

Implementasi Algoritma C4.5 Untuk Klasifikasi Penjualan Barang di Swalayan Dutalia

Aprilianus Kristianus Lalo^{*1}, Patrisius Batarius², Yovinia Carmeneja Hoar Siki³

^{1,2,3}Program Studi Ilmu Komputer, Fakultas Teknik, Universitas Katolik Widya Mandira
Jl. San Juan No 1, Penfui Timur 85361, Kupang Tengah, Kupang – NTT, Telp. (0380) 833395
Email: ¹[rickylalo29@gmail.com](mailto:rickylo29@gmail.com), ²patrisbatarius@unwira.ac.id, ³yoviniacarmeneja@gmail.com

ABSTRAK

Semua perusahaan ritel, termasuk Swalayan Dutalia, dituntut untuk mengambil keputusan secara tepat dalam strategi pemasaran dengan melihat kondisi pasar. Namun, pihak swalayan kerap kali mengalami kesulitan dalam memprediksi angka penjualan produk, terlebih lagi pihak swalayan belum memiliki teknik khusus guna melakukan prediksi angka penjualan barang di masa depan. Hal ini menyebabkan pihak swalayan tidak mampu menentukan stok minimum barang yang dijual.

Oleh karena itu, penelitian ini dilakukan untuk merancang bangun sebuah sistem yang menerapkan algoritma C4.5 untuk mengolah data dalam jumlah besar untuk menemukan pola penjualan barang. Berdasarkan decision tree yang terbentuk, akan dilakukan prediksi terhadap kumpulan data penjualan barang baru. Sehingga, pihak Swalayan Dutalia dapat menentukan stok minimum barang. Atribut yang digunakan berjumlah enam buah, yaitu jenis barang, harga, jumlah jual, waktu jual, momen jual, dan status penjualan sebagai atribut target. Berdasarkan pengujian terhadap data training dan data hasil prediksi pada aplikasi RapidMiner Studio Version 9.7, akurasi perhitungan yang dihasilkan mencapai 100% sehingga sistem yang dirancang bangun dapat digunakan untuk memprediksi penjualan barang di masa depan.

Kata Kunci: Swalayan Dutalia, Algoritma C4.5, Decision Tree, Prediksi Penjualan

ABSTRACT

All of retail companies, including Dutalia Supermarket, are required to take the right decision in marketing strategy by observing market condition. However, the supermarket often has a difficulty to predict numbers of product sale, even it doesn't have a special technique to predict the numbers of sale for the future. So, it causes the supermarket can't determine minimum number of supplies.

Because of that, this research is made to build a system that applying C4.5 algorithm to process a big data of products sale to see pattern of sale. Based on decision tree, a prediction of new collection of products sale data can be made. So, Dutalia Supermarket can determine minimum numbers of supplies. The research uses six attributes, they are kinds of product, price, total of products sale, time of sale, moment of sale, and sales status as the target attribute. Based on data training and prediction data test using RapidMiner Studio Version 9.7 application, the accuration of the calculation reaches 100%, so it can be concluded that the established system can be used to predict products sale in the future.

Keywords: Dutalia Supermarket, C4.5 Algorithm, Decision Tree, Sale Prediction

1. PENDAHULUAN

Perkembangan persaingan di bidang bisnis disertai kemajuan di bidang teknologi informasi mengantarkan perusahaan-perusahaan ritel saat ini, termasuk di Indonesia, pada persaingan yang ketat dan lebih terbuka. Setiap perusahaan dituntut untuk mengambil keputusan secara tepat dalam strategi pemasaran dengan melihat kondisi pasar.

Pada perusahaan ritel, seperti pada Swalayan Dutalia, salah satu cara yang bisa dilakukan untuk mengetahui kondisi pasar adalah dengan mengamati data transaksi penjualan. Data transaksi penjualan disimpan dalam *data warehouse* swalayan. Namun, tumpukan data transaksi tersebut nyatanya belum dimanfaatkan secara efektif. Data yang begitu besar hanya dimanfaatkan sebagai bahan pembuatan laporan penjualan per bulan tanpa menggali informasi atau pengetahuan baru dari pola transaksi yang ada.

Belum adanya cara mengolah informasi yang efektif dari kumpulan data penjualan di tengah angka penjualan produk yang fluktuatif banyak kali menjadi sumber masalah baru bagi pihak swalayan. Pihak swalayan kerap kali mengalami kesulitan dalam memprediksi angka penjualan produk, terlebih lagi pihak swalayan belum memiliki teknik khusus guna melakukan peramalan angka penjualan di masa depan. Hal ini mengakibatkan swalayan kesulitan menentukan stok minimum tiap barang secara efektif, sehingga yang terjadi adalah adanya penimbunan stok barang di gudang perusahaan ataupun habisnya stok barang yang paling diminati konsumen karena kebutuhan pasar tidak diimbangi persediaan barang.

Oleh sebab itu, Swalayan Dutalia membutuhkan suatu teknik otomatis di bidang teknologi informasi guna mengatasi masalah tersebut, yaitu *data mining*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. (Harahap, 2015)

Teknik klasifikasi dengan Algoritma C4.5 dalam *Data Mining* adalah salah satu metode yang dapat dimanfaatkan dalam kaitannya dengan prediksi angka penjualan produk atau barang. Algoritma C4.5 digunakan untuk menghasilkan model berupa *decision tree* (pohon keputusan) dimana pohon atau aturan yang terbentuk berguna dalam membaca prediksi, khususnya dalam kaitan dengan prediksi penjualan produk terlaris di Swalayan Dutalia.

