

Klasifikasi Penjualan Makanan Hewan Peliharaan di Muezza Petshop01 Bogor Menggunakan Metode Algoritma C4.5

¹*Ahmad Arif, ²Agus Yulianto

^{1,2}Universitas Nusa Mandiri
Jakarta, Indonesia

¹ahmadarifadhe@gmail.com, ²agus.aag@nusamandiri.ac.id

*Penulis Korespondensi

Diajukan : 20/10/2022

Diterima : 27/10/2022

Dipublikasi : 27/10/2022

ABSTRAK

Ketersedian data penjualan di toko *Muezza Petshop01* Bogor tidak digunakan semaksimal mungkin, sehingga data penjualan tersebut tidak dimanfaatkan secara maksimal untuk mengetahui makanan apa yang paling disukai dan digemari konsumen atau paling laris terjual. Tujuan penelitian ini yaitu untuk menerapkan metode *Algoritma C4.5* pada toko *Muezza Petshop01* Bogor dan semoga dapat menghasilkan suatu pengetahuan berupa klasifikasi penjualan makanan yang paling disukai banyak pelanggan dataset yang ada (laris dan kurang laris). Metode yang dipakai untuk membuat klasifikasi penjualan makanan yaitu *Algoritma C4.5*, prosesnya memakai aturan lima langkah dalam KDD (Knowledge Discovery in Database), yang meliputi beberapa aktivitas yaitu seleksi, praproses, transformasi, data mining, interpretasi dan evaluasi. Selain membuat perhitungan dalam bentuk manual, penelitian ini saya ujikan juga menggunakan tools RapidMiner. Dari output penelitian mencari hasil pohon keputusan menggunakan metode *Algoritma C4.5* dihasilkan nilai entrophy dan untuk nilai gain. Pada atribut jumlah terjual pada perhitungan manual. Sedangkan menggunakan tool RapidMiner dihasilkan pohon keputusan seperti pada gambar 4.11 Jumlah Terjual – Harga – Nama Makanan. Hasil dari penelitian ini adalah penjualan terlaris yaitu Beauty Chicken & Salmon Adult 1kg, Cat Choize Salmon Adult 800gr dan Cat Choize Tuna Adult 800gr, Penelitian dengan metode *Algoritma C4.5* dengan tool RapidMiner memiliki tingkat nilai Accuracy sebesar 93.58%, Percision 95.88%, recall sebesar 97.07%.

KataKunci: *Algoritma C4.5, Data Mining, RapidMiner, Klasifikasi*

I. PENDAHULUAN

Makan adalah hal yang mendasar yang pasti setiap hewan membutuhkan dalam bertahan hidup, sehingga akan terus muncul berbagai bisnis yang menawarkan jenis makanan hewan dari harga, rasa dan bentuk kemasan. Makanan kucing harus memiliki kebutuhan gizi yang cukup dan dibutuhkan oleh kucing domestik, makanan yang banyak mengandung vitamin dan asam amino. Makanan yang mengandung asam amino taurin contohnya adalah daging. Taurin adalah sejenis asam amino yang merupakan bahan penyusun semua protein. Kandungan zat ini sangat penting untuk penglihatan, pencernaan, fungsi otot jantung, dan menjaga sistem kekebalan tubuh yang sehat.

Penjualan merupakan kegiatan yang dilakukan oleh suatu perusahaan dalam mempertahankan bisnisnya untuk berkembang dan mendapatkan laba atau keuntungan yang diinginkan. (Siwalankerto dan petra, 2013)

Penjualan makanan hewan di toko *Muezza Petshop01* setiap harinya bisa terjual kurang lebih 50kg, itu terbagi kedalam jenis makanan basah, makanan kering dan bentuk kalengan. Dalam seminggu bisa sekitar 100kg terjual, tapi data penjualan tersebut hanya dicatat atau diinput saja



setiap harinya tidak dimanfaatkan untuk mencari tahu kira-kira makanan apa yang paling disukai konsumen untuk hewan peliharaannya, informasi ini bisa membantu dalam membuat strategi penjualan kedepannya agar bisa meningkatkan omset penjualan. Mungkin penjualan bisa melakukan inovasi lagi terhadap jenis-jenis produk makanan hewan peliharaan yang paling disukai konsumen atau bisa juga mendirikan beberapa cabang toko Muezza Petshop01 dengan mengutamakan makanan yang paling disukai konsumen, sehingga toko Muezza Petshop01 bisa berkembang lagi.

Untuk memecahkan masalah tersebut maka penulis melakukan penelitian dengan menggunakan penerapan teknik data mining Decision tree Algoritma C4.5 pada bisnis Muezza Petshop01 dan diharapkan bisa memberikan suatu pengetahuan berupa klasifikasi jenis makanan yang disukai banyak pelanggan dari dataset yang ada yang ada (Laris dan Kurang Laris). Sehingga kedepannya pemilik bisnis harus melakukan analisa klasifikasi atau prediksi jenis makanan mengikuti trend dan kegemaran pelanggannya.

II. STUDI LITERATUR

1. Penelitian Terdahulu

Dalam penelitian ini yang digunakan (1) Data *Mining*, (2) *Decision Tree*, (3) *Algoritma C4.5*, (4) Klasifikasi. Dari pohon keputusan yang telah terbentuk, variabel tertinggi dalam memprediksi penjualan adalah variabel penjualan via, artinya pendistribusian motor dari berbagai tempat sangat mempengaruhi terhadap penjualan motor. Harga dalam *range* Sangat Terjangkau Sekali (STS) yaitu harga dibawah Rp 17.699.000,00 masih mendominasi motor *matic*. Artinya, *matic* masih menguasai pasar penjualan. “Analisis *Algoritma C4.5* untuk memprediksinya penjualan Motor Pada PT. Capella Dinamik Nusantara Cabang Muka Kuning” (Azwanti, 2018).

Adapun Penelitian Algoritma C4.5, juga diterapkan di dalam berbagai kasus tentang penerapan *algoritma C4.5* untuk indexing data pasien dalam mendiagnosa penyakit gangguan kepribadian dengan hasil akurasi sebesar 80%. Penelitian tentang penerapan algoritma C4.5 untuk deteksi dini gangguan autisme pada anak dengan hasil akurasi sebesar 72. Penelitian tentang penerapan *data mining* untuk klasifikasi gangguan jiwa menggunakan metode *fuzzy logic* di telah dilakukan oleh [24]. Dalam penelitian ini diharapkan algoritma C4.5 dapat melakukan klasifikasi gangguan psikologis berdasarkan gejala-gejala yang telah ditentukan, sehingga dapat membantu dalam proses kegiatan diagnosis secara cepat dan tepat. “Analisis Data Hasil Diagnose Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5” (Febrian dan Sulistiani, 2021).

