

## IMPLEMENTASI ALGORITME C4.5 UNTUK PREDIKSI KETEPATAN WAKTU LAYANAN YES AGEN GLC GEMILANG

Afif Bangkit Nur Rahmaan<sup>1\*</sup>, Gunawan Pria Utama<sup>2</sup>

<sup>1</sup>Teknologi Informasi, Fakultas Teknologi Informatika, Universitas Budi Luhur, Jakarta, Indonesia

Email: <sup>1</sup>afif2916@gmail.com, <sup>2</sup>gunawan.priautama@budiluhur.ac.id  
(\* : corresponding author)

**Abstrak-** Agen JNE GLC Gemilang adalah sebuah agen ekspedisi dari perusahaan JNE yang bergerak di bidang logistik yang beroperasi di daerah Tangerang Kota Sejak Tahun 2015. Terdapat beberapa layanan pengiriman yang ada di Agen ini. Salah satunya adalah layanan YES( Yakin Esok Sampai). Namun dalam operasional pengiriman layanan YES masalah yang sering terjadi adalah masih ada beberapa pengiriman yang mengalami kendala keterlambatan. Oleh karena itu perlu adanya sistem yang dapat memprediksi ketepatan pengiriman YES yang dapat dihitung berdasarkan faktor-faktor yang dapat dipertimbangkan seperti waktu, jenis barang pengiriman, tujuan pengiriman dan lain lain. Data mining sendiri adalah proses mencari pola atau informasi menarik dalam data yang terpilih menggunakan metode tertentu. Metode yang digunakan untuk memprediksi ketepatan pengiriman layanan YES tersebut adalah algoritme C4.5. Dari data latih yang didapatkan menghasilkan sebuah *decision tree*. *Decision tree* adalah kumpulan rules yang menentukan apakah nomor resi yang menggunakan service YES tepat waktu atau tidak. Setelah metode klasifikasi data mining Algoritme C4.5 sudah diimplementasikan di aplikasi web, pengujian data diuji terhadap *decision tree* untuk melihat keakuratan *decision tree* yang sudah dibentuk dalam aplikasi tersebut. Untuk melihat keakuratan membutuhkan perhitungan *confusion matrix* agar dapat melihat keakuratan *decision tree* yang sudah terbentuk oleh aplikasi. Dengan adanya penelitian ini, menggunakan data mining dapat membantu meningkatkan ketepatan dari layanan YES, model pembelajaran mesin dapat mengenali pola yang mengarah pada pengiriman yang tepat waktu dan memprediksi kiriman yang kemungkinan gagal. Dataset yang digunakan bersumber dari Agen JNE GLC Gemilang sebanyak 265 data, hasil analisis berdasarkan dataset yang digunakan hasil pengujian memiliki tingkat akurasi sebesar 74.68%, *recall* sebesar 98.18% dan presisi sebesar 73.97% dan diketahui rata-rata keterlambatan dialami untuk resi dengan destinasi ke luar pulau dan paket dalam pulau yang tidak di asuransi.

**Kata Kunci:** Algoritme C4.5, layanan YES, JNE, data mining, prediksi

## IMPLEMENTATION OF C4.5 ALGORITHM FOR PREDICTING SERVICE TIMELINESS OF YES AT AGENT GLC GEMILANG

**Abstract-** Agent JNE GLC Gemilang is a logistics company's expedition agent from JNE, operating in logistics field located at Tangerang City area since 2015. There are several delivery services available at this agent, such as Regular delivery, JNE trucking service, international shipment and the YES (Yakin Esok Sampai) service. However, in the operational aspect of the YES delivery service, there are still some deliveries that experience delays. Therefore, there is a need for a system that can predict the accuracy of YES deliveries based on various factors such as time, type of goods, delivery destination and others. Data mining is the process of finding interesting patterns of information in selected data using specific methods. The method used to predict the accuracy of the YES delivery service is the C4.5 algorithm. From obtained training data, a decision tree will be generated. The decision tree consists of a set of rules that will determine whether a shipment with a YES service will be on time or not. Once the data mining classification method, the C4.5 algorithm is implemented in the web application, data testing will be conducted against the decision tree to assess the accuracy of the decision tree formed in the application. To evaluate accuracy, a confusion matrix will be calculated to examine the accuracy of the decision tree formed by the application. Through this research using data mining, it is expected to improve the accuracy of the YES service. The machine learning model will be able to recognize patterns that lead to on-time deliveries and predict potential failures delivery. The dataset used for this study consists of 265 data points sourced from JNE GLC Gemilang Agent. Based on the analysis of the dataset used, the testing result indicates an accuracy rate of 74.68 recall rate of 98.18% and precision rate of 73.97%. It is found that the average delay occurs for shipments with destinations outside the island and destinations inside the island where packages are uninsured..