Penggunaan algoritma C4.5 tentu harus dilengkapi dengan variabel-variabel atau atribut-atribut yang turut mengindikasikan sebuah produk dengan status penjualan laris atau tidak laris sehingga pihak swalayan dapat menyiapkan stok yang sesuai dengan kebutuhan pelanggan. Atribut-atribut yang dimaksudkan adalah jenis barang, jumlah penjualan, harga barang, waktu jual, momen penjualan (hari raya, hari biasa, hari libur), dan status penjualan (laris, tidak laris) sebagai atribut kelas (atribut target).

Berdasarkan uraian persoalan yang dihadapi oleh pihak Swalayan Dutalia, maka penting bagi peneliti untuk mengevaluasi data penjualan barang agar dapat menerapkan Algoritma C4.5 untuk klasifikasi produk sehingga pihak swalayan dapat memperoleh informasi berupa pola penjualan produk atau barang berdasarkan riwayat penjualan barang di masa lampau yang akan turut mendukung dalam pengambilan kebijakan perusahaan di masa yang akan datang.

2. KLASIFIKASI

Klasifikasi adalah proses menemukan kumpulan pola atau fungsi-fungsi yang mendeskripsikan dan memisahkan kelas data satu dengan lainnya agar dapat digunakan untuk memprediksi data yang belum memiliki kelas data tertentu. (Rosela, 2019)

2.1 Algoritma C4.5

Algoritma C4.5 merupakan salah satu teknik klasifikasi pada *machine learning* yang digunakan pada proses *data mining* dengan membentuk sebuah pohon keputusan (*decision tree*) yang direpresentasikan dalam bentuk aturan. (Pritalia, 2018)

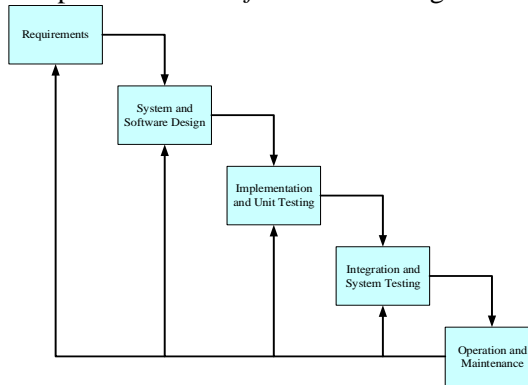
Berikut adalah langkah-langkah perhitungan:

1. data dikelompokkan berdasarkan atribut beserta nilai di dalamnya;
2. menghitung jumlah data pada setiap nilai atribut yang ada;
3. mengklasifikasi data yang sudah dihitung menjadi dua kelompok berdasarkan target tujuan, yaitu laris (L) dan tidak laris (TL);
4. menghitung *entropy* total dari 20 data penjualan barang;
5. menghitung *entropy* dari masing-masing nilai atribut;
6. menghitung *gain* dari tiap atribut;
7. menghitung *gain ratio* dari tiap atribut;
8. mencari atribut dengan *gain ratio* tertinggi untuk dijadikan *root*;
9. menentukan nilai atribut yang akan dijadikan cabang;
10. menentukan *node* selanjutnya dari atribut yang terpilih berdasarkan nilai *gain ratio* tertinggi.

3. METODE PENELITIAN

Dalam penelitian ini, metode yang diterapkan adalah metode rekayasa perangkat lunak model *waterfall*. Model *Waterfall* adalah suatu proses pengembangan perangkat lunak berurutan, dimana kemajuan dipandang sebagai arus yang terus mengalir ke bawah (seperti air terjun) melewati fase-fase perencanaan, pemodelan, implementasi (konstruksi), dan pengujian. (Trisianto, 2018)

Tahap-tahap dalam model penelitian *waterfall* adalah sebagai berikut.



Gambar 1. Tahap-Tahap Metode *Waterfall*

Penelitian ini membutuhkan data atribut, data nilai atribut, data *training*, dan data prediksi seperti terlihat pada Tabel 1, Tabel 2 dan Tabel 3.

Tabel 1. Data Atribut dan Nilai Atribut

Atribut	Nilai Atribut
Jenis Barang	Beras Jeruk 20 kg
	Bimoli 1 Ltr Rfl
	Bimoli 2 Ltr Rfl
	Mie Sedap Goreng
	Gula Pasir Hasta 1
	Terigu 1 kg
	Terigu 500 g
	Kopi TB 450
	Teh Sari Wangi 25
	Mie Sedap Soto
	Telur Ayam
	Daia Bunga 850 g
	Rinso An 900 g
	Dancow Cokelat Forts
	Dancow Cokelat 800 g
	Garam C Kpl 250
	Bear Brand 189 ML
	Mama Lemon Jrk Nps 400 ml
	Lifebuoy Lemonfresh
	SGM Ananda 400g 6-12 bln
	You C1000 Vit. Orange
	Aqua 1500 ml/fc
	Bir Bintang Btl 620
Coca Cola Btl 1.5 Lt	
Pulpy Orange 350	
Le Minerale Btl 600 ml	
Mamy Poko Pant Standar	

Atribut	Nilai Atribut
	Nescafe Classic 10x2g
	DCW 1 + MD 1000 g
	Kecap Manis Sedap 135 ml
	Krisbee Fries Kentang Grg 75 g
	Kopi Top White 12x21 g
	Jas Jus 10 g Melon
	Morinaga Bmt Lacto 400g
	Chilml 2 DHA 200 Original
	Chilml DHA 400 g
Harga	< 100 ribu
	>= 100 ribu
Jumlah Jual	<= 250
	> 250
Waktu Jual	Januari
	Februari
	Maret
	April
	Mei
	Juni
	Juli
	Agustus
	September
	Oktober
	November
	Desember
Momen Jual	Hari Raya
	Hari Biasa
	Hari Libur

Data atribut dan nilai atribut merupakan kriteria yang memengaruhi omzet penjualan barang. *Data training* dan data prediksi yang di-*input* harus mencakup atribut dan nilai atribut tersebut.