Begitupun dengan penelitian ini yang digunakan (1) Data *Mining*, (2) *Decision Tree*, (3) *Algoritma C4.5*, (4) Prediksi penjualan. Dari hasil perhitungan secara manual maupun dengan bantuan *tools* RapidMiner dari data training sebanyak 58 didapatkan pohon keputusan yang menciptakan rule untuk produk rokok yang termasuk ke dalam minat Laku adalah produk rokok dengan kondisi: berukuran Sedang; berukuran Kecil dan berkategori SKT; berukuran Besar dan berkategori SPM; berukuran Kecil, berkategori SKM dan bejenis mild. Berdasarkan evaluasi hasil menggunakan data testing sebanyak 37 data pada penelitian prediksi penjualan produk rokok menggunakan Algoritma C4.5. didapatkan nilai *accuracy* sebesar 92,11% dan AUC 0,878 sehingga mendapatkan tingkat akurasi *Good Classification*. Dalam penelitiannya yang berjudul “Prediksi penjualan produk rokok pada PT. Indomarco Prismatama menggunakan Algoritma C4.5” (Leonardi, Emilda, Katrin, Yulianto, 2021).

2. Algoritma C4.5

Algoritma C4.5 yaitu sebuah algoritma yang digunakan untuk membangun *decision tree* (pengambilan keputusan). *Algoritma C4.5* adalah salah satu *algoritma* induksi pohon kepuasan yaitu *ID3* (*Iterative Dichotomiser 3*). *ID3* dikembangkan oleh J. Ross Quinlan. Dalam prosedur algoritma *ID3*, input berupa sampel *training*, label *training* dan atribut. *Algoritma C4.5* merupakan pengembangan dari *ID3*. Beberapa pengembangan yang dilakukan pada C4.5 adalah sebagian antara lain bisa mengatasi *missing value*, bisa mengatasi *continue* data, dan *pruning*, “Implementasi Data Mining Untuk Pengenalan Karakteristik Transaksi Dengan Customer” (Faradillah, 2013).



Banyak *algoritma* yang dapat digunakan dalam membuat pohon keputusan, antara lain ID3, CART, dan C.45. Pada *algoritma C4.5* Jumlah data berbanding lurus dengan tingkat keakuratan, sehingga semakin banyak data maka tingkat keakuratan akan semakin tinggi. “Penerapan Algoritma C4.5 untuk Penentuan Ketersediaan Barang E-commerce” (Pritalia, 2018). Metode *algoritma C4.5* bertujuan untuk membangun pohon keputusan diantaranya (*Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C45.* (Eska, J., 2018) : (a) Tentukan atribut yang nantinya akan berperan sebagai akar pohon. (b) Kemudian, setiap nilai dibentuk sebuah cabang. (c) Bagi kasus pada sebuah cabang.

Untuk memilih atribut sebagai akar, didasarkan pada nilai *Gain* tertinggi dari atribut- atribut yang ada. Sehingga akan diperoleh nilai *gain* dari atribut yang paling tertinggi. *Gain* adalah salah satu *attribute selection measure* yang digunakan untuk memilih *test attribute* tiap *node* pada *tree*. Atribut dengan *information gain* tertinggi dipilih sebagai *test attribute* dari suatu *node*. Setelah mendapatkan nilai *Gain*, ada satu hal lagi yang perlu dilakukan perhitungan yaitu mencari nilai *Entrophy*. *Entrophy* digunakan untuk menentukan seberapa informatif sebuah input atribut untuk menghasilkan *output* atribut.

a. Entropy

Secara istilah *Entrophy* adalah perbedaan atau keberagaman. Dalam data *mining*, *Entrophy* didefinisikan sebagai suatu parameter untuk mengukur heterogenitas (keberagaman) dalam suatu himpunan data. Semakin heterogen himpunan suatu data, semakin besar pula nilai *Entrophynya* (*Data Mining Untuk Klasifikasi Dan Klasterisasi Data*. Bandung: Informatika Bandung, Suyanto, 2017)

Keterangan :

S = himpunan kasus n

n = jumlah partisi

π_i = proporsi Si terhadap S

Dimana n adalah jumlah nilai yang terdapat pada atribut target (jumlah kelas). Sedangkan P_i menyatakan porsi atau rasio antara jumlah sampel di kelas i dengan jumlah sampel pada himpunan data.

b. Gain

Secara istilah *Gain* adalah perolehan informasi. Dalam data *mining* *Gain* diartikan sebagai ukuran efektivitas suatu atribut dalam mengklasifikasikan data. Secara matematis, *gain* dirumuskan sebagai berikut :

Keterangan :

S ≡ Himpuan kasus

A = Atribut

N = Jumlah Partisi Atribut A

| Si | = Jumlah Kasus pada partisi ke-i

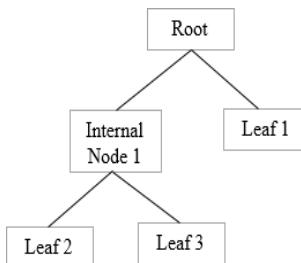
| S | = Jumlah Kasus dalam S

Dengan atribut yang nilai *gain* yang paling tertinggi dipilih sebagai atribut tes untuk sampul.

3. Pohon Keputusan (*Decision Tree*)

Pohon keputusan menggunakan representasi struktur pohon (*tree*) di mana setiap *node* mempresentasikan atribut, cabangnya mempresentasikan nilai dari atribut dan daun mempresentasikan kelas. *Node* yang paling atas dari pohon keputusan disebut sebagai root. Salah satu keuntungan yang paling signifikan dari pohon keputusan adalah kenyataan bahwa keputusan adalah kenyataan bahwa pengetahuan dapat diekstrak dan direpresentasikan dalam bentuk klasifikasi aturan *if-then*. “Optimization of C4.5 Decision Tree Algorithm for Data Mining Application.” (Agrawal and Gupta, 2013).





Gambar II. 2. Konsep Dasar Pohon Keputusan (Defiyanti & Pardede, 2010)

Pohon keputusan merupakan metode klasifikasi yang paling popular digunakan. Selain karena pembangunannya relative cepat, hasil dari model yang dibangun mudah untuk di pahami. “Penerapan Data Mining Untuk Pengolahan Data Siswa Dengan Menggunakan Metode Decision Tree” (Sijabat, 2015)

Tabel II. 1. Frekuensi Penggunaan Algoritma Pohon Keputusan

No	Algoritma Pohon Keputusan	Frekuensi
1	ID3	68 %
2	C4.5	54.55 %
3	CART	40.9 %
4	SPRINT	31.84 %
5	SLIQ	27.27 %
6	PUBLIC	13.6 %
7	C5.0	9 %
8	CLS	9 %
9	RANDOM FOREST	9 %
10	RANDOM TREE	4.5 %
11	ID3+	4.5 %
12	OCI	4.5 %
13	CLOUDS	4.5 %

Sumber : Han & Kamber, 2006

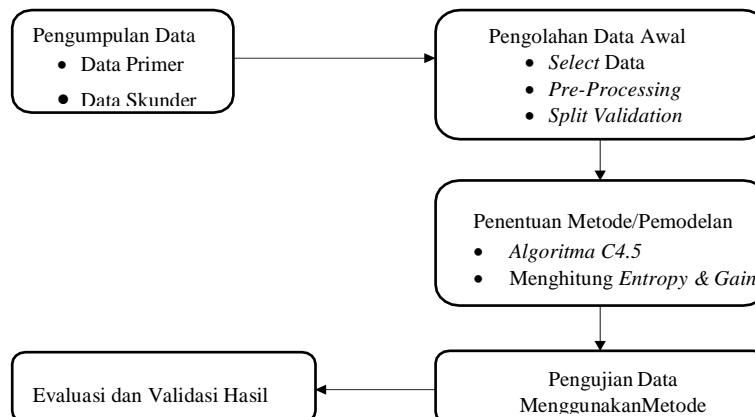
Pohon keputusan bekerja dengan membentuk pohon keputusan yang dapat disimpulkan aturan-aturan klasifikasi tertentu, salah satunya adalah C4.5 “Data Mining Untuk Menganalisa Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Metode Decision Tree C4.5,” (Untari, 2010).

