

Klasifikasi Jasa Pengiriman JNE Service Menggunakan Algoritma Decision Tree

Muhammad Akbar Ramadhan¹, Saruni Dwiasnati²

*Jurusan Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana²
Jl. Raya Meruya Selatan, Kembangan, Jakarta 11650*

¹141515110117@student.mercubuana.ac.id, ²saruni.dwiasnati@mercubuana.ac.id

Abstract

A company engaged in the field of freight forwarding services, or better known as expedition services. Is one area that is growing rapidly with the need for freight forwarding services from national to international scale. With this research, the author is aiming to implement a number of existing algorithms in data mining, but the author focuses more on the C4.5 algorithm which functions to be able to produce complex decisions to be simpler such as a data for customers to make customers more secure at all times. great so as not to switch or move services. The results showed that the tree produced from software created by the decision tree method with the C4.5 algorithm has a 96.28% percentage between the truth tree to 100% that can be obtained from a prediction. A large percentage of the truth tree is strongly influenced by the training data used to build a decision tree model.

Keywords: *Delivery, Data Mining, Decision Tree, C4.5 Algorithm*

Abstrak

Sebuah perusahaan yang bergerak di bidang jasa pengiriman barang ini atau lebih dikenal seperti jasa ekspedisi. Merupakan salah satu bidang yang berkembang pesat dengan kebutuhan pelayanan jasa pengiriman barang dari skala nasional sampai internasional. Akan tetapi dari pengiriman titipan yang telah dicatat hanya disimpan sebagai laporan pencatatan saja. Dengan masalah yang ada maka Penelitian saat ini yang penulis lakukan bertujuan untuk mengimplementasikan salah satu algoritma yang ada pada data mining, namun penulis lebih memfokuskan pada algoritma C4.5 yang berfungsi untuk mampu menghasilkan keputusan kompleks menjadi lebih sederhana seperti sebuah data untuk pelanggan agar lebih mempertahankan nasabah dengan golongan skala besar agar tidak berpaling atau pindah ke layanan jasa ekspedisi lain. Hasil penelitian menunjukkan bahwa pohon yang dihasilkan dari perangkat lunak yang dibuat dengan metode decision tree dengan algoritma C4.5 memiliki persentase 96.28% antara pohon kebenaran sampai 100% yang dapat didapatkan dari sebuah prediksi. Persentase besar pohon kebenaran sangat dipengaruhi oleh data training yang digunakan untuk membangun suatu model pohon keputusan. Pengolahan data yang dilakukan dalam penelitian ini menggunakan bantuan tools Rapid Miner untuk mendapatkan model, hasil dari pohon keputusan dan mendapatkan nilai accuracynya.

Kata Kunci: *Pengiriman, Data Mining, Decision Tree, Algoritma C4.5*

1. PENDAHULUAN

Data Mining merupakan sebuah database yang salah satu bidang yang sedang berkembang pesat karena besarnya kebutuhan akan nilai tambah dari database skala besar. Pada perusahaan yang bergerak dalam bidang ekspedisi pelayanan jasa pengiriman barang dari skala nasional sampai internasional. PT Jalur Nugrahaeka Kurir yaitu sebuah perusahaan yang bergerak dalam bidang jasa pengiriman ekspedisi paket barang nasional sampai internasional dalam melayani dalam skala kecil atau skala besar. Salah satu jasa pengiriman paket terbesar dalam ekspedisi paket pengiriman yang sudah hampir tersebar di seluruh Indonesia [1].

Penerapan permasalahan tersebut dapat di atasi melalui dengan dikembangkannya konsep-konsep data mining yang lebih aktual. Penelitian ini sebelumnya dikembangkannya melalu konsep data mining yaitu metode Klasifikasi C4.5 menggunakan Algoritma Decision Tree untuk menganalisa pola data dari hasil database tersebut.

Kemajuan teknologi dalam sistem pengiriman barang menghitung untuk menentukan transaksi dalam frekuensi yang cukup sering dalam proses transaksi yang sedang terjadi. Proses transaksi pengiriman barang tentunya harus menguntungkan kedua belah pihak. Tingkatnya dapat dipengaruhi oleh kepuasan transaksi dalam pengiriman yang cukup banyak dalam intensitas pengiriman yang tinggi. Salah satunya, menggunakan kegiatan promosi sehingga menarik minat pengiriman yang akan menarik pelanggan menjadi lebih aktif karena disediakannya tawaran atau promosi dalam pengiriman barang yang dilakukan. Misalnya berupa pemberian discount atau point disetiap transaksinya dengan ketentuan yang berlaku. Promosi yang dilakukan bertujuan untuk menambah dan memperluas area pemasaran dalam kegiatan pengiriman.