**Keywords:** algorithm C4.5, YES service, JNE, data mining, prediction

## 1. PENDAHULUAN

Agen GLC Gemilang merupakan salah satu agen resmi dari perusahaan logistik PT Tiki Jalur Nugraha Ekakurir yang memberikan layanan pengiriman paket dengan berbagai jenis layanan, salah satunya adalah layanan YES (Yakin Esok Sampai). Masalah yang sering didapatkan oleh agen GLC Gemilang adalah masih adanya keterlambatan untuk kiriman dengan layanan YES. Menggunakan data mining diharapkan dapat membantu menentukan strategi pengiriman paket layanan YES agar tidak terjadi keterlambatan.

Data mining merupakan suatu metode untuk menemukan pengetahuan dalam suatu tumpukan data yang cukup besar dengan proses menggali dan menganalisis sejumlah data yang sangat besar[1]. Informasi yang dihasilkan diperoleh dengan cara mengekstrasi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data[2].

*Preprocessing* adalah tahap untuk melakukan sebuah proses awal dalam pengolahan data[8]. Didalam tahap *preprocessing* terdapat tahap data *Transformation*, data *Selection* dan juga Pengecekan *Outlier*. Data *Transformation* adalah tahapan dimana data ditransformasikan dan dikonsolidasikan ke dalam bentuk yang sesuai untuk *mining*[11]. Dalam proses nya dilakukan transformasi data agar dapat di proses aplikasi mining yang sudah di buat.

Deteksi/pengecekan *outlier* sebenarnya menemukan *anomaly* atau mendeteksi data yang janggal dari sebuah penelitian untuk mendeteksi kasus kasus abnormal pada *database*[12].

Metode yang digunakan dalam penelitian adalah algoritme C4.5. Algoritme C4.5 merupakan algoritme yang digunakan untuk melakukan klasifikasi data dengan menggunakan teknik keputusan.[3]. Algoritme C4.5 menggunakan *Gain Ratio* agar tidak bias dalam penentuan atribut pemilah terbaik. Algoritme C4.5 merupakan algoritme yang digunakan untuk melakukan proses klasifikasi data dengan menggunakan struktur pohon atau struktur berhirarki. Konsep pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan aturan keputusan.[4]. Langkah langkah yang dilakukan untuk melakukan algoritme C4.5 adalah perhitungan jumlah data, perhitungan entropy, perhitungan gain, perhitungan gain ratio dan membuat pohon keputusan berdasarkan gain tertinggi lalu ulangi proses setiap cabang sampai semua kasus pada cabang memiliki kelas[5]. Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon[10]. Yang dimulai dari pemilihan variable atau atribut dengan *gain* tertinggi sebagai akar, membangun cabang untuk setiap nilai, membagi kasus dalam cabang kemudian melakukan pengulangan proses untuk setiap cabang sampai seluruh kasus pada cabang memiliki kelas yang sama. Setelah dilakukan pembentukan pohon bentuk pohon diubah menjadi rule *IF-THEN*. *IF* adalah kondisi dan *THEN* adalah kesimpulan aturan ini terdiri dari dua bagian *IF* yang mengandung satu atau lebih kondisi dan *THEN* yang mengandung keputusan[9].

Dalam penelitian ini juga dilakukan perhitungan akurasi yang sudah di buat oleh algoritme C4.5 menggunakan *Confusion matrix*. *Confusion matrix* adalah sebuah metode yang kegunaannya untuk melakukan perhitungan akurasi pada konsep dalam data mining[6]. Tujuan *confusion matrix* menganalisis kualitas kinerja model klasifikasi dalam mengenali variabel dari seluruh kelas.[7].

Tujuan dari penelitian ini adalah membantu Agen GLC Gemilang memprediksi tepat atau tidaknya estimasi sampai dari paket yang menggunakan layanan YES menggunakan Algoritme C4.5.

Secara keseluruhan implementasi algoritme C4.5 dalam prediksi ketepatan waktu layanan YES akan membawa manfaat bagi agen GLC Gemilang seperti pengambilan keputusan untuk memperbaiki alur kerja / alur kiriman paket dengan layanan YES, meningkatkan kepuasan pada pelanggan dan juga meningkatkan daya saing terhadap perusahaan kompetitor lain.