III. METODE

A. Tahapan Penelitian

Metode penelitian merupakan salah satu kegiatan ilmiah untuk mendapatkan data yang valid dengan langkah-langkah yang teratur dan sistematis. Penelitian ini dilakukan untuk menemukan, mengembangkan, dan membuktikan suatu pengetahuan tertentu sehingga dapat ditarik kesimpulan berdasarkan faktor-faktor yang mempengaruhi (Sugiyono, 2016).

Pada penelitian ini, tahapan yang akan digunakan untuk melakukan klasifikasi dan prediksi terhadap penjualan makanan dapat dilihat pada gambar III.1 Tahap ini dilakukan untuk mempermudah penelitian sehingga penelitian bisa berjalan dengan baik dan sistematis, serta memenuhi tujuan yang diinginkan.

**Gambar III. 1** Tahapan Penelitian**B. Metode Penelitian****a. Pengumpulan Data**

Pada tahap ini menjelaskan tentang bagaimana dan dari mana sumber data didapatkan, diantaranya adalah :

1. Sumber Data

- Data Primer : Jurnal, buku, survei, internet, dan lain-lain. Semua sumber data diambil dari sesi *interview* atau wawancara, dan studi pustaka untuk memperkuat bahan penelitian sebagai representasi teori.

- Data Sekunder : wawancara, studi literatur dan internet, tetapi menggunakan metode pengumpulan data yang diperoleh langsung dari sumber objek penelitian. Data yang akan digunakan dalam penelitian hasil penjualan pada tanggal 3 Januari sampai tanggal 31 Februari 2022. Jumlah data yang digunakan adalah 1.862 data.

2. Variabel Data

- Variabel Input dinyatakan sebagai data sampel dari label. Berikut ini adalah variabel input yang terdapat pada data label yang akan digunakan untuk mengklasifikasi yaitu item jenis makanan, nama makanan, harga, jumlah terjual.

- Variabel Output dinyatakan sebagai hasil dari penelitian yang dilakukan terhadap variabel *input*. Sedangkan variabel target dari penelitian ini adalah atribut "Status Penjualan" yang berisikan label "Laris" dan "Kurang Laris".

b. Pengolahan Data Awal

Beberapa tahapan tersebut yaitu :

1. *Select Data* Pada tahap ini akan dilakukan pemilihan variabel data yang akan dianalisis, karena sering ditemukan bahwa tidak semua data yang dibutuhkan dengan mempertimbangkan tujuan penulisan, sehingga diperoleh beberapa variabel yang akan digunakan untuk menjadi masukan variabel *input*. 1.862 data *record* akan diambil 5 variabel, yaitu : jenis makanan, nama makanan, harga, jumlah terjual, status penjualan. Data hasil seleksi akan digunakan dalam proses data *mining*.

2. *Pre-processing*

Pada proses *pre-processing* akan dilakukan pembersihan data untuk membuang data yang *missing value* yaitu data yang tidak konsisten dan juga memperbaiki data yang rusak. Proses pembersihan data dilakukan secara manual untuk memastikan bahwa data yang telah dipilih layak untuk dilakukan proses pemodelan.

3. *Split Validation*

Data menjadi dua bagi secara acak, sebagian data *training* sebagian data *testing*. Data yang sudah disiapkan untuk klasifikasi dibagi menjadi dua untuk data *training* dan *data testing* menggunakan *sampling random* sistematis (*Systematic Random Sampling*), yakni ketika menetukan unsur pertama dari sampling yang akan diambil. Penentuan unsur *sampling* selanjutnya ditempuh dengan cara memanfaatkan interval

sampel. Interval sampel atau juga disebut sampling rasio diperoleh dengan cara membagi ukuran populasi dengan ukuran sampel yang dikehendaki (N/n). Hasil perhitungan untuk mengambil data *testing*.

$$\text{Jumlah Data (N)} = 1.862$$

$$\text{Jumlah Data Testing} = 30\% \times 1.862 = 558,6 \text{ dibulatkan jadi } 559$$

$$\text{Jumlah Sampel (n)} = 559$$

$$\text{Interval Sampling} = N/n = 1.862/559 = 3,33 \text{ dibulatkan jadi } 3$$

$$\text{Unsur pertama yang diambil untuk data testing (s)} = 1$$

$$\text{Unsur Kedua} = s + k$$

$$\text{Unsur Ketiga} = s + 2k$$

$$\text{Unsur Keempat} = s + 3k, \text{ dan seterusnya hingga unsur ke-n}$$

Pembagian data menjadi data *training* dan data *testing* pada penelitian ini menggunakan *split ratio* 70% untuk data *training* dan 30% untuk data *testing*. Dari hasil diatas diperoleh data testing sebanyak 559 data, maka sisanya dijadikan data *training* yaitu 500 data.

c. Pemodelan

Metode *Algoritma C4.5* dipilih karena salah satu kelebihannya adalah dapat menangani data numerik dan diskret. *Algoritma C4.5* menggunakan rasio perolehan (*gain ratio*). Sebelum menghitung rasio perolehan, perlu dilakukan perhitungan nilai informasi dalam satu bit dari suatu kumpulan objek, yaitu dengan menggunakan konsep *entropy* untuk membentuk pohon keputusan. Data kemudian dihitung menggunakan *algoritma* sesuai dengan metodenya kemudian dicari hasil akurasinya.

d. Pengujian Data Metode

Perhitungan yang dianalisa dan untuk melihat apakah fungsi bekerja dengan baik atau tidak. Setelah data dihitung secara manual, kemudian data diuji menggunakan *tools RapidMiner* untuk memastikan apakah hasil perhitungan manual dengan hasil yang diperoleh *RapidMiner* sesuai atau tidak.

e. Evaluasi Validasi Data

Evaluasi dapat dilakukan dengan cara mengamati dan menganalisa hasil dari algoritma yang digunakan untuk memastikan bahwa hasil pengujian itu benar atau tidak sesuai dengan pembahasan. Sedangkan validasi dilakukan dengan mengukur hasil prediksi untuk mengetahui tingkat akurasi, presisi, dan *recall*.

IV. HASILDAN PEMBAHASAN

Tahapan pengolahan data awal antara lain : *select* data, *cleaning* data, *transformation* serta akan dilakukan *spilt validation*.

a. Proses *Select* Data

Data yang sudah tersedia selanjutnya pemilihan terhadap parameter yang akan dianalisis dan akan digunakan untuk menjadi masukan atau *variable input*.