Tabel 2. *Data Training*

NO.	Jenis Barang	Harga	Jumlah Jual	Waktu Jual	Momen Jual	Status
1	Beras Jeruk 20 kg	>= 100 ribu	<= 250	Januari	Hari Raya	Tidak Laris
2	Bimoli 1 Liter Rfl	< 100 ribu	<= 250	Januari	Hari Raya	Laris
3	Bimoli 2 Liter Rfl	< 100 ribu	<= 250	Januari	Hari Biasa	Tidak Laris
4	Mie Sedap Goreng	< 100 ribu	> 250	Januari	Hari Biasa	Laris
5	Gula Pasir Hasta 1	< 100 ribu	> 250	Januari	Hari Raya	Laris

NO.	Jenis Barang	Harga	Jumlah Jual	Waktu Jual	Momen Jual	Status
6	Terigu 1 kg	< 100 ribu	< = 250	Januari	Hari Raya	Tidak Laris
7	Terigu 500 g	< 100 ribu	< = 250	Januari	Hari Libur	Tidak Laris
8	Kopi TB 450	< 100 ribu	< = 250	Januari	Hari Biasa	Laris
9	Teh Sari Wangi 25	< 100 ribu	> 250	Januari	Hari Biasa	Laris
10	Mie Sedap Soto	< 100 ribu	> 250	Januari	Hari Biasa	Laris
....						
430	Morinaga Bmt Lacto 400g	< 100 ribu	< = 250	Desember	Hari Libur	Tidak Laris
431	Chilml 2 DHA 200 Original	< 100 ribu	< = 250	Desember	Hari Libur	Tidak Laris
432	Chilml DHA 400 g	< 100 ribu	< = 250	Desember	Hari Biasa	Tidak Laris

Data training digunakan dalam proses *training* untuk menemukan pola penjualan menggunakan Algoritma C4.5. Pola yang dihasilkan berupa pohon keputusan yang berisikan *rules* penjualan. *Rules* yang dihasilkan akan digunakan untuk melakukan prediksi status penjualan barang yang baru.

Tabel 3. Data Prediksi

NO.	Jenis Barang	Harga	Jumlah Jual	Waktu Jual	Momen Jual
1	Beras Jeruk 20 kg	> = 100 ribu	< = 250	Januari	Hari Raya
2	Bimoli 1 Liter Rfl	< 100 ribu	< = 250	Januari	Hari Raya
3	Bimoli 2 Liter Rfl	< 100 ribu	< = 250	Januari	Hari Biasa
4	Mie Sedap Goreng	< 100 ribu	> 250	Februari	Hari Biasa
5	Gula Pasir Hasta 1	< 100 ribu	> 250	Februari	Hari Biasa
6	Terigu 1 kg	< 100 ribu	< = 250	Februari	Hari Libur
7	Terigu 500 g	< 100 ribu	< = 250	Maret	Hari Raya
8	Kopi TB 450	< 100 ribu	< = 250	Maret	Hari Libur
9	Teh Sari Wangi 25	< 100 ribu	> 250	Maret	Hari Biasa
10	Mie Sedap Soto	< 100 ribu	> 250	April	Hari Biasa
.....					
40	Coca Cola Btl 1.5 Lt	< 100 ribu	> 250	Desember	Hari Raya
41	Pulpy Orange 350	< 100 ribu	> 250	Desember	Hari Libur
42	Le Minerale Btl 600 ml	< 100 ribu	> 250	Desember	Hari Libur
43	Mamy Poko Pant Standar	< 100 ribu	< = 250	Desember	Hari Biasa
44	Nescafe Classic 10x2g	< 100 ribu	< = 250	Desember	Hari Biasa
45	DCW 1 + MD 1000 g	< 100 ribu	< = 250	Desember	Hari Libur
46	Kecap Manis Sedap 135 ml	< 100 ribu	< = 250	Desember	Hari Biasa
47	Krisbee Fries Kentang Grg 75 g	< 100 ribu	< = 250	Desember	Hari Biasa
48	Kopi Top White 12x21 g	< 100 ribu	< = 250	Desember	Hari Biasa
49	Jas Jus 10 g Melon	< 100 ribu	< = 250	Desember	Hari Biasa
50	Morinaga Bmt Lacto 400g	< 100 ribu	< = 250	Desember	Hari Libur

Data prediksi berisikan data barang beserta atribut yang memengaruhi omzet penjualannya. Data prediksi belum memiliki status penjualan, sehingga harus dilakukan prediksi untuk menentukan status penjualannya.

4. HASIL DAN PEMBAHASAN

4.1 Perhitungan *Root Decision Tree* Dari Data Training

Pada perhitungan penentuan *root* (akar), data akan dikelompokkan berdasarkan atribut dan nilai atribut. Kemudian, setiap nilai atribut akan dicari nilai *entropy*-nya. Setelah diketahui nilai *entropy*, setiap atribut akan dicari nilai *gain*, *split info*, dan *gain ratio*.