Berdasarkan penelitian yang dilakukan[2]. Dilakukan penelitian tentang kesuksesan pengiriman barang dari beberapa perusahaan logistik, data yang didapatkan oleh peneliti adalah data yang berasal dari sebuah perusahaan *e-commerce* namun tidak disebutkan layanan yang di pergunakan dan label yang digunakan adalah sukses dan tidak suksesnya pengiriman. Berdasarkan penelitian[2] hasil yang didapatkan tingkat akurasi sebesar 93% dan menghasilkan 9 *rules*. Dalam penelitian ini data yang dimiliki lebih di perkecil cakupannya menjadi salah satu sebuah perusahaan logistik, dan layanan tertentu yaitu layanan YES.

## 2. METODE PENELITIAN

### 2.1 Data Penelitian

Data penelitian yang didapatkan dari Agen GLC Gemilang, Dalam melakukan penelitian ini dilakukan pengumpulan data Resi dengan Layanan YES di agen GLC Gemilang pada tanggal 1-7 April 2023. Sehingga

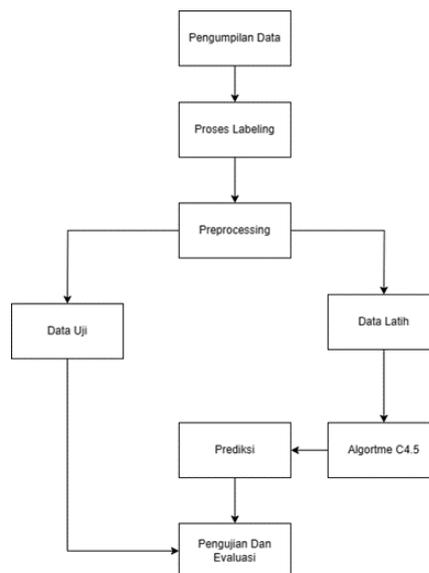
didapatkan jumlah data sebanyak 265 data. Pada tabel berikut merupakan contoh data yang didapatkan dari agen GLC Gemilang:

**Tabel 1.** Contoh Data Resi

AWB	Connote Date	Insurance Flag	Service	Destination	Qty	Handling	type	Pod date	POD receiving	COD flag	COD Amount
541990019606923	1/4/2023 10:07	Y	YES19	Jabodetabek	1	General	Dokumen	1/4/2023 22:21	TANIA	No	0
541990019611223	1/4/2023 10:41	N	YES19	Jabodetabek	1	General	Paket	2/4/2023 16:20	Anwar	No	0
541990019612123	1/4/2023 10:47	N	YES19	Jabodetabek	1	General	Dokumen	1/4/2023 23:25	Edi	No	0

## 2.2 Penerapan Metode

Dalam membangun aplikasi data mining yang diimplementasikan dalam penelitian ini, dilakukan beberapa langkah, Langkah-langkah tersebut mewakili setiap proses dan desain penelitian ini dari awal hingga akhir aplikasi bekerja. Tahapan yang dilakukan dapat dilihat pada gambar 1 berikut :



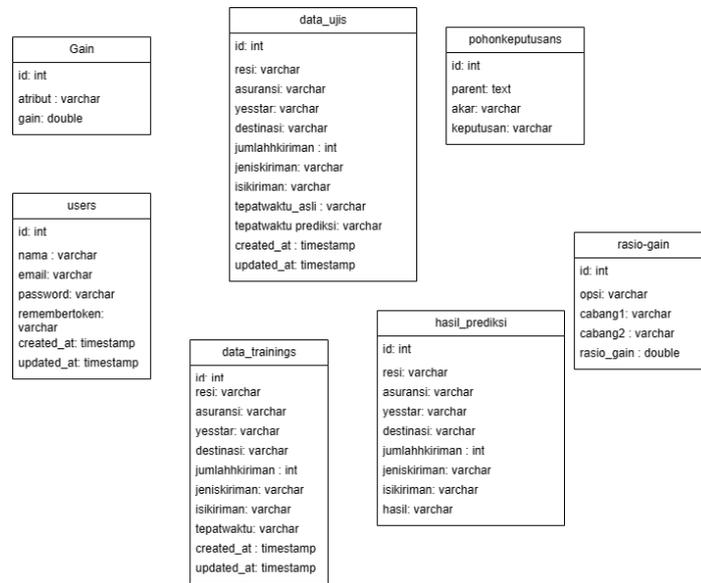
**Gambar 1.** Penerapan Metode Keseluruhan

Berdasarkan gambar 1, pengumpulan data dilakukan melalui staff Agen GLC Gemilang lalu data tersebut masuk ke tahap *preprocessing*. Didalam tahap *Preprocessing* terdapat tahap data *Transformation*, data *Selection* dan juga pengecekan *Outlier*.