Berdasarkan permasalahan tersebut, maka dalam penelitian yang ini penulis akan menggunakan metode *decision tree* algoritma C4.5. Setelah diolah dengan data mining diharapkan dapat menemukan rencana strategis bagi perusahaan dapat menentukan kebijakan strategi untuk menentukan tawaran promo yang berguna.

2. LANDASAN TEORI DAN METODE

Landasan Teori

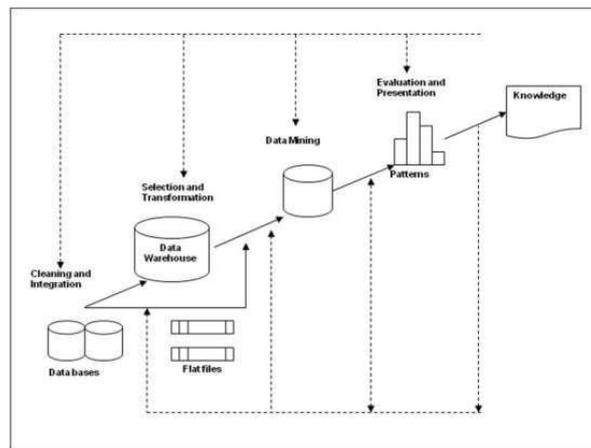
1) Data Mining

Data Mining merupakan suatu istilah yang digunakan untuk menguraikan penemuan didalam databases untuk mendapatkan hasil akurasi yang sesungguhnya. Data mining adalah proses yang menggunakan teknik statistic, matematika, kecerdasan buatan dan mechine learning untuk mengekstaksi dan mengidentifikasi infomasi yang bermanfaat dan pengetahuan yang terkait dari berbagai databases besar[2].

Definisi umum dari Data Mining itu sendiri adalah serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data.

2) Knowledge Discovery in Database (KDD)

Istilah data mining dan knowledge discovery in data bases (KDD) Sering kali 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.



Gambar 1. Data Mining sebagai presos knowledge Discovery

Perusahaan memerlukan kecerdasan bisnis untuk mengembangkan proses bisnis, memonitor waktu, biaya kualitas, da pengendalian. Kecerdasan bisnis merupakan proses pengubahan data menjadi informasi. Dari kumpulan informasi yang ada akan diambil polanya menjadi pengetahuan. Tujuan kecerdasan bisnis adalah mengubah data banyaj menjadi nilai bisnis melalui laporan analistik.

3) Decision Tree

Decision Tree merupakan sebuah metode diagram yang berbentuk seperti struktur pola pohon yang setiap internal note dalam pengujiannya menentukan fungsi-fungsi pendekatan yang bernilai diskrit terhadap suatu atributnya. Metode Decision tree juga digunakan untuk melakukan klasifikasi terhadap sekumpulan suatu fungsi pola data ke dalam label data yang belum diketahui labelnya ke dalam label-label yang sudah ditentukan.

Decision tree merupakan salah satu teknik yang dapat digunakan untuk melakukan klasifikasi terhadap sekumpulan objek atau record.[3]

4) Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Teknik klasifikasi juga merupakan pendekatan sistematis untuk membangun model klasifikasi dari kumpulan data masukan.

Klasifikasi sangat berperan penting dalam proses algoritma C4.5 dimana algoritma C4.5 mampu menghasilkan pohon atauran model yang mudah diinterpretasikan dan diubah untuk strukur aturan. Tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya diketahui. Model itu sendiri bisa berupa keputusan “ Jika - Maka”, Berupa pohon keputusan, formula matematis atau neural network. Proses klasifikasi biasanya dibagi menjadi dua fase: *learning* dan *test*.

Pada fase *learning*, sebagian data yang telah diketahui kelas datanya diumpungkan untuk membentuk model perkiraan. Kemudian pada fase *test* model yang sudah terbentuk diuji dengan sebagian data lainnya untuk mengetahui accuracy dan model tersebut. Bila akurasiya mencukupi model ini dapat dipakai untuk prediksi kelas data yang belum diketahui[4].

5) Algoritma C4.5

Algoritma C4.5 merupakan sebuah algoritma yang digunakan untuk membangun decision tree (Pohon Keputusan). Algoritma C4.5 membuat sebuah pola pohon keputusan dari atas ke bawah. Sebagaimana menentukan atribut paling atas merupakan akan sedangkan yang paling bawah dinamakan daun. Setiap pola berisikan nilai dari atribut-atribut untuk sebuah kelas.

Tiga prinsip cara kerja algoritma C4.5 adalah:

- a. Tujuan dari algoritma pohon keputusan adalah mengkonstruksi struktur data pohon yang digunakan memprediksi kelas dari sebuah kasus atau record yang sesuai.
- b. Pemangkasan pola data juga bertujuan untuk mengurangi tingkat kesalahan prediksi pada kasus (record).
- c. Dalam pohon keputusan terdapat aturan-aturan dalam bentuk *if-then* atau *jika-maka* diturunkan dari pohon keputusan dengan melakukan penelusuran dari akar sampai ke daun [5].