Tabel 4. Perhitungan Penentuan *Root*

No de	Atribut	Nilai Atribut	S	Laris	Tidak Laris	Entropy	Gain	Split Info	Gain Ratio
0	Total		432	246	186	0.986			
	Jenis Barang	Beras Jeruk 20 kg	12	11	1	0.414			
		Bimoli 1 Ltr Rfl	12	10	2	0.650			
		Bimoli 2 Ltr Rfl	12	2	10	0.650			
		Mie Sedap Goreng	12	10	2	0.650			
		Gula Pasir Hasta 1	12	12	0	0			
		Terigu 1 kg	12	11	1	0.414			
		Terigu 500 g	12	7	5	0.980			
		Kopi TB 450	12	12	0	0			
		Teh Sari Wangi 25	12	12	0	0			
		Mie Sedap Soto	12	12	0	0			
		Telur Ayam	12	12	0	0			
		Daia Bunga 850 g	12	10	2	0.650			
		Rinso An 900 g	12	0	12	0			
		Dancow Cokelat Forts	12	11	1	0.414			
		Dancow Cokelat 800 g	12	0	12	0			
		Garam C Kpl 250	12	12	0	0			
		Bear Brand 189 ML	12	12	0	0			
		Mama Lemon Jrk Nps 400 ml	12	1	11	0.414			
		Lifebuoy Lemonfresh	12	12	0	0			
		SGM Ananda 400g 6-12 bln	12	4	8	0.918			
		You C1000 Vit. Orange	12	12	0	0			
		Aqua 1500 ml/fc	12	12	0	0			
		Bir Bintang Btl 620	12	12	0	0			
		Coca Cola Btl 1.5 Lt	12	9	3	0.811			
		Pulpy Orange 350	12	12	0	0			
		Le Minerale Btl 600 ml	12	12	0	0			
		Mamy Poko Pant Standar	12	2	10	0.650			
	Nescafe Classic 10x2g	12	0	12	0				
	DCW 1 + MD 1000 g	12	0	12	0				
	Kecap Manis Sedap 135 ml	12	4	8	0.918				
	Krisbee Fries Kentang Grg 75 g	12	5	7	0.980				

No de	Atribut	Nilai Atribut	S	Laris	Tidak Laris	Entropy	Gain	Split Info	Gain Ratio
		Kopi Top White 12x21 g	12	2	10	0.650	0.692	5.170	0.134
		Jas Jus 10 g Melon	12	0	12	0			
		Morinaga Bmt Lacto 400g	12	1	11	0.414			
		Chilml 2 DHA 200 Original	12	0	12	0			
		Chilml DHA 400 g	12	0	12	0			
	Harga	< 100 ribu	420	235	185	0.990	0.012	0.183	0.067
		>= 100 ribu	12	11	1	0.414			
	Jumlah Jual	<= 250	267	84	183	0.898	0.381	0.959	0.397
		> 250	165	162	3	0.131			
	Waktu Jual	Januari	36	16	20	0.991	0.010	3.585	0.003
		Februari	36	18	18	1.000			
		Maret	36	19	17	0.998			
		April	36	19	17	0.998			
		Mei	36	20	16	0.991			
		Juni	36	23	13	0.944			
		Juli	36	21	15	0.980			
		Agustus	36	23	13	0.944			
		September	36	20	16	0.991			
		Oktober	36	22	14	0.964			
		November	36	22	14	0.964			
		Desember	36	23	13	0.944			
	Momen Jual	Hari Raya	54	45	9	0.650	0.043	1.280	0.034
		Hari Biasa	276	136	140	1.000			
		Hari Libur	102	65	37	0.945			

Perhitungan *Entropy* Total tabel penentuan *ROOT* adalah sebagai berikut.

$$Entropy (Total) = \left(-\frac{246}{432} * \log_2\left(\frac{246}{432}\right)\right) + \left(-\frac{186}{432} * \log_2\left(\frac{186}{432}\right)\right) = 0.986$$

Perhitungan *Entropy* pada tiap nilai atribut dihitung dengan cara yang sama.

$$Entropy (Momen Jual, Hari Raya) = \left(-\frac{45}{54} * \log_2\left(\frac{45}{54}\right)\right) + \left(-\frac{9}{54} * \log_2\left(\frac{9}{54}\right)\right) = 0.65$$

Lalu, perhitungan nilai *Gain* untuk atribut Momen Jual dan atribut lainnya dihitung dengan persamaan *gain* sebagai berikut.

$$Gain (Total, Momen Jual) = 0.986 - \left(\frac{54}{432} * 0.65 + \frac{276}{432} * 1 + \frac{102}{432} * 0.945\right) = 0.043$$

Selanjutnya, perhitungan nilai *Split Info* dan *Gain Ratio* untuk atribut Momen Jual dan atribut lainnya dihitung dengan cara sebagai berikut.

$$SplitInfo (Momen Jual) = -\left(\left(\frac{54}{432} * \log_2\left(\frac{54}{432}\right)\right) + \left(\frac{276}{432} * \log_2\left(\frac{276}{432}\right)\right) + \left(\frac{102}{432} * \log_2\left(\frac{102}{432}\right)\right)\right) = 1.28$$