Setelah data sudah di *preprocessing* dilakukan pembagian menjadi data latih dan juga data uji. Kemudian data latih dan data uji dimasukkan ke dalam basis data yang nantinya data latih akan dilakukan pemodelan menggunakan data latih yang tersedia. Untuk data uji sendiri dipersiapkan untuk pengujian akurasi pohon keputusan yang sudah dibuat oleh algoritme C4.5, pengujian tingkat akurasi, dan juga prediksi.

## 2.3 Rancangan Basis Data

Dalam proses pembuatan aplikasi ini, dibutuhkan basis data yang berisikan semua data untuk menjalankan aplikasi berikut merupakan rancangan *Logical Record structure* basis data yang akan digunakan :



Gambar 2. LRS basis data

Berdasarkan gambar 2 diatas basis data yang digunakan memiliki 7 Tabel,

#### 2.4 Algoritme C4.5

Pada tahapan ini dilakukan perhitungan *Entropy*, *Gain*, *Rasio Gain* dalam proses ini memiliki beberapa proses yang dilalui untuk membuat sebuah pohon keputusan antara lain: menghitung jumlah data, menghitung entropy, menghitung gain, menghitung rasion gain, memilih data dengan gain tertinggi dan membuat pohon keputusan dan *Rule*.

a. Perhitungan *Entropy*

Rumus yang digunakan untuk mendapat nilai *entropy* adalah :

$$Entropy (S) = \sum_{i=1}^n -p_i * \log_2 p_i \tag{1}$$

Keterangan :

- S : Himpunan kasus
- A : Fitur
- N : jumlah partisi S
- pi : Proporsi dari Si terhadap S

b. Perhitungan *Gain*

Rumus yang digunakan untuk mendapatkan nilai *gain* adalah :

$$Gain (S,A) = \sum_{i=1}^n \frac{S_i}{S} * Entropy(S_i) \tag{2}$$

Keterangan :

- S : Himpunan kasus
- A : Fitur
- n : Jumlah Partisi Atribut A
- Si : Jumlah kasus pada partisi ke-i
- S : Jumlah kasus dalam S

c. Perhitungan *Split Info*

Rumus yang digunakan untuk mendapatkan nilai *Split Info* adalah :

$$SplitInfo (S,A) = - \sum_{i=1}^n \frac{S_i}{S} * \log_2 \frac{S_i}{S} \tag{3}$$

Keterangan :

- S : Himpunan kasus
- A : Fitur

- n : Jumlah Partisi Atribut A
- Si : Jumlah kasus pada partisi ke-i
- S : Jumlah kasus dalam S

d. Perhitungan *Gain Ratio*

Rumus yang digunakan untuk mendapatkan nilai *gain ratio* adalah :

$$\text{GainRatio}(S,A) = \frac{\text{Gain}(S,A)}{\text{SplitInfo}(S,A)} \quad (4)$$

Keterangan :

- S : Himpunan kasus
- A : Fitur
- n : Jumlah Partisi Atribut A
- Si : Jumlah kasus pada partisi ke-i
- S : Jumlah Kasus dalam S

e. Pembentukan pohon

Pada proses ini pembentukan pohon dilakukan berdasarkan *gain* tertinggi. Data yang memiliki *gain* tertinggi akan dijadikan *node* dan proses ini akan dilakukan secara rekursif atau berulang ulang hingga data dan kondisi yang dimiliki sudah di hitung seluruhnya.

## 2.5 Confusion Matrix

Untuk pengujian akurasi algoritme C4.5 digunakan metode *Confusion matrix* untuk mengetahui akurasi, *recall* dan juga presisi. Berikut merupakan tabel *confusion matrix* :

**Tabel 2.** Confusion Matrix

		Tepatwakt tu	Tidak Tepatwakt tu	Jumlah
Kelas	Tepatwakt tu	TP	FN	P
	Tidaktep atwakt u	FP	TN	N
Aktual	Jumlah	P	N	P+N

Berdasarkan tabel 2 diatas nilai TP, FN, FP, dan TN akan digunakan untuk Perhitungan nilai akurasi, presisi dan recall. Perhitungan Akurasi, presisi dan recall dapat dinyatakan menggunakan persamaan berikut :

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (5)$$

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \quad (6)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (7)$$

Keterangan :

- TP (True Positive) : Jumlah data positif yang diprediksi sebagai data positif
- TN (True Negative) : Jumlah data negatif yang diprediksi sebagai data negatif
- FP (False Positive) : Jumlah data negatif yang diprediksi sebagai data positif
- FN (False Negative) : Jumlah data positif yang diprediksi sebagai data negatif
- P : Jumlah Sampel Positif
- N : Jumlah Sampel Negatif

### 3. HASIL DAN PEMBAHASAN

Pada bagian ini berisi analisis, hasil implementasi ataupun pengujian serta pembahasan dari topik penelitian, yang bisa dibuat terlebih dahulu metodologi penelitian. Bagian ini juga merepresentasikan penjelasan yang berupa penjelasan, gambar, tabel dan lainnya.