Secara umum algoritma C4.5 untuk membangun pohon keputusan berdasarkan beberapa keputusan seperti:

- a. Pilih atribut sebagai akar.
- b. Buat cabang untuk setiap nilai
- c. Bagi kasus dalam cabang
- d. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Berdasarkan akar dari pohon keputusan algoritma C4.5 adalah atribut data dengan nilai gain teratas.

$$Entropy(s) = \sum_{i=1}^n -p_i \cdot \log_2(p_i) \dots\dots\dots(1)$$

Rumus entropy yang digunakan dalam perhitungannya dimana:

- S : Himpunan kasus
- N : Jumlah partisi S
- Pi : Proporsi dari Si terhadap S

Sementara untuk mendapatkan hasil nilai gain, terlebih dahulu menghitung nilai entropy atribut dalam data.

$$GAIN(S,A)=Entropy(S) - \sum_{i=1}^n |S_i|/|S| * Entropy(S) \dots\dots\dots(2)$$

Rumus gain merupakan rumus yang digunakan dalam perhitungannya dimana :

- S : Himpunan kasus
- A : Atribut
- N : Jumlah partisi atribut A
- |Si| : Jumlah kasus pada partisi ke - i
- |S| : Jumlah kasus dalam S

Dengan mengetahui rumus-rumus diatas, kita dapat memproses data menggunakan metode algoritma decision tree[6].

6) Rapid Miner

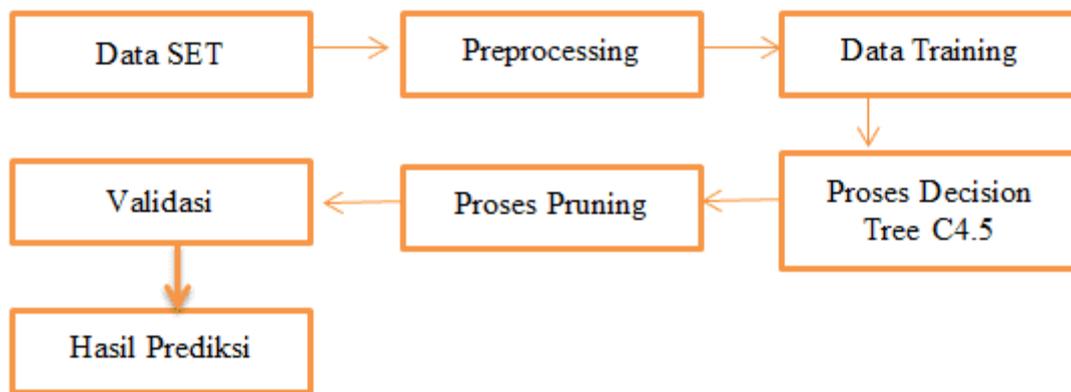
Rapid miner adalah sebuah alat media yang digunakan dalam menentukan perhitungan untuk mendapatkan hasil perhitungan yang lebih bersifat aktif dalam menentukan hasil akurasi yang didapatkan dalam data set yang ada didalam datanya[7].

Sedangkan Rapid Miner merupakan platform perangkat lunak ilmu data yang dikembangkan oleh perusahaan yang bernama sama dengan pihak yang menyediakan lingkungan terintegrasi untuk persiapan data, penambangan teks dan analisis prediktif.

Metode

Dalam menggunakan metode ini penelitian ini bertujuan untuk lebih memahami pengimplementasian data mining menggunakan algoritma C4.5 dalam menentukan keputusan dalam permasalahan yang dibahas. Penelitian ini juga akan menggambarkan accuracy baru untuk sebuah keputusan calon nasabah yang berpotensi. Data yang digunakan dalam penelitian ini diambil pada 2019 bulan april, Data yang digunakan diambil dari data pengiriman sekitar 15.000an namun untuk data testing untuk penelitian ini penulis hanya mengambil sampel data sebanyak 1000an data.

Tahapan metodologi dalam penelitian ini dijelaskan seperti gambar dibawah ini :



Gambar 2. Rancangan Proses Penelitian

Dari rancangan proses penelitian ini dapat dijelaskan sebagai berikut :