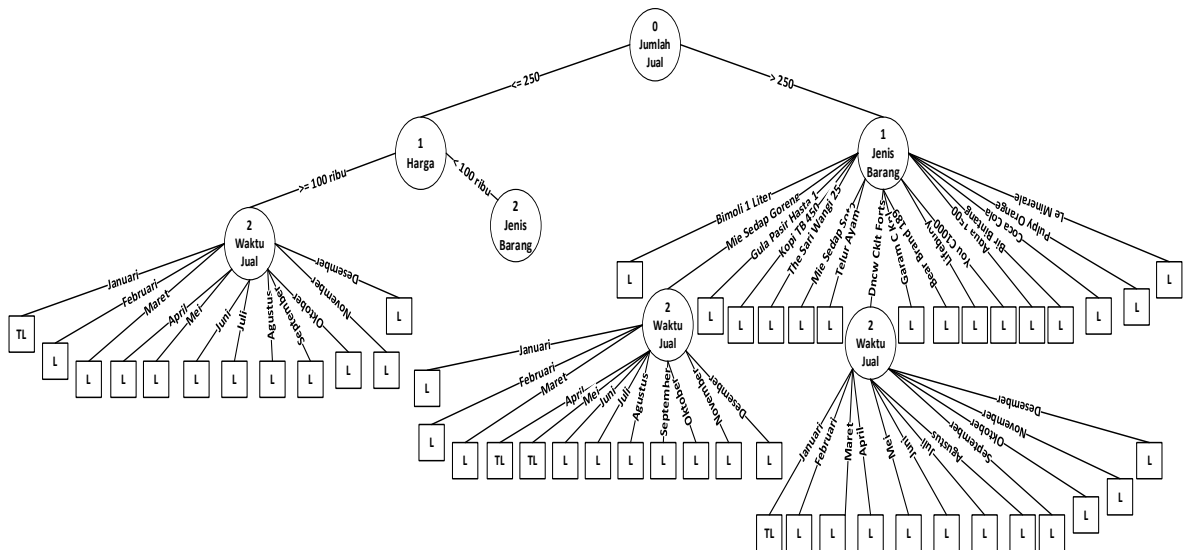
$$Gain\ Ratio\ (Momen\ Jual) = \frac{Gain(S,A)}{SplitInfo(S,A)} = \frac{0.043}{1.28} = 0.034$$

Berdasarkan hasil perhitungan pada tabel, atribut yang memiliki nilai *Gain Ratio* terbesar adalah Jumlah Jual, yaitu sebesar 0.397. Maka, Jumlah Jual menjadi node akar (*root*). Terdapat dua nilai atribut Jumlah Jual, yaitu (≤ 250) dan (> 250). Dua nilai atribut Jumlah Jual diklasifikasikan berdasarkan nilai Laris (L) dan Tidak Laris (TL). Apabila nilai L = 0 dan TL > 0, maka akan menghasilkan rule Tidak Laris (TL). Sebaliknya, apabila nilai L > 0 dan TL = 0, maka menghasilkan rule Laris. Berdasarkan tabel, nilai atribut (≤ 250) dan (> 250) sama-sama memiliki nilai L > 0 dan TL > 0, maka perlu dihitung lagi untuk mencari *Node* selanjutnya.

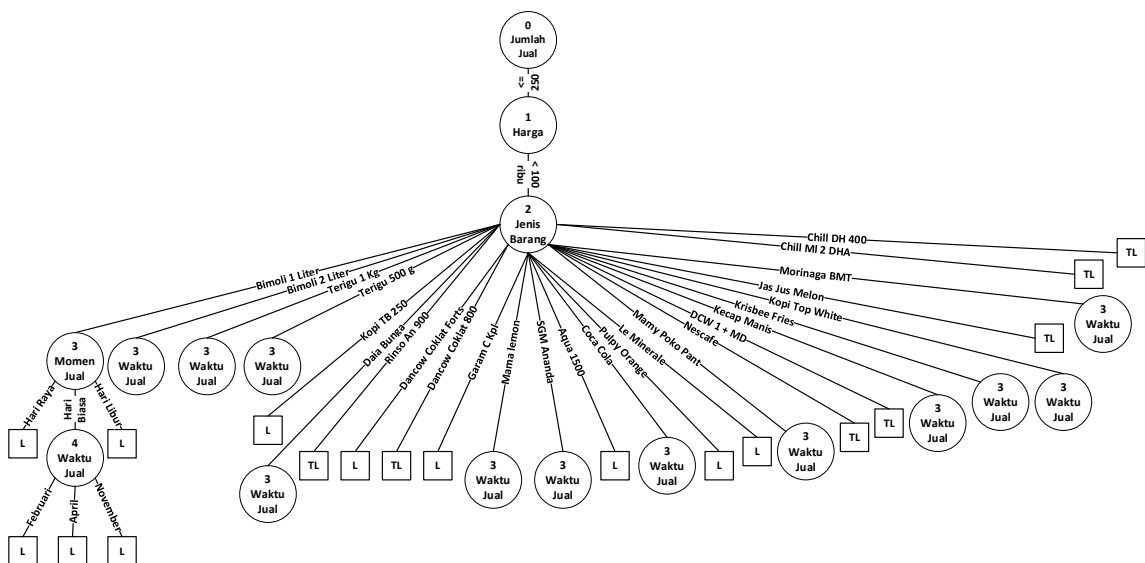
Metode perhitungan untuk *node* selanjutnya sama dengan perhitungan menentukan *node* yang ada. Perbedaannya adalah atribut dengan nilai terbesar pada perhitungan sebelumnya tidak dihitung lagi. Atribut total diganti dengan total dari nilai atribut terbesar pada perhitungan sebelumnya.

4.2 Decision Tree yang Terbentuk

Berdasarkan perhitungan secara keseluruhan, maka dihasilkan sebuah pohon keputusan akhir yang merangkum semua atribut beserta nilai atributnya pada Gambar 2 dan Gambar 3 berikut.



Gambar 2. Decision Tree yang Terbentuk



Gambar 3. Decision Tree Lanjutan

Secara keseluruhan, pohon keputusan yang telah terbentuk akan mencapai daun (*leaf*). Daun adalah status penjualan barang, yaitu Laris atau Tidak Laris. Jika cabang telah mencapai daun, maka cabang tersebut akan membentuk sebuah *rule*.

4.3 Rules yang Dihasilkan

Rules yang terbentuk dari hasil *training* berjumlah 206 *rules* yang dijabarkan sebagai berikut.