#### 3.1 Implementasi Metode

a. Tahap Labeling

Tahap labeling adalah sebuah tahap pemberian label atau kelas pada data yang kita miliki, Dalam penelitian ini data yang dimiliki akan diberikan label “Ya” untuk yang tepat waktu dan “Tidak” untuk yang tidak tepat waktu.

b. Tahap Preprocessing

Pada tahap ini terdiri dari beberapa tahap yaitu :

1. Data *Transformation*

Pada proses Data *Transformation* dilakukan transformasi data yang mengandung “Yes” dan “no” di transformasi menjadi “Ya” dan “Tidak”, data yang mengandung “YES\*” dan “YES18” menjadi “Ya” dan “Tidak” data yang berisikan “Y” menjadi “Ya” dan data yang berisikan “N” menjadi “Tidak” dan untuk atribut jumlah kiriman yang memiliki nilai 1 menjadi “Satu” dan yang lebih dari 1 menjadi “Lebihdarisatu”.

2. Data *Selection*

Pada proses data *selection* dilakukan perhitungan *information gain* menggunakan aplikasi WEKA pada gambar 3 berikut merupakan hasil yang ditunjukkan oleh aplikasi WEKA:

```
=== Attribute Selection on all input data ===

Search Method:
  Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 12 tepatwaktu):
  Information Gain Ranking Filter

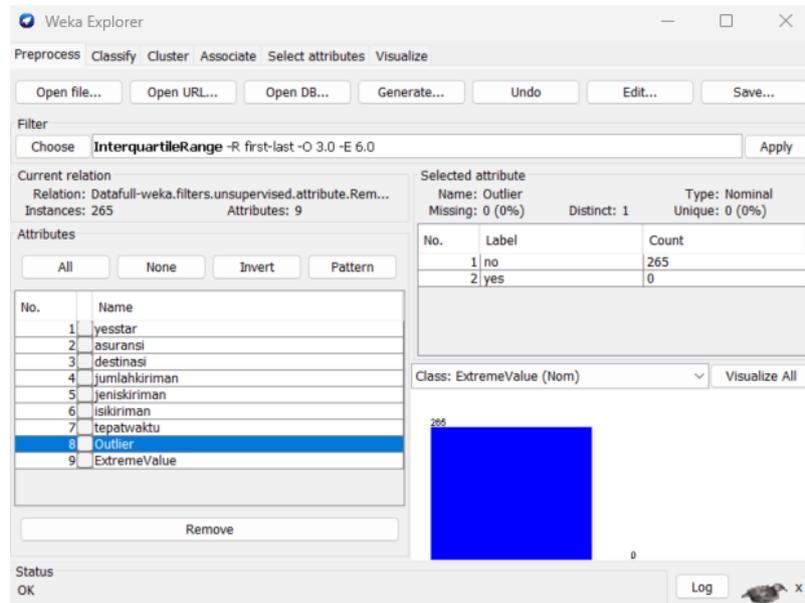
Ranked attributes:
0.70677  9 nama penerima
0.69167  8 tanggal terima
0.59441  1 tanggal input
0.0624   4 destinasi
0.0257   5 jumlahkiriman
0.01759  2 yesstar
0.00971  3 asuransi
0.00804  7 isikiriman
0.00109  6 jeniskiriman
0        10 COD
0        11 hargacod
```

**Gambar 3.** Hasil *Information Gain*

Setelah dilakukan seleksi data menggunakan *information gain* menggunakan aplikasi WEKA, dilakukan Kembali seleksi berdasarkan validasi lapangan sehingga atribut yang terpilih adalah atribut yesstar, asuransi, isikiriman, jeniskiriman, destinasi dan jumlah kiriman.

3. Pengecekan *Outlier*

Setelah dilakukan tahap seleksi data, langkah berikutnya adalah pengecekan *outlier*. Pada langkah ini pengecekan dilakukan menggunakan aplikasi WEKA. Berikut merupakan hasil pengecekan *outlier* dari aplikasi WEKA :



Gambar 3. Hasil Pengecekan Outlier

Berdasarkan gambar 3 diatas aplikasi WEKA menunjukkan tidak ada data outlier jadi tidak perlu dilakukan pembersihan data pada dataset.