- 1) Data Set
Merupakan data yang digunakan dalam data pengiriman ekspedisi barang dalam frekuensi periode satu bulan.
- 2) Preprocessing
Tahap ini adalah tahap untuk memudahkan supaya data bisa terbaca dalam algoritma yang mau digunakan dalam inputannya yang akan dikeolah. Pada penelitian ini, data akan diolah seperti pembersihan data yang tidak digunakan akan disingkirkan atau dibuang. Pembersihan data ini akan menghilangkan atribut data yang tidak diperlukan dalam penelitian.
- 3) Data Training
Sehabis tahap preprocessing diselesaikan, data training ini yang akan dijadikan atau digunakan untuk dikelolah dalam proses ke algoritma C4.5 agar menemukan model prediksi yang diinginkan.
- 4) Proses Decision Tree C4.5
Metode ini merupakan metode klasifikasi yang digunakan dalam penelitian untuk pohon keputusan yang hasilnya akan mempresentasikan aturannya.
- 5) Proses Pruning
Pruning adalah bagian proses pembentkan decision tree. Pembentukan decision tree ini ada beberapa node maupun hasil dari noisedata. Penerapan pruning ini dapat mengurangi outlier maupun noise data pada decision tree awal sehingga dapat meningkatkan nilai akurasi pada proses klasifikasi data yang lebih maksimal.
- 6) Validasi
Pengujian ini dilakukan menggunakan metode cross validasi agar dapat hasil prediksi dengan nilai keakuratan data pengujianya.
- 7) Hasil Prediksi
Dalam hasil output yang akan dihasilkan merupakan hasil dari algoritma C4.5 [8]

3. HASIL PENELITIAN DAN PEMBAHASAN

Proses Pengumpulan Data

Dalam proses pengumpulan data yang digunakan diambil dari sample data pengiriman ekspedisi dalam periode satu bulan. Data yang akan diolah mempunyai beberapa karakteristik data yang merupakan syarat dalam pengolahan data mining dengan menggunakan teknik algoritma C4.5. Dalam penggunaan data asli secara umum ada lima belas ribuan data, Tetapi untuk data yang digunakan hanya kurang lebih seribuanan data sebagai data testing data yang telah diklasifikasikan berdasarkan variable yang ada. Masalah yang harus dipecahkan pada penelitian ini adalah, untuk mendapatkan hasil nilai accuracy yang lebih optimal untuk mengklasifikasi prediksinya.

Tabel 1. Struktur data

Date	Waktu	Waktu	Service	Weight	Weight	Destination	Kota	Shipper Name	Shipper Name	18 000.00	KETERANGAN
01042019	08.00 - 11.00	PAGI	YES	1	Sedikit	BABELAN	BEKASI	NAGAPUSPABAR	SELALU	18 000.00	TIDAK DISKON
01042019	08.00 - 11.00	PAGI	YES	1	Sedikit	BABELAN	BEKASI	NAGAPUSPABAR	SELALU	18 000.00	TIDAK DISKON
01042019	08.00 - 11.00	PAGI	YES	1	Sedikit	BEJI	DEPOK	NAGAPUSPABAR	SELALU	18 000.00	TIDAK DISKON
01042019	08.00 - 11.00	PAGI	REG	1	Sedikit	BINTARA	BEKASI BARAT	NAGAPUSPABAR	SELALU	9 000.00	TIDAK DISKON
01042019	08.00 - 11.00	PAGI	REG	1	Sedikit	JAKAMILYA	BEKASI SELATAN	NAGAPUSPABAR	SELALU	9 000.00	TIDAK DISKON
01042019	08.00 - 11.00	PAGI	REG	1	Sedikit	AREAJAYA	BEKASI TIMUR	NAGAPUSPABAR	SELALU	9 000.00	TIDAK DISKON
01042019	08.00 - 11.00	PAGI	REG	1	Sedikit	AREAJAYA	BEKASI TIMUR	NAGAPUSPABAR	SELALU	9 000.00	TIDAK DISKON
01042019	08.00 - 11.00	PAGI	REG	1	Sedikit	AREAJAYA	BEKASI TIMUR	DAMARA	JARANG	9 000.00	TIDAK DISKON
01042019	08.00 - 11.00	PAGI	CKE	1	Sedikit	AREAJAYA	BEKASI TIMUR	DAMARA	JARANG	8 000.00	TIDAK DISKON
01042019	08.00 - 11.00	PAGI	REG	1	Sedikit	AREAJAYA	BEKASI TIMUR	SCARLETT_WHIT	SELALU	9 000.00	TIDAK DISKON
01042019	08.00 - 11.00	PAGI	REG	1	Sedikit	AREAJAYA	BEKASI TIMUR	FELI SKIN	SERING	9 000.00	TIDAK DISKON
01042019	08.00 - 11.00	PAGI	REG	1	Sedikit	HARAPANBARU	BEKASI UTARA	NINO	SELALU	9 000.00	TIDAK DISKON

1) Pengolahan Data

Dataset ini terdapat tahap preprocessing memiliki 3 proses seperti :

- a. Membuang data yang terduplikasi
- b. Memeriksa data tidak sesuai
- c. Memperbaiki kesalahan pada data

Dan Hasil akhir dari data ini sudah menjadi kumpulan data yang sudah bersih atau tidak ada missing value-nya.

