY, PPV DAN NPV

Hasil perhitungan Accuracy, Sensitivity, Specifity, PPV dan NPV dalam (%)	
<i>Accuracy</i>	93
<i>Sensitivity</i>	95,55
<i>Specifity</i>	70
PPV	96,63
NPV	63,64

Berdasarkan tabel IX menunjukkan bahwa, tingkat akurasi menggunakan metode algoritma C4.5 adalah sebesar 93% dengan data yang diuji sebanyak 100 data pengiriman. Hasil perhitungan divisualisasikan dengan kurva ROC. Perbandingan kedua *class* bisa dilihat

pada gambar 8 yang merupakan kurva ROC untuk algoritma C4.5. Kurva ROC mengekspresikan *confusion matrix*. Garis horizontal adalah false positives dan garis vertikal true positives. Menghasilkan nilai AUC (Area Under Curve) sebesar 0.739 dengan nilai akurasi klasifikasi Cukup (*Fair classification*).



Gambar 8. Hasil AUC Dalam Grafik ROC Algoritman C4.5

IV. KESIMPULAN

Ada klasifikasi keberhasilan pengiriman barang berdasarkan atribut logistik, alamat tujuan, kategori barang dan kondisi barang. Implementasi data mining dengan algoritma C4.5 dapat menghasilkan informasi berupa klasifikasi keberhasilan pengiriman barang dimana dari data training yang digunakan dengan jumlah 100 dapat dibangun sebuah decision tree yang menghasilkan 9 rules. Berdasarkan hasil pengujian dengan algoritma C4.5 menggunakan tools *Rapid Miner* yang diukur tingkat akurasi menggunakan pengujian confusion matrix dan kurva ROC didapat nilai akurasi sebesar 93 persen dan menghasilkan nilai AUC (*Area Under Curve*) sebesar 0.739 dengan nilai akurasi klasifikasi Cukup (*Fair classification*). Data mining dengan algoritma C4.5 dapat diimplementasikan untuk mengklasifikasi keberhasilan dan kegagalan pengiriman. Hal ini dapat menjadi rekomendasi bagian distribusi dalam memilih jasa logistik untuk pengiriman barang berdasarkan alamat pengiriman, dan kategori barang agar resiko kegagalan pengiriman bisa dikurangi.

REFERENSI

- [1] L. N. Rani, "Klasifikasi Nasabah Menggunakan Algoritma C4 . 5 Sebagai Dasar Pemberian Kredit," *INOVTEK Polbeng - Seri Inform.*, vol. 1, pp. 126–132, 2016.
- [2] A. E. Pramadhani and T. Setiadi, "Penerapan Data Mining untuk Klasifikasi Penyakit ISPA dengan Algoritma Desicion Tree," *J. Sarj. Tek. Inform. e-ISSN 2338-5197*, vol. 2, no. 1, pp. 831–839, 2014.
- [3] E. S. Palupi, S. M. Pahlevi, U. Bina, S. Informatika, P. Magister, and I. Komputer, "Inti nusa mandiri," *INTI Nusa Mandiri*, vol. 14, no. 2, pp. 133–138, 2020.
- [4] L. Silvia and J. Sundari, "Penerapan Algoritma Apriori Untuk Menentukan Impor Barang Pada PT . Pafa Mandiri Sakti," *Process. J. Ilm. Sist. Informasi, Teknol. Inf. dan Sist. Komput.*, vol. 15, no. 1, pp. 54–61, 2020.

- [5] A. Muzakir and R. A. Wulandari, "Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree," *Sci. J. Informatics*, vol. 3, no. 1, pp. 19–26, 2016.
- [6] S. Nurajizah, U. Bina, S. Informatika, R. Barat, and J. Barat, "Analisa Transaksi Penjualan Obat menggunakan Algoritma Apriori," *INOVTEK Polbeng - Seri Inform.*, vol. 4, no. 1, pp. 35–44, 2019.
- [7] B. Hermanto, A. Sn, and F. P. Putra, "Analisis Kinerja Decision Tree C4 . 5 dalam Prediksi Potensi Pelunasan Kredit Calon Debitur," *INOVTEK Polbeng - Seri Inform.*, vol. 2, no. 2, pp. 189–197, 2017.
- [8] T. Limbong and J. Simarmata, "Menentukan Matakuliah yang Efektif Belajar Daring (Belajar dan Ujian) dengan Metode Multi-Attribute Utility Theory (MAUT)," *J. RESTI*, vol. 4, no. 2, pp. 370–376, 2020.
- [9] P. P. Putra and A. S. Chan, "Pengembangan Aplikasi Perhitungan Prediksi Stock Motor Menggunakan Algoritma C 4 . 5 Sebagai Bagian dari Sistem Pengambilan Keputusan (Studi Kasus di Saudara Motor)," *INOVTEK Polbeng - Seri Inform.*, vol. 3, pp. 24–33, 2018.
- [10] H. Amalia and Evicienna, "Penentuan Proses Persalinan Ibu Melahirkan Menggunakan Algoritma c4.5," *Semin. Nas. Cendekiawan*, vol. 3, pp. 101–107, 2017.
- [11] M. Iqbal, W. Usino, and T. Triono, "Sistem Pendukung Keputusan Menentukan Hasil Budidaya Udang Vaname Dengan Metode Algoritma C4.5 (Pt Anugerah Sumber Laut Jaya)," *J. Tekno Insentif*, vol. 14, no. 1, pp. 28–39, 2020.
- [12] H. Widayu, S. Darma, N. Silalahi, and Mesran, "Data Mining Untuk Memprediksi Jenis Transaksi Nasabah Pada Koperasi Simpan Pinjam Dengan Algoritma C4.5," *J. Media Inform. Budidarma*, vol. 1, no. 2, pp. 32–37, 2017.
- [13] K. Umam, D. Puspitasari, and A. Nurhadi, "Penerapan Algoritma C4.5 Untuk Prediksi Loyalitas Nasabah PT Erdika Elit Jakarta," *J. Media Inform. Budidarma*, vol. 4, no. 1, p. 65, 2020.
- [14] N. Lutfiyana, "Penerapan Algoritma C4.5 Berbasis Particle Swarm Optmization Untuk Prediksi Hasil Layanan Kemudaha Donasi Zakat Dan Program," *J. Pilar Nusa Mandiri*, vol. 14, no. 1, pp. 103–110, 2018.
- [15] I. Romli and A. T. Zy, "Penentuan Jadwal Overtime Dengan Klasifikasi," *J. Sains Komput. Inform.*, vol. 4, no. 2, pp. 694–702, 2020.

UCAPAN TERIMA KASIH

Peneliti mengucapkan terima kasih kepada Allah SWT serta Tim Jurnal Inovtek Seri Informatika Polbeng yang telah meluangkan waktu untuk mereview dan memberikan masukan untuk jurnal guna menunjang penelitian ini dengan baik.