Tabel 4.1 *select* Data

No.	Atribut	Indikator	Detail Pengguna
1	Date	NO	-
2	Jenis Makanan	OK	ID
3	Nama Makanan	OK	NILAI MODEL
4	Harga	OK	NILAI MODEL
5	Jumlah Terjual	OK	NILAI MODEL
6	Status Penjualan	OK	NILAI MODEL

Tabel diatas menerangkan atribut yang akan dipakai dalam penelitian ini. Indikator "OK" menandakan atribut akan digunakan, sedangkan indikator "NO" menandakan atribut akan dieliminasi pada tahap pengolahan data awal.



b. Proses *Cleaning* Data

Pada tahap ini akan dilakukan proses pemeliharaan data untuk memastikan data yang telah dipilih tersebut telah layak untuk dilakukan proses pemodelan, memperbaiki data yang rusak, membersihkan dan menghapus data yang tidak diperlukan.

No.	Jenis Makanan	Nama Makanan	Harga	Jumlah Terjual	Status Penjualan
1	Makanan Basah	whiskas Tuna Pouch Kitten 1,2 Kg	7000	1	KURANG LARIS
2	Makanan Basah	whiskas Tuna Pouch Adult 1,2Kg	7000	1	KURANG LARIS
3	Makanan kering	Whiskas Junior Ocean Fish 1,2Kg	85000	2	LARIS
4	Makanan kering	Whiskas Adult Ocean Fish 1,2Kg	80000	2	LARIS
5	Makanan kering	Happy Cat Urinary Care 1,2Kg	130000	2	LARIS
6	Makanan kering	Excel Rasa Ayam dan Tuna 500gr	11000	4	LARIS
7	Makanan kering	Excel Rasa Ikan Tuna 500gr	11000	2	LARIS
8	Makanan kering	Felibete Adult 500gr	12500	2	LARIS
9	Makanan kering	Cat Choize Tuna Adult 800gr	20000	2	LARIS
10	Makanan kering	Cat Choize Salmon Adult 800gr	20000	5	LARIS
11	Makanan kering	Beauty Chicken & Salmon Adult 1Kg	30000	1	KURANG LARIS
12	Makanan kering	Grain Free Adult 1,2Kg	115000	1	KURANG LARIS
13	Makanan kering	Snack Temptations Tuna 85gr	30000	1	KURANG LARIS
14	Makanan kering	Snack Temptations Salmon Flavour 85gr	30000	1	KURANG LARIS
15	Makanan kering	Ori Cat Adult Cat Food 1Kg	25000	1	KURANG LARIS
16	Makanan kering	Momo Tuna Flavor 1Kg	30000	2	LARIS
17	Makanan kering	Momo Tuna Flavor 20Kg	510000	2	LARIS
-----	-----	-----	-----	-----	-----
1856	Makanan Kering	Me-O Gourmet Adult 1,2Kg	75000	2	LARIS
1857	Makanan Kering	Me-O Chicken & Vegetable Adult 1,2Kg	75000	3	LARIS
1858	Makanan Kering	Me-O Beef & Vegetable Adult 1,2Kg	75000	4	LARIS
1859	Makanan Basah	Me-O Kaleng Tuna Adult 400gr	20000	2	LARIS
1860	Makanan Basah	Life Cat Kaleng Tuna Adult 400gr	15000	2	LARIS
1861	Makanan Basah	Life Cat Kaleng Tuna Kitten 400gr	15000	1	KURANG LARIS
1862	Makanan Basah	Life Cat Kaleng Chicken Salmon Adult 400gr	15000	4	LARIS

c. Proses Transformation Data

Setelah data sudah dipilih maka akan dilakukan tahapan untuk melakukan *transformasi* terhadap atribut. Jenis atribut yang ada pada data awal penelitian ini berupa atribut *biner* (Jenis Makanan, Nama Makanan, Harga, Jumlah Terjual, Status penjualan), atribut numeric (Harga, Jumlah Terjual). Selanjutnya atribut data hasil proses *cleaning* data akan di *transformasi* ke dalam bentuk kordinat agar memudahkan proses pemodelan. Berikut adalah penjelasan proses *transformasi* atribut :

Tabel 4.3 Preoses *Transformasi* Data

Jenis Makanan	Harga	Range Harga	Jumlah Terjual	Range Jumlah	Status
A	Murah	1000 <= X < 10000	Sedikit	1 <= X < 20	Kurang Laris
B	Mahal	X >= 10000	Banyak	X >= 20	Laris

Penjelasan :

- 1000 <= X < 10000 = Maksudnya nilai HARGA kurang dari sama dengan 1000 dan X kurang dari sepuluh ribu maka HARGA ditransformasi menjadi “MURAH”.
- X >= 10000 = Nilai HARGA lebih dari sama dengan sepuluh ribu maka hasil transformasinya “MAHAL”.



Tabel diatas adalah penjelasan dari proses *transformasi* data berupa perubahan data numerik menjadi *biner* dima HARGA (MURAH, MAHAL), JUMLAH TERJUAL (SEDIKIT, BANYAK), dan STATUS (LARIS, KURANG LARIS). Berikut adalah hasil dari *transformasi* data :

Tabel 4.4 Data Hasil *Transformasi*

No.	Jenis Makanan	Nama Makanan	Harga	Jumlah Terjual	Status Penjualan
1	Makanan Basah	Whiskas Tuna Pouch Kitten 1,2 Kg	Murah	Sedikit	Kurang Laris
2	Makanan Basah	Whiskas Tuna Pouch Adult 1,2Kg	Murah	Sedikit	Kurang Laris
3	Makanan Kering	Whiskas Junior Ocean Fish 1,2Kg	Mahal	Sedikit	Kurang Laris
4	Makanan Kering	Whiskas Adult Ocean Fish 1,2Kg	Mahal	Sedikit	Kurang Laris
5	Makanan Kering	Happy Cat Urinary Care 1,2Kg	Mahal	Sedikit	Kurang Laris
6	Makanan Kering	Excel Rasa Ayam dan Tuna 500gr	Murah	Banyak	Laris
7	Makanan Kering	Excel Rasa Ikan Tuna 500gr	Murah	Banyak	Laris
8	Makanan Kering	Felibete Adult 500gr	Murah	Banyak	Laris
9	Makanan Kering	Cat Choize Tuna Adult 800gr	Murah	Banyak	Laris
10	Makanan Kering	Cat Choize Salmon Adult 800gr	Murah	Banyak	Laris
11	Makanan Kering	Beauty Chicken & Salmon Adult 1Kg	Murah	Sedikit	Kurang Laris
12	Makanan Kering	Grain Free Adult 1,2Kg	Mahal	Sedikit	Laris
13	Makanan Kering	Snack Temptations Tuna 85gr	Murah	Sedikit	Kurang Laris
14	Makanan Kering	Snack Temptations Salmon Flavour 85gr	Murah	Sedikit	Kurang Laris
15	Makanan Kering	Ori Cat Adult Cat Food 1Kg	Murah	Sedikit	Kurang Laris
16	Makanan Kering	Momo Tuna Flavor 1Kg	Murah	Sedikit	Kurang Laris
17	Makanan Kering	Momo Tuna Flavor 20Kg	Mahal	Sedikit	Kurang Laris
-----	-----	-----	-----	-----	-----
1856	Makanan Kering	Me-O Gourmet Adult 1,2Kg	Mahal	Banyak	Laris
1857	Makanan Kering	Me-O Chicken & Vegetable Adult 1,2Kg	Mahal	Banyak	Laris
1858	Makanan Kering	Me-O Beef & Vegetable Adult 1,2Kg	Mahal	Banyak	Laris
1859	Makanan Basah	Me-O Kaleng Tuna Adult 400gr	Murah	Banyak	Laris
1860	Makanan Basah	Life Cat Kaleng Tuna Adult 400gr	Murah	Banyak	Laris
1861	Makanan Basah	Life Cat Kaleng Tuna Kitten 400gr	Murah	Sedikit	Kurang Laris
1862	Makanan Basah	Life Cat Kaleng Chicken Salmon Adult 400gr	Murah	Banyak	Laris