2) Transformasi Data

Transformasi data ini digunakan untuk supaya mengubah nilai atribut yang akan digunakan akan menjadi lebih bervariasi dengan nilai nominal menjadi atribut yang dapat disesuaikan dengan data yang akan dilakukan proses perhitungan algoritma klasifikasi C4.5. Seperti ada proses cleaning data agar lebih akan mendapatkan nilai akurasi lebih aktual diantaranya :

- a. Menggunakan variable atribut yang akan digunakan tidak ada data yang missing.
- b. Membersihkan data set yang akan diambil menjadi lebih akurat, lalu data yang digunakan dalam proses data klasifikasi C4.5 tidak semua digunakan. Hanya atribut dan data yang bersih yang akan digunakan untuk data yang eror atau missing akan disingkirkan dalam proses tersebut.
- c. Pengkelompokan sesuai penggolongan dalam atribut variable seperti service, weight, kota, shippname dalam penggolongan data tersebut agar data memperhitungkan nilai entropy dan information gain sebagai berikut :

Tabel 2. Tabel Transformasi Data

Date	Waktu	Waktu	Service	Weight	Kota	KOTA	Shipper Name	Shipper Name	KETERANGAN
01/04/2019	05.00 - 10.00	PAGI	REG	Sedikit	BEKASI TIMUR	BEKASI	NAGAPUSPABARU	SELALU	TIDAK DISKON
01/04/2019	05.00 - 10.00	PAGI	REG	Sedikit	BEKASI TIMUR	BEKASI	DAMARA	JARANG	TIDAK DISKON
01/04/2019	05.00 - 10.00	PAGI	OKE	Sedikit	BEKASI TIMUR	BEKASI	DAMARA	JARANG	TIDAK DISKON
01/04/2019	05.00 - 10.00	PAGI	REG	Sedikit	BEKASI TIMUR	BEKASI	SCARLETT_WHITENING	SELALU	TIDAK DISKON
01/04/2019	05.00 - 10.00	PAGI	REG	Sedikit	BEKASI TIMUR	BEKASI	FELI SKIN	SERING	TIDAK DISKON
01/04/2019	05.00 - 10.00	PAGI	REG	Sedikit	BEKASI UTARA	BEKASI	NINO	SELALU	TIDAK DISKON
01/04/2019	05.00 - 10.00	PAGI	REG	Sedikit	BEKASI UTARA	BEKASI	NINO	SELALU	TIDAK DISKON
01/04/2019	10.01 - 14.00	SIANG	REG	Sedikit	BEKASI	BEKASI	MHAQUABIO	SERING	TIDAK DISKON

3) Perhitungan nilai Entropy dan Informasion Gain

Dalam untuk menentukan nilai Entropy dan Informasion gain ini, Dapat menggunakan perhitungan dalam rumus nilai (1) dan rumus (2). Menentukan akan dari pohon keputusan akar ini diambil dari atribut terpilih dengan cara menghitung menggunakan rumus (1) dan rumus (2) berdasarkan masing-masing atribut yang digunakan dengan persamaan sebagai berikut contohnya :

Langkah pertama hitung nilai entropy diantaranya sebagai berikut :

- Jumlah kasus dalam data : 1021
- Keputusan data diskon : 73
- Keputusan data tidak diskon : 948

$$Entropy (Total) \left(-\frac{73}{1021} * \log_2 \left(\frac{73}{1021} \right) \right) + \left(-\frac{948}{1021} * \log_2 \left(\frac{948}{1021} \right) \right) = 0,371491154$$

Sedangkan untuk selanjutnya melakukan perhitungan dalam nilai entropy pada masing-masing atribut yang akan digunakan. Berikut untuk entropy pada Waktu :

$$Entropy (Total) \left(-\frac{0}{235} * \log_2 \left(\frac{0}{235} \right) \right) + \left(-\frac{235}{235} * \log_2 \left(\frac{235}{235} \right) \right) = 0$$

Keseluruhan data pagi : 235

Data Pagi Diskon : 0

Data Pagi Tidak Diskon : 235

Sementara untuk mendapatkan hasil nilai gain, terlebih dahulu menghitung nilai Information Gain dengan cara menggunakan rumus (2), Seperti dengan cara sebagai berikut :

$$Gain (Total, Waktu) = 0,371491154 - \left(\left(\frac{235}{1021} * 0 \right) + \left(\frac{223}{1021} * 0 \right) + \left(\frac{242}{1021} * 0,121480478 \right) + \left(\frac{321}{1021} * 0,121480478 \right) \right) = 0,275373451$$

Kemudian setelah hasil keseluruhan nilai Entrophy dan Information Gain berhasil didapatkan hasilnya dari perhitngn rumusnya. Maka kemudia dapat diketahui nilai dari terbesar merupakan weight sedangkan nilai terendah merupakan shippname.