c. Pembagian Data

Setelah tahap *preprocessing* data yang sudah ada di lakukan pembagian menggunakan ratio 7:3 atau 70% untuk data training dan 30% untuk data uji. Dalam penelitian ini tahap pembagian dilakukan terhadap 265 data resi. Dengan rasio data 7:3 makadapat diketahui jumlah data yang menjadi data latih adalah 186 dan data uji sebanyak 79.

d. Tahap Algoritme C4.5

Setelah tahap pembagian data selesai. Tahapan berikutnya adalah tahap memproses data latih menggunakan algoritme C4.5. Langkah pertama yang dilakukan untuk menghitung menggunakan Algoritme C4.5 adalah menghitung *Entropy* pada tabel 2 berikut merupakan hasil perhitungan *entropy* menggunakan persamaan (1) :

Tabel 2. Hasil Hitung *entropy*

Atribut		Jumlah	Tepatwaktu	Tidak	Eentropy
Asuransi	Ya	68	6	62	0.431
	Tidak	118	96	22	0.694
Yesstar	Ya	3	1	2	0.918
	Tidak	183	157	26	0.59
Destinasi	Jabodetabek	123	111	12	0.461
	LuarPulau	123	111	12	0.461
	DalamPulau	35	25	10	0.863
JumlahKiriman	Satu	137	113	24	0.669
	Lebihdarisatu	49	45	4	0.408
Jeniskiriman	General	139	117	22	0.63
	High Value	15	13	2	0.567
	Dangerous Goods	35	23	10	0.863
Isi Kiriman	Dokumen	37	29	8	0.753
	Paket	149	129	20	0.569

Berdasarkan pada tabel 2 diatas nilai-nilai *entropy* yang sudah di hitung akan digunakan untuk menghitung *gain* pada tabel 3. Berikut merupakan hasil perhitungan *gain* menggunakan persamaan (2) :

Tabel 3. Hasil Hitung *gain*

Atribut		Tepatwaktu = Ya	Tepatwaktu = Tidak	Gain
Asuransi	Ya	6	62	0.013
	Tidak	96	22	

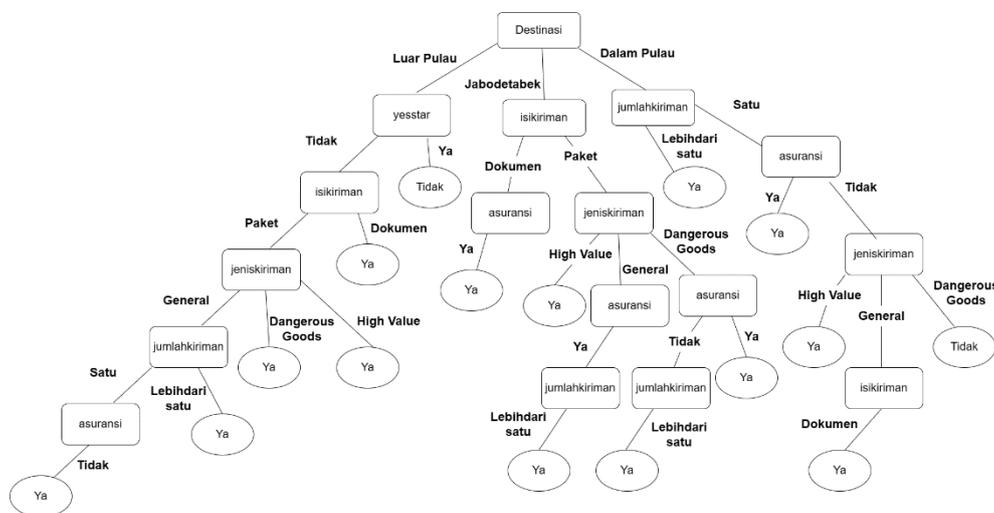
Yesstar	Ya	1	2	0.016
	Tidak	157	26	
Destinasi	Jabodetabek	111	12	0.031
	LuarPulau	111	12	
	DalamPulau	25	10	
JumlahKiriman	Satu	113	24	0.011
	Lebihdarisatu	45	4	
Jeniskiriman	General	117	22	0.001
	High Value	13	2	
	Dangerous Goods	23	10	
Isi Kiriman	Dokumen	29	8	0.005
	Paket	129	20	

Berdasarkan tabel 3 diatas nilai gain tertinggi di peroleh oleh atribut destinasi, dikarenakan destinasi memiliki 3 nilai atribut makaperlu dilakukan perhitungan *split info* untuk mendapatkan nilai *rasio gain* menggunakan persamaan (3) dan (4). Pada tabel 4. Berikut merupakan hasil perhitungannya :