c. Split Validation

Split Validation merupakan teknik validasi yang membagi data menjadi dua bagian, sebagian data *training* dan *data testing*. Data di klasifikasi menjadi dua menggunakan teknik sampling random untuk data *training* (70%) dan data *testing* (30%). Contoh perhitungan untuk pengambilan data *testing* adalah sebagai berikut :

$$\text{Jumlah data Keseluruhan (N)} = 1862$$

$$\text{Jumlah data testing} = 30\% \times 1862 = 558,6 \text{ dibulatkan jadi } 559$$

$$\text{Jumlah sample (n)} = 559$$

$$\text{Interval sampling (k)} = N/n = 1862/559 = 3,33 \text{ dibulatkan jadi } 3$$

$$\text{Unsur pertama yang diambil untuk data testing (s)} = 1$$

$$\text{Unsur kedua} = s+k = 1 + 3 = 4$$

$$\text{Unsur ketiga} = s+2k$$



Unsur Keempat

= s+3k, dan seterusnya

Dari hasil diatas diperoleh data *testing* sebanyak 559 data penjualan makanan, maka sisanya dijadikan data *training* sebanyak $1862 - 559 = 1303$ data penjualan makanan.

Tabel 4.6 Cuplikan Data *Testing*

No.	Jenis Makanan	Nama Makanan	Harga	Jumlah Terjual	Status Penjualan
1304	Makanan Kering	Grain Free Adult 1,2Kg	Mahal	Sedikit	Kurang Laris
1305	Makanan Kering	Snack Temptations Tuna 85gr	Murah	Sedikit	Kurang Laris
1306	Makanan Kering	Snack Temptations Salmon Flavour 85gr	Murah	Sedikit	Kurang Laris
1307	Makanan Kering	Ori Cat Adult Cat Food 1Kg	Murah	Banyak	Laris
1308	Makanan Kering	Momo Tuna Flavor 1Kg	Murah	Banyak	Laris
1309	Makanan Kering	Momo Tuna Flavor 20Kg	Mahal	Sedikit	Kurang Laris
1310	Makanan Kering	Cattie Care Tuna 22Kg	Mahal	Banyak	Laris
1311	Makanan Kering	Cattie Care Tuna 1Kg	Murah	Sedikit	Kurang Laris
1312	Makanan Kering	Bolt Adult Tuna 20Kg	Mahal	Banyak	Laris
1313	Makanan Kering	Bolt Adult Tuna 1Kg	Murah	Banyak	Laris
-----	-----	-----	-----	-----	-----
1856	Makanan Kering	Me-O Gourmet Adult 1,2Kg	Mahal	Banyak	Laris
1857	Makanan Kering	Me-O Chicken & Vegetable Adult 1,2Kg	Mahal	Banyak	Laris
1858	Makanan Kering	Me-O Beef & Vegetable Adult 1,2Kg	Mahal	Banyak	Laris
1859	Makanan Basah	Me-O Kaleng Tuna Adult 400gr	Murah	Banyak	Laris
1860	Makanan Basah	Life Cat Kaleng Tuna Adult 400gr	Murah	Banyak	Laris
1861	Makanan Basah	Life Cat Kaleng Tuna Kitten 400gr	Murah	Sedikit	Kurang Laris
1862	Makanan Basah	Life Cat Kaleng Chicken Salmon Adult 400gr	Murah	Banyak	Kurang Laris

Tabel 4.7 Cuplikan Data *Training*

No.	Jenis Makanan	Nama Makanan	Harga	Jumlah Terjual	Status Penjualan
1	Makanan Basah	Whiskas Tuna Pouch Kitten 1,2 Kg	Murah	Sedikit	Kurang Laris
2	Makanan Basah	Whiskas Tuna Pouch Adult 1,2Kg	Murah	Sedikit	Kurang Laris
3	Makanan Kering	Whiskas Junior Ocean Fish 1,2Kg	Mahal	Sedikit	Kurang Laris
4	Makanan Kering	Whiskas Adult Ocean Fish 1,2Kg	Mahal	Sedikit	Laris
5	Makanan Kering	Happy Cat Urinary Care 1,2Kg	Mahal	Sedikit	Kurang Laris
6	Makanan Kering	Excel Rasa Ayam dan Tuna 500gr	Murah	Banyak	Laris
7	Makanan Kering	Excel Rasa Ikan Tuna 500gr	Murah	Banyak	Laris
8	Makanan Kering	Felibete Adult 500gr	Murah	Banyak	Laris
9	Makanan Kering	Cat Choize Tuna Adult 800gr	Murah	Banyak	Laris
10	Makanan Kering	Cat Choize Salmon Adult 800gr	Murah	Banyak	Laris
11	Makanan Kering	Beauty Chicken & Salmon Adult 1Kg	Murah	Sedikit	Kurang Laris
12	Makanan Kering	Grain Free Adult 1,2Kg	Mahal	Sedikit	Laris
13	Makanan Kering	Snack Temptations Tuna 85gr	Murah	Sedikit	Kurang Laris
14	Makanan Kering	Snack Temptations Salmon Flavour 85gr	Murah	Sedikit	Kurang Laris
15	Makanan Kering	Ori Cat Adult Cat Food 1Kg	Murah	Sedikit	Kurang Laris



16	Makanan Kering	Momo Tuna Flavor 1Kg	Murah	Sedikit	Kurang Laris
17	Makanan Kering	Momo Tuna Flavor 20Kg	Mahal	Sedikit	Kurang Laris
-----	-----	-----	-----	-----	-----
1296	Makanan Kering	Whiskas Adult Ocean Fish 1,2Kg	Mahal	Sedikit	Kurang Laris
1297	Makanan Kering	Happy Cat Urinary Care 1,2Kg	Mahal	Banyak	Laris
1298	Makanan Kering	Excel Rasa Ayam dan Tuna 500gr	Murah	Banyak	Laris
1299	Makanan Kering	Excel Rasa Ikan Tuna 500gr	Murah	Banyak	Laris
1300	Makanan Kering	Felibete Adult 500gr	Murah	Banyak	Laris
1301	Makanan Kering	Cat Choize Tuna Adult 800gr	Murah	Sedikit	Kursng Laris
1302	Makanan Kering	Whiskas Adult Ocean Fish 1,2Kg	Mahal	Banyak	Laris

d. Pembahasan

1. Perhitungan *Pada Performance*

Pada perhitungan Performa *confusion matrix* harus ada 2 data *training* dan *testing*. Data *training* kali ini memakai 1303 data dan data *testing* memakai *sample* 559 data. Sebelum proses *Confusion matrix*, maka kita harus menentukan terlebih dahulu prediksi label. Jika sudah diketahui, maka langkah yang harus dilakukan adalah menentukan data *actual* dengan prediksi tersebut. Setelah perbandingan antar dan actual prediksi dihitung sesuai golongannya, maka langkah selanjutnya adalah proses *Confusion matrix*.