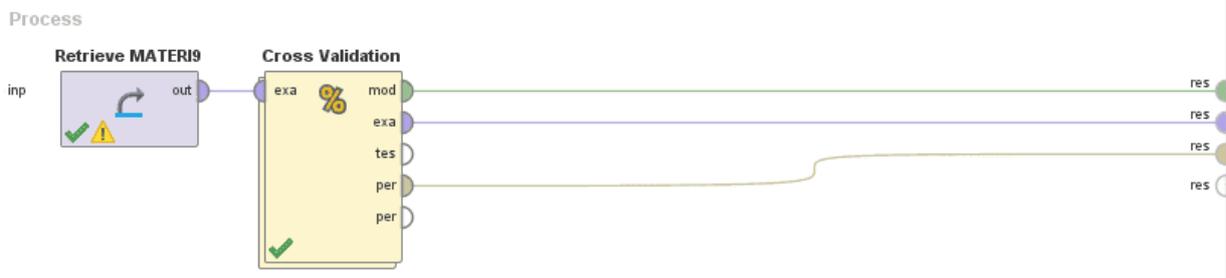
Tabel 3. Hasil perhotungan *Entropy* dan *Information Gain*

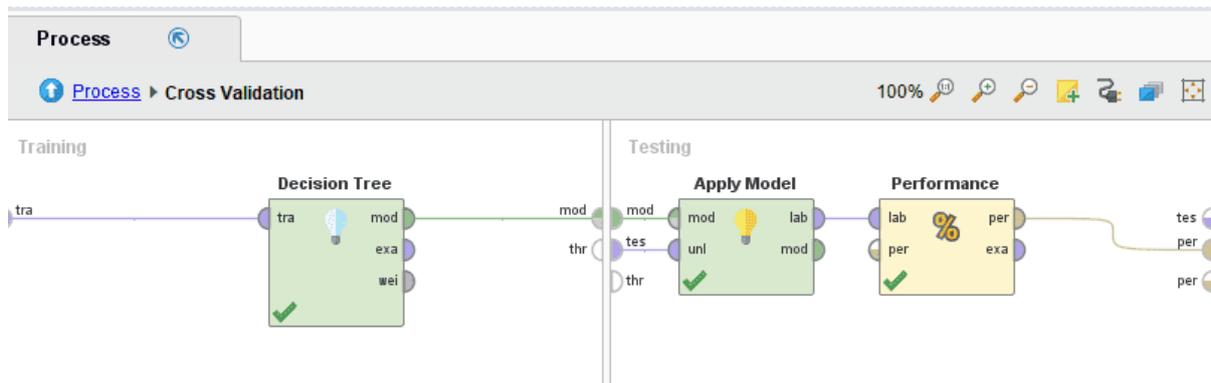
Variable	Nilai	Jumlah Ka	Diskon	Tidak Disk	Entropy	Information Gain
Total		1021	73	948	0,371491154	
Waktu						0,275373451
	PAGI	235	0	235	0	
	SIANG	223	0	223	0	
	SORE	242	4	238	0,121480478	
	MALAM	321	69	252	0,121480478	
Service						0,170405126
	JTR	49	30	19	0,963335546	
	OKE	15	0	15	0	
	REG	916	38	878	0,24905863	
	YES	41	5	36	0,534943699	
Weight						0,170347727
	Banyak	26	13	13	1	
	Lebih Banyak	37	31	6	0,639457132	
	Sering	958	29	292	0,195751089	
Kota						0,016867623
	Bekasi	225	20	205	0,432750159	
	Bogor	72	9	63	0,543564443	
	Depok	92	14	78	0,615253891	
	DKI Jakarta	440	14	426	0,203429561	
	Tangerang	192	16	176	0,41381685	
Shipper Name						-0,042325696
	Jarang	114	6	108	0,297472249	
	Selalu	828	63	765	0,388238978	
	Sering	79	4	75	0,28907946	

4) Validasi dan Evaluasi

Dari hasil data yang diperoleh sesuai katerogi pada variabel yang ada berdasarkan pada data training dan data testing. Terbetuk juga sesuai aturan yang diperoleh dengan pengujian dengan rapidminer didapat dari decision tree dalam prediksi. Desain model yang akan digunakan seperti ada beberapa seperti:

- a. *Retreiving data* : Proses yang digunakan untuk mengimport dataset yang akan digunakan, Diimport ini dalam data import dari file excel.
- b. *Cross Validation* : Metode Validation yang yang digunakan berdasarkan penelitian yang dipakai.
- c. *Decision Tree* : Metode algoritma yang digunakan
- d. *Apply Model* : Operator yang digunakan setelah penerapan dalam data yang akan digunakan.
- e. *Perfomence* : Operator yang digunakan untuk mengukur perfomance nilai akurasi dari model.