**Tabel 4.**Hasil Hitung *Gain Ratio*

Opsi	JumlahCabang1	Jumlahcabang2	Split	Rasio
Opsi1	LuarPulau/Dalampulau = 63	Jabodetabek = 123	0.942	0.034
Opsi2	DalamPulau/Jabodetabek = 158	Luarpulau = 28	0.611	0.051
Opsi3	Jabodetabek/LuarPulau = 151	DalamPulau = 35	0.698	0.044

Berdasarkan dari tabel 4, maka langkah berikutnya adalah membentuk pohon dari perhitungan-perhitungan yang sudah dilakukan secara rekursif atau berulang ulang berdasarkan kondisi yang di dapatkan berdasarkan perhitungan *gain* tertinggi. Pada gambar 4. Berikut merupakan pohon keputusan yang sudah di hasilkan berdasarkan perhitungan menggunakan data set yang sudah di siapkan :



**Gambar 4.** Pohon Keputusan

Berdasarkan Pohon Keputusan Pada gambar 4. Dapat dilihat untuk paket yang data/paket yang memiliki keterlambatan ada di resi/data dengan destinasi ke dalam pulau dengan kondisi tidak di asuransi dan kiriman ke luarpulau dengan kondisi layanan YES\*.

### 3.2 Pengujian Akurasi Prediksi

Pada Penelitian ini dilakukan pengujian dari sisi akurasi, presisi, dan recall pada implementasi algoritme C4.5 dalam memprediksi ketepatan waktu layanan YES. Sampel prediksi yang dapat dilihat pada tabel berikut :

**Tabel 4.**Sampel Data hasilPrediksi

NomorResi	Asuran	YES	Destinasi	JumlahKiri	Jeniskiirim	Isikirim	Tepatwak	PrediksiTepatwa
	si	*		man	an	an	tu	ktu

541990020547 923	Tidak	Tidak	LuarPulau	Lebihdarisat u	General	Paket	Tidak	Ya
541990020551 223	Tidak	Tidak	DalamPul au	Satu	General	Dokume n	Ya	Ya

Pada tabel 4 tersebut sampel data hasil prediksi, kolom prediksi tepat waktu adalah hasil dari prediksi menggunakan rule yang sudah di buat berdasarkan data latih yang sudah digunakan untuk membuat pohon keputusan. Keseluruhan data uji sebanyak 79 data kemudian dapat di representasikan di *Confusion Matrix*, yang dapat dilihat pada tabel 5 berikut :

**Tabel 5.**Sampel Data hasil Prediksi

		Positif	Negatif
Nilai Prediksi	Positif	54	5
	Negatif	19	1

BerdasarkanTabel 5 dapat diketahui hasil dari masing-masing hasil perbandingan antara nilai tepat waktu asli dan tepat waktu prediksi, setelah diketahui langkah berikutnya adalah menghitung nilai Recall, Akurasi dan presisi menggunakan persamaan (5), (6) dan (7) yang dapatdilihat pada tabel berikutini :

**Tabel 6.**Sampel Data hasilPrediksi

		Prediksi	
Akurasi	$(54+5)/79$	74.68%	
Presisi	$54/(54+19)$	73.97%	
Recall	$54/(54+1)$	98.18%	

Berdasarkan tabel 6, hasil pengujian yang dapat diketahui dan disimpulkan bahwa algoritme C4.5 menunjukkan nilai pengujian dengana kurasi 74.68%, Presisi sebesar 73.97% dan recall sebesar 98.18%.

### 3.3 Pengujian Prediksi

Pengujian ini menggunakan resi baru dengan patokan pohonke putusan untuk mengetahui prediksi resi tersebut. Pada tabel 7 berikut merupakan tabel sampel pengujian :

**Tabel 7.**Sampel Data hasilPrediksi

NomorResi	Asuransi	YES *	Destinasi	JumlahKiriman	Jeniskiiriman	Isikiriman
541990041318223	Tidak	Ya	LuarPulau	Satu	General	Paket

Berdasarkan data sampel yang sudahdigunakan, aplikasi yang sudah di buatkan menyocokkan atribut data baru berdasarkan pohon keputusan yang sudah di buat, setelah di cocokkan aplikasi akan memberikan prediksi waktu atau tidaknya. Pada tabel 8 berikut merupakan tabel sampel yang sudah di prediksi :

**Tabel 8.**Sampel Data hasilPrediksi

NomorResi	Asuransi	YES *	Destinasi	JumlahKiriman	Jeniskiiriman	Isikiriman	Prediksi
541990041318223	Tidak	Ya	LuarPulau	Satu	General	Paket	Tidak

Berdasarkan tabel 8 diatasprediksi data sampel yang digunakan di prediksi “Tidak” tepat waktu untuk *rule* yang terpilih untuk memprediksi data sampel adalah *IF*(destinasi=Luar Pulau) *AND* (yesstar='Ya') *THEN* tepatwaktu = Tidak.