2. Perhitungan *Confusion Matrix*

Pada proses *Confusion Matrix* untuk penelitian ini menampilkan nilai *accuracy*, *precision*, dan *recall*. Dari table tersebut sebanyak 1303 data penjualan makanan, dengan hasil prediksi laris 883 data penjualan dan hasil prediksi kurang laris sebanyak 420 penjualan. Setelah dilakukan pengujian, diprediksi dengan hasil laris sebanyak 883 data dan hasil kurang laris 420 data. Maka table *confusion matrix* pada penelitian ini adalah :

Tabel 4.7 Perhitungan *Confusion Matrix*

Accuracy	100%		
	True Kurang Laris	True Laris	Class Precision
Pred. Laris	379	26	93.58%
Pred. Kurang Laris	37	861	96.39%
Class Recall	91.11%	97.07%	

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} = \frac{(883 + 258)}{(201 + 258 + 41)} = \frac{459}{1303} = 0,918 = 91,80\%$$

a. Perhitungan seharusnya (*Positive Class* : Kurang laris)

$$Precision = \frac{TP}{(TP + FN)} = \frac{201}{(201 + 41)} = \frac{201}{402} = 0,8629 = 86,29\%$$

$$Recall = \frac{TP}{(TP + FP)} = \frac{201}{(201 + 258)} = \frac{201}{459} = 0,9053 = 90,53\%$$

Setelah dilakukan seluruh tahapan evaluasi tahapan untuk *confusion matrix* di dapatkan hasil pengujian diatas, dikelompokan seperti table di bawah ini :



Tabel 4.8 Hasil Evaluasi dan Validasi

<i>Accuracy</i>	91,80%
<i>Precision</i>	86,29%
<i>Recall</i>	90,53%

3. AUC (*Area Under Curve*) *Optimistic*

Hasil perhitungan divisualisasikan kurva ROC (*Receiver operating Characteristic*) atau AUC (*Area Under Curva*).

ROC memiliki tingkat diagnose yaitu :

1. Akurasi bernilai $0.90 - 1.00 = \text{excellent classification}$
2. Akurasi bernilai $0.80 - 0.90 = \text{fair classification}$
3. Akurasi bernilai $0.70 - 0.80 = \text{poor classification}$
4. Akurasi bernilai $0.50 - 0.60 = \text{failure}$

Nilai dari kurva ROC diketahui hasil dari AUC optimistic pada gambar 4.2 diatas itu nilainya 0.948 masuk dalam range ROC nomor 1, itu berarti hasil klasifikasi penelitian ini masuk ke dalam tingkat diagnose *excellent classification*.

e. Proses Data Mining Menggunakan Algoritma C4.5

1. Hasil Perhitungan Nilai Entropy dan Gain

Berikut adalah hasil dari perhitungan *Entropy* dan *Gain* menggunakan bantuan *Microsoft Excel* agar lebih cepat dan mudah.

Tabel 4.9 Hasil perhitungan *Entropy* dan *Gain Node 1*

		Jumlah (s)	Laris (Si)	Kurang Laris (Si)	Entropy	Gain
Total		1303	888	415	0.902738166	
Jenis Makanan						0.000121725
	Makanan Basah	206	151	55	0.837112664	
	Makanan Kering	1097	735	362	0.914917059	
Nama Makanan						0.860025349
	Whiskas Tuna Pouch Kitten 1,2 Kg	35	24	11	0.898058793	
	Whiskas Tuna Pouch Adult 1,2Kg	35	26	9	0.822404226	
	Whiskas Junior Ocean Fish 1,2Kg	35	21	14	0.970950594	
	Whiskas Adult Ocean Fish 1,2Kg	35	24	11	0.898058793	
	Happy Cat Urinary Care 1,2Kg	35	20	15	0.985228136	
	Excel Rasa Ayam dan Tuna 500gr	35	31	4	0.512709142	
	Excel Rasa Ikan Tuna 500gr	35	33	2	0.315997133	
	Felibete Adult 500gr	35	21	14	0.970950594	
	Cat Choize Tuna Adult 800gr	35	30	5	0.591672779	
	Cat Choize Salmon Adult 800gr	35	26	9	0.822404226	
	Beauty Chicken & Salmon Adult 1Kg	35	26	9	0.822404226	
	Grain Free Adult 1,2Kg	34	14	20	0.977417818	
	Snack Temptations Tuna 85gr	34	28	6	0.672294817	
	Snack Temptations Salmon Flavour 85gr	34	29	5	0.602430802	
	Ori Cat Adult Cat Food 1Kg	34	29	5	0.602430802	
	Momo Tuna Flavor 1Kg	34	20	14	0.977417818	
	Momo Tuna Flavor 20Kg	34	11	23	0.908178347	
	Cattie Care Tuna 22Kg	34	15	19	0.989992792	
	Cattie Care Tuna 1Kg	34	25	9	0.833764907	
	Bolt Adult Tuna 20Kg	34	25	9	0.833764907	
	Bolt Adult Tuna 1Kg	34	25	9	0.833764907	
	Royal Canin Persian Adult 2Kg	34	8	26	0.787126586	
	Royal Canin Persian Kitten 2Kg	34	13	21	0.959686894	
	Me-O Pouch Mackarel Adult 80gr	34	25	9	0.833764907	
	Me-O Pouch Tuna Adult 80gr	34	27	7	0.733537929	
	Me-O Ocean Fish Kitten 1,2Kg	34	18	16	0.997502546	
	Me-O Tuna Adult 7kg	34	26	8	0.787126586	
	Me-O Tuna Adult 1,2kg	34	24	10	0.873981048	
	Me-O Salmon Adult 7kg	34	24	10	0.873981048	
	Me-O Seafood Adult 7kg	34	23	11	0.908178347	
	Me-O Gourmet Adult 7Kg	34	19	15	0.989992792	



	Me-O Gourmet Adult 1,2Kg	34	22	12	0.936667382	
	Me-O Chicken & Vegetable Adult 1,2Kg	34	25	9	0.833764907	
	Me-O Beef & Vegetable Adult 1,2Kg	34	28	6	0.672294817	
	Me-O Kaleng Tuna Adult 400gr	34	24	10	0.873981048	
	Life Cat Kaleng Tuna Adult 400gr	34	27	7	0.733537929	
	Life Cat Kaleng Tuna Kitten 400gr	34	27	7	0.733537929	
	Life Cat Kaleng Chicken Salmon Adult 400gr	34	25	9	0.833764907	
Harga						-32.76954642
	Mahal	615	347	268	0.988064265	
	Murah	688	530	158	0.777402706	
Jumlah Terjual						0.988064265
4	Banyak	379	26	353	0.027990489	
.	Sedikit	37	816	824	0.001199107	
1						
0						