Gambar 3. Desain metode algoritma C4.5

PerformanceVector

```

PerformanceVector:
accuracy: 96.28% +/- 1.29% (micro average: 96.28%)
ConfusionMatrix:
True:  TIDAK DISKON  DISKON
TIDAK DISKON:  942    32
DISKON:  6         41
precision: 90.06% +/- 13.83% (micro average: 87.23%) (positive class: DISKON)
ConfusionMatrix:
True:  TIDAK DISKON  DISKON
TIDAK DISKON:  942    32
DISKON:  6         41
recall: 55.89% +/- 18.11% (micro average: 56.16%) (positive class: DISKON)
ConfusionMatrix:
True:  TIDAK DISKON  DISKON
TIDAK DISKON:  942    32
DISKON:  6         41
AUC (optimistic): 0.996 +/- 0.004 (micro average: 0.996) (positive class: DISKON)
AUC: 0.769 +/- 0.123 (micro average: 0.769) (positive class: DISKON)
AUC (pessimistic): 0.604 +/- 0.158 (micro average: 0.604) (positive class: DISKON)
    
```

Gambar 4. Hasil dari validasi cross validation

Berdasarkan metode Ten-Fold Cress Validation menghasilkan nilai akurasi yang hampir sempurna sebesar 96,28%, Precision sebesar 90,06%, dan recall sebesar 55,89% . Hal tersebut menunjukkan bahwa dari hasil proses kalifikasi yang dilakukan akan dapat diterapkan.

Table View Plot View

Changes to a table showing the confusion matrix.

accuracy: 96.28% +/- 1.29% (micro average: 96.28%)

	true TIDAK DISKON	true DISKON	class precision
pred. TIDAK DISKON	942	32	96.71%
pred. DISKON	6	41	87.23%
class recall	99.37%	56.16%	

Gambar 5. Confusion Matrix

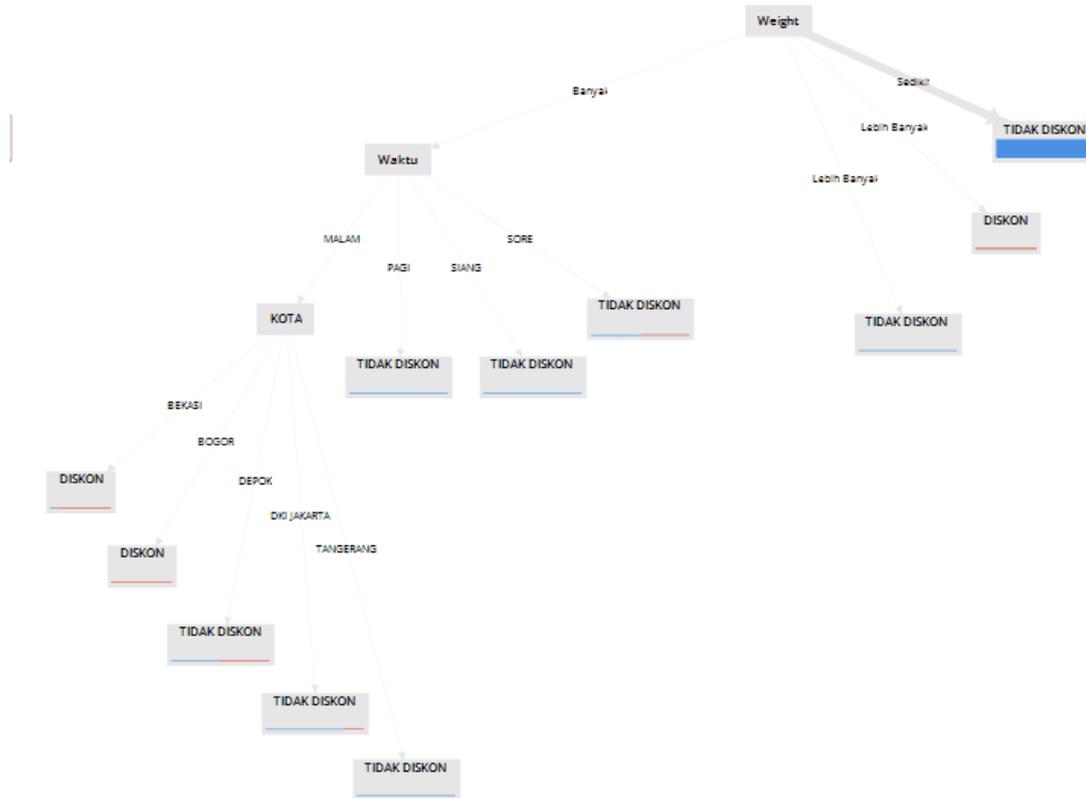
Dari hasil *confusion matrix* diatas, selanjutnya dilakukan perhitungan nilai *accuracy*, *precision*, dan *recall*. Perbandingan nilai *accuracy*, *precision*, dan *recall* yang telah dihitung untuk algoritma C4.5.