1. IF Jumlah Jual (≤ 250) AND Harga (≥ 100 ribu) AND Waktu Jual (Januari) THEN Tidak Laris.
2. IF Jumlah Jual (≤ 250) AND Harga (≥ 100 ribu) AND Waktu Jual (Februari) THEN Laris.
3. IF Jumlah Jual (≤ 250) AND Harga (≥ 100 ribu) AND Waktu Jual (Maret) THEN Laris.
4. IF Jumlah Jual (≤ 250) AND Harga (≥ 100 ribu) AND Waktu Jual (April) THEN Laris.
5. IF Jumlah Jual (≤ 250) AND Harga (≥ 100 ribu) AND Waktu Jual (Mei) THEN Laris.
6. IF Jumlah Jual (≤ 250) AND Harga (≥ 100 ribu) AND Waktu Jual (Juni) THEN Laris.
7. IF Jumlah Jual (≤ 250) AND Harga (≥ 100 ribu) AND Waktu Jual (Juli) THEN Laris.
8. IF Jumlah Jual (≤ 250) AND Harga (≥ 100 ribu) AND Waktu Jual (Agustus) THEN Laris.
9. IF Jumlah Jual (≤ 250) AND Harga (≥ 100 ribu) AND Waktu Jual (September) THEN Laris.
10. IF Jumlah Jual (≤ 250) AND Harga (≥ 100 ribu) AND Waktu Jual (Oktober) THEN Laris.
-
201. IF Jumlah Jual (> 250) AND Jenis Barang (You C1000 Vit. Orange) THEN Laris.
202. IF Jumlah Jual (> 250) AND Jenis Barang (Aqua 1500 ml/fc) THEN Laris.
203. IF Jumlah Jual (> 250) AND Jenis Barang (Bir Bintang Btl 620) THEN Laris.
204. IF Jumlah Jual (> 250) AND Jenis Barang (Coca Cola Btl 1.5 Lt) THEN Laris.
205. IF Jumlah Jual (> 250) AND Jenis Barang (Pulpy Orange 350) THEN Laris.
206. IF Jumlah Jual (> 250) AND Jenis Barang (Le Minerale Btl 600 ml) THEN Laris.

4.4 Prediksi Data Penjualan Menggunakan Aplikasi yang Dirancang Bangun

Proses *training data* yang telah dilakukan menghasilkan sebuah pohon keputusan dan *rules*. *Rules* tersebut berfungsi dalam menentukan prediksi status penjualan barang di Swalayan Dutalia. Operator akan menginput data prediksi melalui aplikasi. Data barang yang diinput tidak menyertakan atribut target, yaitu status. Sistem akan melakukan prediksi status penjualan (laris atau tidak laris) berdasarkan *rules* yang sudah terbentuk. Tabel 5 berikut adalah *interface* hasil prediksi pada aplikasi yang dibangun.

Nomor	Jenis Barang	Harga	Jumlah Jual	Waktu Jual	Momen Jual	Status
1	Beras Jeruk 20 kg	≥ 100 ribu	≤ 250	Januari	Hari Raya	Tidak Laris
2	Bimoli 1 Liter Rfl	< 100 ribu	≤ 250	Januari	Hari Raya	Laris
3	Bimoli 2 Liter Rfl	< 100 ribu	≤ 250	Januari	Hari Biasa	Tidak Laris
4	Mie Sedap Goreng	< 100 ribu	> 250	Februari	Hari Biasa	Laris
5	Gula Pasir Hasta 1	< 100 ribu	> 250	Februari	Hari Biasa	Laris
6	Terigu 1 kg	< 100 ribu	≤ 250	Februari	Hari Libur	Tidak Laris
7	Terigu 500 g	< 100 ribu	≤ 250	Maret	Hari Raya	Laris
8	Kopi TB 450	< 100 ribu	≤ 250	Maret	Hari Libur	Laris
9	Teh Sari Wangi 25	< 100 ribu	> 250	Maret	Hari Biasa	Laris
10	Mie Sedap Soto	< 100 ribu	> 250	April	Hari Biasa	Laris

Gambar 4. Data Prediksi yang Di-input

Berdasarkan hasil *training* pada aplikasi dan perhitungan manual menggunakan *Microsoft Office Excel*, masing-masing menghasilkan *result* yang sama. Selain itu, *decision tree* dan *rules* yang terbentuk pun sama. Sehingga, sistem yang dirancang dapat digunakan untuk melakukan proses *training data mining* menggunakan algoritma C4.5. Setelah diuji coba menggunakan aplikasi RapidMiner Studio Version 9.7, tingkat akurasi hasil *training data* dengan hasil prediksi mencapai 100%. Data hasil prediksi diuji kembali sebagai *data testing* pada aplikasi RapidMiner Studio dan menghasilkan hasil serupa dengan dengan sistem yang dirancang.

5. ANTARMUKA SISTEM

5.1 Training Data

No	Referensi	Jenis Barang	Harga	Jumlah Jual	Waktu Jual	Momen Jual	Status	Action
1	6021000300071	Beras Jeruk 20 Kg	> = 100 rba	< = 200	Januari	Hari Raya	Tidak Laris	👍👎
2	6010003000007	Bimoli 1 Liter RFI	< 100 rba	< = 200	Januari	Hari Raya	Laris	👍👎
3	6010003000000	Bimoli 2 Liter RFI	< 100 rba	< = 200	Januari	Hari Biasa	Tidak Laris	👍👎
4	1030003000001	Mie Sedap Goreng	< 100 rba	> 200	Januari	Hari Biasa	Laris	👍👎
5	6021000300071	Beras Jeruk 20 Kg	> = 100 rba	< = 200	Februari	Hari Biasa	Laris	👍👎
6	6010003000007	Bimoli 1 Liter RFI	< 100 rba	< = 200	Februari	Hari Biasa	Laris	👍👎
7	6010003000000	Bimoli 2 Liter RFI	< 100 rba	< = 200	Februari	Hari Libur	Tidak Laris	👍👎
8	1030003000001	Mie Sedap Goreng	< 100 rba	> 200	Februari	Hari Biasa	Laris	👍👎
9	6021000300071	Beras Jeruk 20 Kg	> = 100 rba	< = 200	Maret	Hari Biasa	Laris	👍👎
10	6010003000007	Bimoli 1 Liter RFI	< 100 rba	< = 200	Maret	Hari Libur	Tidak Laris	👍👎