Tabel 4.10 Hasil perhitungan Entropy dan Gain Node 1.1

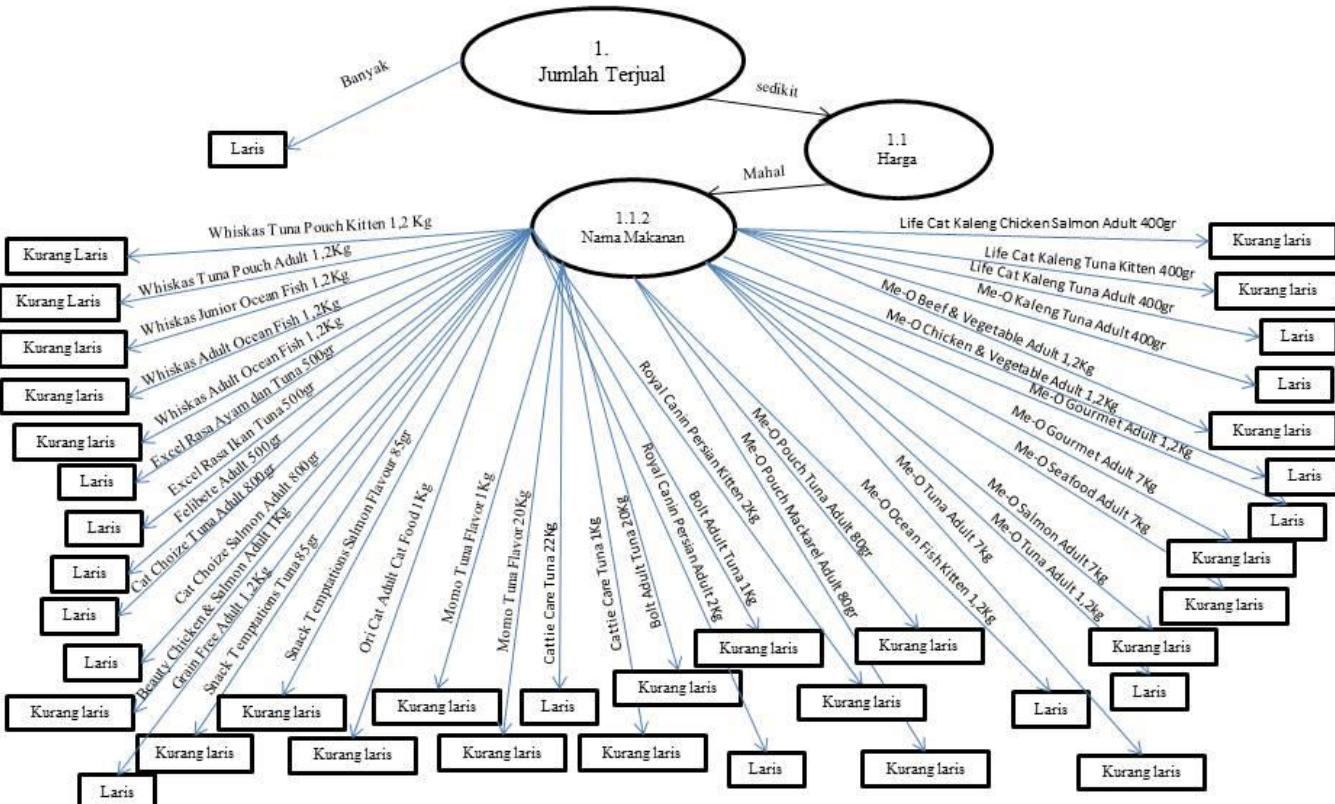
		Jumlah (s)	Laris (Si)	Kurang Laris (Si)	Entropy	Gain
Total		1303	888	415	0.902738166	
Jenis Makanan						0.000121725
	Makanan Basah	206	151	55	0.837112664	
	Makanan Kering	1097	735	362	0.914917059	
Nama Makanan						0.860025349
	Whiskas Tuna Pouch Kitten 1,2 Kg	35	24	11	0.898058793	
	Whiskas Tuna Pouch Adult 1,2Kg	35	26	9	0.822404226	
	Whiskas Junior Ocean Fish 1,2Kg	35	21	14	0.970950594	
	Whiskas Adult Ocean Fish 1,2Kg	35	24	11	0.898058793	
	Happy Cat Urinary Care 1,2Kg	35	20	15	0.985228136	
	Excel Rasa Ayam dan Tuna 500gr	35	31	4	0.512709142	
	Excel Rasa Ikan Tuna 500gr	35	33	2	0.315997133	
	Felibete Adult 500gr	35	21	14	0.970950594	
	Cat Choize Tuna Adult 800gr	35	30	5	0.591672779	
	Cat Choize Salmon Adult 800gr	35	26	9	0.822404226	
	Beauty Chicken & Salmon Adult 1Kg	35	26	9	0.822404226	
	Grain Free Adult 1,2Kg	34	14	20	0.977417818	
	Snack Temptations Tuna 85gr	34	28	6	0.672294817	
	Snack Temptations Salmon Flavour 85gr	34	29	5	0.602430802	
	Ori Cat Adult Cat Food 1Kg	34	29	5	0.602430802	
	Momo Tuna Flavor 1Kg	34	20	14	0.977417818	
	Momo Tuna Flavor 20Kg	34	11	23	0.908178347	
	Cattie Care Tuna 22Kg	34	15	19	0.989992792	
	Cattie Care Tuna 1Kg	34	25	9	0.833764907	
	Bolt Adult Tuna 20Kg	34	25	9	0.833764907	
	Bolt Adult Tuna 1Kg	34	25	9	0.833764907	
	Royal Canin Persian Adult 2Kg	34	8	26	0.787126586	
	Royal Canin Persian Kitten 2Kg	34	13	21	0.959686894	
	Me-O Pouch Mackarel Adult 80gr	34	25	9	0.833764907	
	Me-O Pouch Tuna Adult 80gr	34	27	7	0.733537929	
	Me-O Ocean Fish Kitten 1,2Kg	34	18	16	0.997502546	
	Me-O Tuna Adult 7kg	34	26	8	0.787126586	
	Me-O Tuna Adult 1,2kg	34	24	10	0.873981048	
	Me-O Salmon Adult 7kg	34	24	10	0.873981048	
	Me-O Seafood Adult 7kg	34	23	11	0.908178347	
	Me-O Gourmet Adult 7Kg	34	19	15	0.989992792	
	Me-O Gourmet Adult 1,2Kg	34	22	12	0.936667382	
	Me-O Chicken & Vegetable Adult 1,2Kg	34	25	9	0.833764907	
	Me-O Beef & Vegetable Adult 1,2Kg	34	28	6	0.672294817	
	Me-O Kaleng Tuna Adult 400gr	34	24	10	0.873981048	
	Life Cat Kaleng Tuna Adult 400gr	34	27	7	0.733537929	
	Life Cat Kaleng Tuna Kitten 400gr	34	27	7	0.733537929	
	Life Cat Kaleng Chicken Salmon Adult 400gr	34	25	9	0.833764907	
Harga						-32.76954642
	Mahal	615	347	268	0.988064265	
	Murah	688	530	158	0.777402706	0.988064265

2. Decission Tree (Pohon Keputusan)



This is a Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Dari table 4.9 dan 4.10 daitas tentang hasil perhitungan *entropy* dan *gain* maka didapat nilai *gain* tertinggi dijadikan node akar pertama. Pada table diatas akar tertinggi adalah *attribute* "Jumlah Terjual" dengan nilai 0.9880642 pada *attribute* Jumlah Terjual itu ada *variable* banyak dan sedikit. Untuk *variable* banyak nilai *entropynya* kosong maka dianggap sudah selesai, tapi untuk variabel sedikit ada nilainya maka saya cari lagi nilai *entropy* dan *gain* untuk *root node* 1.1 sampai semua *variable* memiliki kelas yang sama atau nilai *entropynya* kosong, maka pencarian *node* dinyatakan sudah selesai. Maka dapat dilihat pohon keputusan dengan *root node* seperti pada gambar dibawah ini :



Gambar 4.3 Pohon keputusan

Pada gambar diatas dapat disimpulkan bahwa pohon keputusan tersebut memiliki *root node*. Jumlah Terjual dimana kalau jumlah terjual banyak maka dikatakan laris tetapi kalau sedikit itu dilanjutkan dengan *root node 1.1*. yaitu Harga , dan dibawahnya ada *root node 1.2* yaitu Nama Menu. Hasil pohon keputusan tersebut sesuai dengan pengujian oleh *tools Rapidminer 9.10*.