## 4. KESIMPULAN

Berdasarkan hasil pengujian dan evaluasi aplikasi yang sudah dibuat maka dataset dan algoritme yang diusulkan. Berdasarkan pohon keputusan menggunakan 186 data latih dan 79 data uji rata-rata keterlambatan ada di resi dengan destinasi ke dalam pulau dengan kondisi tidak di asuransi dan kiriman ke luarpulau dengan kondisi layanan YES\*. Oleh karena itu untuk kiriman dengan kondisi yang telah disebutkan Agen GLC Gemilang harus mempercepat jam serah terima kepada kurir gudang agar dapat mengurangi keterlambatan pengiriman YES.

Menggunakan Teknik *Confusion Matrix* untuk mengevaluasi algoritme C4.5 dalam penelitian ini memiliki tingkat akurasisesbesar 74.68%, *Recall* sebesar 98.18% dan presisi sebesar 73.97%.

## UCAPAN TERIMA KASIH

Ucapan Terima kasih diucapkan sebesar besarnya kepada Agen GLC Gemilang yang sudah memberikan dan mempercayakan data pengiriman untuk bahan penelitian penulis agar dapat menyelesaikan penelitian ini.

## DAFTAR PUSTAKA

- [1] R. W. Nasution, I. O. Kirana, I. Gunawan, and I. P. Sari, "Penerapan Data Mining Untuk Pengelompokan Minat Konsumen Terhadap Pengguna Jasa Pengiriman Pada PT . Jalur Nugraha Ekakurir ( JNE ) Pematangsiantar," *Resolusi*, vol. 1, no. 4, pp. 274–281, 2021.
- [2] G. Taufik and D. Jatmika, "Penerapan Algoritma C45 Untuk Klasifikasi Keberhasilan Pengiriman Barang," *INOVTEK Polbeng - Seri Inform.*, 2021, doi: 10.35314/isi.v6i1.1446.
- [3] C. Rizky, A. Nugroho1, and T. Kristiana2, "Penerapan Algoritma C4.5 Untuk Kepuasan Pelanggan Toko Online Parfume Chantik," *J. Algoritm.*, vol. 3, no. 1, pp. 10–21, 2022.
- [4] R. Haqmanullah Pambudi, B. Darma Setiawan, and Indriati, "Penerapan Algoritma C4.5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 7, pp. 2637–2643, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [5] M. A. Ramadhan and S. Dwiasnati, "Klasifikasi Jasa Pengiriman JNE Service Menggunakan Algoritma Decision Tree," *J. Ilm. Fak. Komput.*, vol. 8, no. 2, p. 1, 2019.
- [6] V. S. Ginting, K. Kusriani, and E. Taufiq, "Implementasi Algoritma C4.5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 1, pp. 36–44, 2020, doi: 10.35585/inspir.v10i1.2535.
- [7] S. Abdurrohman and A. Wibowo, "Klasifikasi Data Mining Untuk Memprediksi Status Penerimaan Di Perguruan Tinggi Negeri Bagi Lulusan Bimbel NF Dengan Algoritme Naive Bayes," *Semin. Nas. Mhs. Fak. Teknol. Inf.*, vol. 1, no. 1, pp. 84–92, 2022, [Online]. Available: <http://senafti.budiluhur.ac.id/index.php/senafti/article/view/111>.
- [8] F. Alghifari and D. Juardi, "PENERAPAN DATA MINING PADA PENJUALAN MAKANAN DAN MINUMAN MENGGUNAKAN METODE ALGORITMA NAIVE BAYES," *J. Ilm. Inform.*, 2021, doi: 10.33884/jif.v9i02.3755.
- [9] Uminingsih, I. Suraya, and I. Nugroho, "Implementasi Algoritma Decision Tree C4.5 Untuk Menentukan Kelas Calon Siswa Di Lembaga Kursus Bahasa Inggris Berbasis Web," *J. Teknol.*, vol. 12, no. 1, pp. 35–44, 2019.
- [10] M. F. Arifin and D. Fitriana, "Penerapan Algoritma Klasifikasi C4.5 Dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : PT Atria Artha Persada," *InComTech*, 2018.
- [11] G. Sani *et al.*, "PENERAPAN DATA TRANSFORMATION PADA DATABASE SISTEM INFORMASI," pp. 1–5, 2019.
- [12] J. Nasional, S. Informasi, and S. Artikel, "Optimasi Pengelompokan Data Pada Metode K-Means dengan Analisis Outlier," vol. 02, pp. 88–95, 2019.