Gambar 5. Interface Data Training

Halaman *training data* bertujuan menampilkan tabel data *training* yang akan melalui proses *training* dengan algoritma C4.5. Data yang ditampilkan adalah atribut dan setiap nilainya.

```

Perhitungan
- Perhitungan Gini Index
  Jenis Barang:
  Beras Jeruk 20 Kg Gini(0,281) = 0,722
  Bimoli 1 Liter RFI Gini(0,281) = 0,373
  Bimoli 2 Liter RFI Gini(0,281) = 0,373
  Mie Sedap Goreng Gini(0,719) = 0,475
  Gini = 0,462
  Nilai Entropi = 2
  Gini Index = 0,903

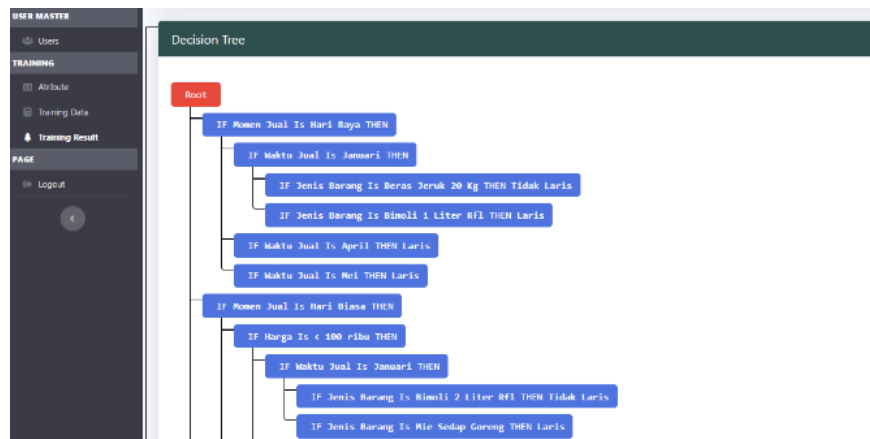
  Harga:
  > = 100 rba Gini(0,281) = 0,722
  < 100 rba Gini(0,719) = 0,397
  Gini = 0,462
  Split Entro = 0,515
  Gini Index = 0,462

  Jumlah Jual:
  < = 200 Gini(0,281) = 0,352
  > 200 Gini(0,719) = 0,463
  Gini = 0,462
  Split Entro = 0,505
  Gini Index = 0,462

  Waktu Jual:
  Januari Gini(0,281) = 1
  Februari Gini(0,719) = 0,411
  Maret Gini(0,281) = 1
  April Gini(0,719) = 1
  Mei Gini(0,281) = 0,833
    
```

Gambar 6. Interface Hasil Perhitungan Data Training

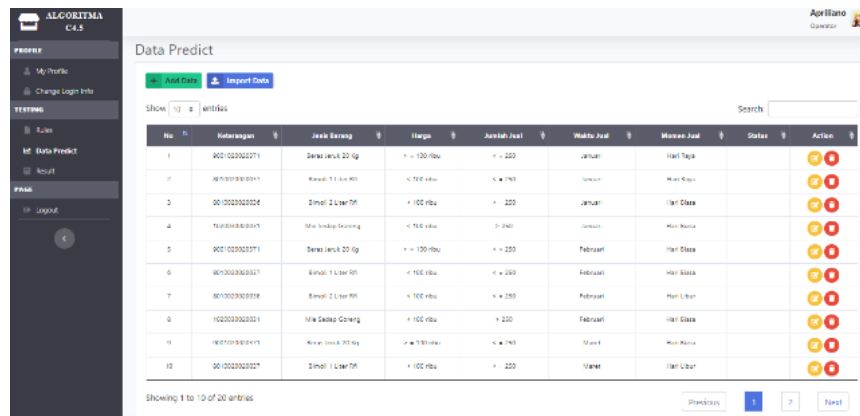
Training result (perhitungan) akan menampilkan hasil perhitungan terhadap *data training* yang tersedia. Perhitungan yang ditampilkan tersebut dirinci untuk setiap atributnya.



Gambar 7. Interface Decision Tree yang Dihasilkan

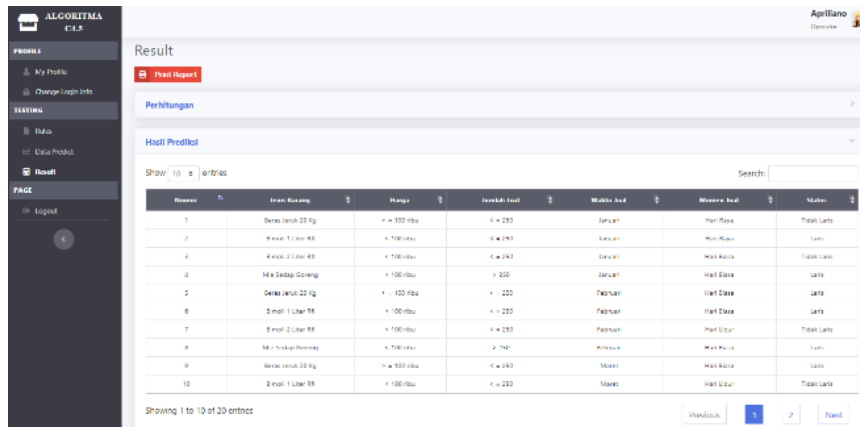
Training result (*decision tree*) menampilkan pohon keputusan yang terbentuk dari hasil perhitungan yang ada. Pohon keputusan ini yang membentuk *rules* terhadap data penjualan.