3. Pengujian Hasil RapidMiner

Pada tahap ini metode data *mining* diterapkan untuk menemukan pengetahuan tersembunyi dan berharga dari data. Metode yang digunakan adalah klasifikasi dengan algoritma C4.5. berikut adalah langkah-langkah pada *tools Rapidminer* untuk mengetahui bentuk dari descision tree pada data *training*:

1. Langkah awal yaitu membuka aplikasi *Rapidminer* dan akan muncul tampilan *loading* aplikasi. 2. Setelah loading selesai akan muncul tampilan utama dari aplikasi *Rapidminer 9.10*. 3. Dari menu utama memenuhi untuk memulai proses pengolahan data klik *blank process*, dan tampilan. 4. Selanjutnya klik *add Data* untuk meng-*import* data yang akan di proses lalu klik *My Computer* dan pilih data *excel* yang akan di proses pada penyimpanan Komputer. 5. Setelah itu klik data yang akan di proses dan klik *next* sampai *finish*, jika sudah sampai pada tahap itu berarti data yang di-*import* sudah tersimpan di dalam penyimpanan dan aplikasi *RapidMiner 9.10*. 6. Selanjutnya

data yang sudah di *import* kemudian di *drag* ke dalam area proses lalu klik pencarian di area operarator untuk mencari *Set Role* dimana *Set Role* mempunyai fungsi untuk merubah parameter sebulan data, setelah itu *drag Set Role* ke area proses dan ubah parameter *attribute name* di langkah ke 3 pada gambar menggunakan atribut status penjualan dan target *reulenya* di ubah menjadi label. 7. Tahap selanjutnya kembali ke area pencarian operator untuk mencari *descision tree* lalu *drag* ke dalam area proses dan sambungkan garis *set role* dengan *descision tree* dan sambungkan kembali garis *Descision Tree* dengan titik res, selanjutnya *run* aplikasi pada langkah terakhir. Dan ini hasil *Descision Tree* dan *Description Descision Tree* untuk data *training* yang sudah di proses dengan aplikasi *RapidMiner 9.10*.

f. *Performance Vector*

Performance Vector adalah suatu cara untuk mengukur performa dari suatu data mining klasifikasi dengan *Descision Tree Algoritma C4.5*. Hasil nilai *accuracy performance vector*, *AUC (Under Area Curve)* *optimistic*, dan *deskripsi performance vector*. Dari hasil *RapidMiner* didapat nilai *Accuracy* sebesar 93.58% dengan 1030 data *training* menghasilkan prediksi *recall* 97.07% dengan prediksi Laris 95.88% sedangkan Kurang Laris 93.58%.

V. KESIMPULAN

Berdasarkan hasil penelitian Klasifikasi penjualan makanan hewan peliharaan di *Muezza Petshop01* Bogor menggunakan metode *Algoritma C4.5*, penulis dapat mengambil keputusan sebagai berikut :

1. Bawa unutk mengklasifikasikan penjualan makanan yang laris dan kurang laris menggunakan data mining dengan metode Algoritma C4.5. dari penelitian ini dapat disimpulkan penjualan makanan terlaris di *Mezza Petshop01* yaitu, Beauty Chicken & Salmon Adult 1Kg, Cat Choize Salmon Adult 800gr dan Cat Choize Tuna Adult 800gr.
2. Dari hasil pengolahan datset sebanyak 1303 data pada *RapidMiner* bahwa Algoritma C4.5 memiliki tingkat nilai *accuracy* sebesar 93.58%, *Precision* 95.58%, *recall* 97.07%.

VI. REFERENSI

- Siwalankerto dan Petra. (2013). P. T. Damai, J. M. Pemasaran., “No Title,” vol. 2, no. 1, pp. 1–15, 2013.
- A. Sijabat, (2015). “Penerapan Data Mining Untuk Pengolahan Data Siswa Dengan Menggunakan Metode Decision Tree (Studi Kasus : Yayasan Perguruan,” vol. V, pp. 7–12, 2015.
- Defiyanti, S. & Pardede, C. (2010). *Perbandingan Kinerja Algoritma ID3 dan C4.5 dalam Klasifikasi Spam-Mail*. Proceeding Seminar Ilmiah Nasional Komputer dan Sistem Intelijen (KOMMIT 2010), (ISSN: 1411-6286). <http://repository.gunadarma.ac.id/964/>
- Lukhayu Pritalia, G. (2018). Penerapan Algoritma C4.5 untuk Penentuan Ketersediaan Barang E-commerce. *Indonesian Journal of Information Systems*, 1(1), 47–56. <https://doi.org/10.24002/ijis.v1i1.1727>
- Han, J. & Kamber, M. (2006). *Data Mining Concepts and Techniques*. 2nd ed. USA: Elsevier.
- D. Untari. (2010). “Data Mining Untuk Menganalisa Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Metode Decision Tree C4.5,”
- Suyanto. (2017). *Data Mining Untuk Klasifikasi Dan Klasterisasi Data*. Bandung: Informatika Bandung.
- N. Iriadi and N. Nuraeni. (2016). “Kajian Penerapan Metode Klasifikasi Data Kelayakan Kredit Pada Bank,” *J. Tek. Komput. AMIK BSI*, vol. 2, pp. 132–137.



- S. Faradillah. (2013) "Implementasi Data Mining Untuk Pengenalan Karakteristik Transaksi Customer Dengan," pp. 63–70.
- Azwanti, N. (2018). Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada Pt. Capella Dinamik Nusantara Cabang Muka Kuning. *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, 13(1), 33. <https://doi.org/10.30872/jim.v13i1.629>
- Yulianto. A, Irena. K, Emilda. R, Leonardi. M. (2021). "Prediksi Penjualan Produk Rokok Pada PT. Indomarco Prismatama Menggunakan Algoritma C4.5" vol. 23, No.2.
- Eska, J. (2018). *Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5*. 2. <https://doi.org/10.31227/osf.io/x6svc>
- Sugiyono. (2016). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: PT Alfabet.
- G. L. Agrawal and P. H. Gupta. (2013). "Optimization of C4 . 5 Decision Tree Algorithm for Data Mining Application," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 3, no. 3, pp. 341–345.
- A. A. Aldino and H. Sulistiani. (2020). "Decision Tree C4. 5 Algorithm For Tuition Aid Grant Program Classification (Case Study: Department Of Information System, Universitas Teknokrat Indonesia)," *Edutic-Scientific J. Informatics Educ.*, vol. 7, no. 1.