$$Accuracy = \left(\left(\frac{942+41}{1021} \right) * 100\% \right) = 96,28\%$$

$$Precision = \left(\left(\frac{41}{6+41} \right) * 100\% \right) = 90,06\%$$

$$\text{Recall} = \left(\frac{41}{32+41} \right) * 100\% = 55.89\%$$

Berdasarkan perhitungan manual diatas dapat disimpulkan bahwa hasil dari perhitungan accuracy, precision, dan recall tersebut sama dengan hasil pada Gambar 4. Berdasarkan hasil dari implemtasi analisa yang dilakukan dengan titik accuracy 96,28, precision 90.06%, recall 55.89% menunjukkan nilai dalam akurasi dan recall yang masih dalam kategori baik. Bahwa disimpulkan peneliti berhasil dalam mengimplementasi algoritma klasfiksasi C4.5 dengan baik. Dan dapat juga diimplementasikan dalam proses pemasaran service untuk pelanggan agar lebih baik untuk proses kerjasama agar berjalan dengan baik dalam waktu jangka panjang.



Gambar 6. Pola Decision Tree

Setelah didapat aturan decision tree kemudian penelitian ini melalui implementasi data mining ini untuk mengetahui nilai accuracy yang akan dihasilkan dari algoritma C4.5.

4. KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil uraian dari penelitian dan pembahasan yang telah dilakukan pada sebelumnya dapat diambil kesimpulan, antara lain:

- 1) Data yang dapat diambil hanya file yang berekstensi *.xls
- 2) Penelitian menggunakan dataset yang dibentuk dari informasi yang dihasilkan pada data pengiriman dalam jangka waktu periode satu bulan.
- 3) Accuracy dari algoritma klasifikasi C4.5 menghasilkan nilai sebesar 96,28% menggunakan data testing sebanyak 1021 data.

Saran

Mengingat keterbatasan yang ada pada penulis, baik pengetahuan, waktu maupun pemikiran, maka penulis dapat memberikan beberapa gambaran untuk saran. Saran ini dapat digunakan untuk referensi untuk dimasa yang akan data maupun untuk pengembangan lebih lanjut mengenai penelitian analisa ini menggunakan algoritma C4.5 adalah sebagai berikut :

- 1) Perlu adanya penambahan variabel tertentu yang akan dilakukan untuk menganalisa jumlah variabel yang akan digunakan.
- 2) Pada penelitian ini, penulis hanya menggunakan salah satu metode decision tree. Untuk mendapatkan hasil yang lebih baik dapat digunakan menggunakan metode data mining dengan algoritma lainnya.

- 3) Dalam penelitian ini penulis hanya menggunakan software data mining seperti rapidminer , perlu juga dilakukan perbandingan lebih lanjut terhadap software aplikasi data mining lainnya. Seperti weka, orange dan lainnya guna melihat hasil dari pengolah yang dilakukan terhadap software yang digunakan.

5. REFERENSI

- [1] I. Graha, "DATA MINING JASA PENGIRIMMAN TITIPAN KILAT DI PT CITRA VAN TITIPAN KILAT (TIKI) DENGAN METODE DECISION TREE."
- [2] E. Elisa, "Prediksi Profit Pada Perusahaan Dengan Klasifikasi Algoritma C4.5," *Klik - Kumpul. J. Ilmu Komput.*, vol. 5, no. 2, p. 179, 2018.
- [3] F. Dwi Meliani Achmad, Budanis, Slamet, "Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree," *J. IPTEK*, vol. 16, no. 1, pp. 18–23, 2012.
- [4] S. A. Aradea, Ariyan Z, and Yuliana A, "Penerapan Decision Tree Untuk Penentuan Pola Data Penerimaan Mahasiswa Baru," *J. Penelit. Sitrotika*, vol. 7, no. 1, pp. 1693–9670, 2011.
- [5] V. S. Moertini, B. Sitohang, and O. S. Santosa, "Integrasi Algoritma Pohon Keputusan C4.5 Yang Dikembangkan Ke Dalam Object-Relational Dbms," *JUTI J. Ilm. Teknol. Inf.*, vol. 6, no. 2, p. 39, 2007.
- [6] A. I. Jamhur, "Penerapan Data Mining Untuk Menganalisa Jumlah Pelanggan Aktif Dengan Menggunakan Algoritma C4.5," *Maj. Ilm.*, vol. Vol. 23, no. No.2, pp. 12–20, 2016.
- [7] R. Avrizal *et al.*, "Implementasi Data Mining dengan Algoritma," vol. 8, pp. 77–84, 2019.
- [8] D. Rosdiana and A. H. Rismayana, "Prediksi waktu tanam cabai menggunakan algoritma c4.5," pp. 436–442, 2018.