5.2 Prediksi Data Penjualan



Gambar 8. Interface Data Prediction

Halaman *data predict* bertujuan untuk menampilkan data yang akan dilakukan prediksi. Data penjualan yang di-*input* tidak disertakan status penjualannya. Proses prediksi-lah yang akan menentukan status penjualan barang. Selain itu, operator dapat melakukan *insert*, *update*, dan *delete* data prediksi. Proses *insert data predict* juga dapat dilakukan dengan melakukan *import file* berekstensi *xlxs*.



Gambar 9. Interface Tabel Hasil Prediksi

Halaman *result* (hasil prediksi) bertujuan untuk menampilkan tabel hasil prediksi terhadap data penjualan. Hasil prediksi ditandai dengan sudah terdapat status penjualan untuk setiap barang beserta atribut-atribut yang menentukan. Terdapat tombol *print report* yang bertujuan agar operator dapat melakukan pencetakan hasil prediksi yang ada.

6. KESIMPULAN

Berdasarkan hasil pengujian dan analisis hasil, maka dapat disimpulkan beberapa hal sebagai berikut.

1. Sistem yang dirancang bangun bermanfaat bagi pihak Swalayan Dutalia melakukan klasifikasi berupa *training data* penjualan di masa lampau menggunakan algoritma C4.5 guna menghasilkan *rules* yang dapat digunakan untuk melakukan prediksi data penjualan masa depan. Sehingga, pihak swalayan dapat menentukan stok minimum yang harus disediakan.
2. Sistem yang dirancangbangun dapat melakukan *create*, *read*, *update*, dan *delete* (CRUD) data atribut, nilai atribut, *data training*, dan data prediksi. Selain itu, *user* juga dapat melihat hasil prediksi dan hasil *training* berupa perhitungan dan *decision tree*.
3. Sistem yang dirancangbangun dapat melakukan *input data training* dan data prediksi menggunakan fitur *import excel*.
4. Sistem yang dirancangbangun dapat mencetak laporan hasil prediksi berupa tabel data hasil prediksi.
5. Sistem klasifikasi menggunakan algoritma C4.5 yang dirancangbangun memiliki tingkat akurasi 100% setelah 50 data hasil prediksi diuji coba sebagai *data testing* pada aplikasi RapidMiner Studio

Version 9.7. *Decision tree* yang dihasilkan pada perhitungan manual, sistem yang dirancang bangun, dan aplikasi RapidMiner pun sama.

DAFTAR PUSTAKA

- [1]. Anharku. (2016). Flowchart. Retrieved from Analisis dan Perancangan Sistem website: <http://ilmukomputer.org>
- [2]. Eska, J. (2016). PENERAPAN DATA MINING UNTUK PREDIKSI PENJUALAN WALLPAPER MENGGUNAKAN ALGORITMA C4.5. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 2(2), 9–13.
- [3]. Harahap, F. (2015). Penerapan Data Mining dalam Memprediksi Pembelian cat. *Konferensi Nasional Sistem & Informatika 2015 STMIK*, 1(1), 856–862.
- [4]. Kamagi, D. H., & Hansun, S. (2014). Implementasi Data Mining dengan Algoritma C4 . 5 untuk Memprediksi Tingkat Kelulusan Mahasiswa. *ULTIMATICS*, VI(1), 15–20.
- [5]. Mardi, Y. (2016). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, 2(2), 213–219.
- [6]. Nurmalina, R. (2017). Perencanaan dan Pengembangan Aplikasi Absensi Mahasiswa Menggunakan Smart Card Guna Pengembangan Kampus Cerdas (Studi Kasus Politeknik Negeri Tanah Laut). *Jurnal Integrasi*, 9(1), 84–91.
- [7]. Oracle. (2019). MySQL. Retrieved from Wikipedia Ensiklopedia Bebas website: <https://id.wikipedia.org/wiki/MySQL>
- [8]. Pritalia, G. L. (2018). Penerapan Algoritma C4 . 5 untuk Penentuan Ketersediaan Barang E-commerce. *Indonesian Journal of Information Systems (IJIS) Vol.*, 1(1), 47–56.
- [9]. Rosela, Y. (2019). IMPLEMENTASI KLASIFIKASI DECISION TREE MENGANALISA STATUS PENJUALAN BARANG MENGGUNAKAN C4 . 5 (Studi Kasus : Pt . Matahari Department Store Medan Mall). *Jurnal Pelita Informatika*, 18(1), 143–150.
- [10]. Suyanto, D. (2017). *Data Mining Untuk Klasifikasi Dan Klasterisasi Data* (1st ed.; I. Bandung, Ed.). Bandung: Informatika Bandung.
- [11]. Trisianto, C. (2018). PENGGUNAAN METODE WATERFALL UNTUK PENGEMBANGAN SISTEM MONITORING DAN EVALUASI PEMBANGUNAN PEDESAAN. *Jurnal Teknologi Informasi ESIT*, XII(01), 8–22.
- [12]. Turnip, M., & Wijaya, C. (2016). PENERAPAN ALGORITMA C4.5 UNTUK PENENTUAN TINGKAT KONSUMSI KONSUMEN PADA MEDAN SOLUSINDO. *Jurnal Senopati (Seminar Nasional Pascasarjana Teknik Informatika) Vol*, 1(1), 34–42.