



KLASIFIKASI PENJUALAN MAKANAN CEPAT SAJI MENGGUNAKAN METODE ALGORITMA C4.5 (STUDI KASUS : AYAM PENYET NABILA)

Asep Muhidin ¹⁾, Enda Suhenda ²⁾

Program Studi Teknik Informatika Fakultas Teknik Universitas Pelita Bangsa
asep.muhammad@pelitabangsa.ac.id

Abstraksi

Ketersediaan data penjualan di gerai Ayam Penyet Nabila tidak digunakan semaksimal mungkin, sehingga data penjualan tersebut tidak dimanfaatkan secara maksimal untuk mengetahui makanan apa yang paling disukai konsumen atau paling laris terjual. Tujuan penelitian ini yaitu untuk menerapkan metode Algoritma C4.5 pada gerai Ayam Penyet Nabila dan semoga dapat menghasilkan suatu pengetahuan berupa suatu klasifikasi penjualan makanan yang paling disukai banyak pelanggan dari 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 Databases), yang meliputi beberapa aktivitas yaitu seleksi, pra-proses, transformasi, data mining, interpretasi dan evaluasi. Selain membuat perhitungan dalam bentuk manual, penelitian ini saya ujikan juga menggunakan tool RapidMiner. Dari output penelitian mencari hasil pohon keputusan menggunakan metode Algoritma C4.5 dihasilkan nilai entropi dan untuk nilai gain tertinggi yaitu 0,463363648 pada atribut Jumlah Terjual pada perhitungan manual. Sedangkan menggunakan tool RapidMiner dihasilkan pohon keputusan seperti terlihat pada Gambar 4.3 Jumlah Terjual – Harga – Nama Menu. Hasil dari penelitian ini adalah Penjualan terlaris yaitu Ayam Penyet, Ayam Bakar, Ayam Serundeng. Penelitian dengan metode Algoritma C4.5 dengan tool RapidMiner memiliki tingkat nilai accuracy sebesar 86.40%, precision 86.29%, recall sebesar 90.53%.

Kata Kunci: Algoritma C4.5, Data Mining, RapidMiner, Klasifikasi.

Abstract

The availability of sales data at the Ayam Penyet Nabila outlet is not used as much as possible, so that the sales data is not utilized maximally to find out what foods are most preferred by consumers or best-selling. The purpose of this research is to apply the C4.5 Algorithm method at the Penyet Nabila Chicken outlet and hopefully it can produce a knowledge in the form of a classification of food sales which is most preferred by many customers from existing datasets (in demand and in demand). The method used to classify food sales is C4.5 Algorithm, the process uses the five-step rule in KDD (Knowledge Discovery in Databases), which includes several activities namely selection,

preprocessing, transformation, data mining, interpretation and evaluation. Besides making calculations in the form of a manual, I also tested this research using the RapidMiner tool. From the research output looking for decision tree results using the C4.5 Algorithm method the entropy value is generated and for the highest gain value is 0.463363648 on the Sold Amount attribute in manual calculations. Whereas using the RapidMiner tool generated a decision tree as shown in Figure 4.3 Amount Sold - Price - Menu Name. The results of this study are the best-selling sales of Penyet Chicken, Grilled Chicken, Serundeng Chicken. Research using the C4.5 Algorithm method with the RapidMiner tool has an accuracy value of 86.40%, precision 86.29%, recall of 90.53%.

Keywords: *C4.5 Algorithm, Data Mining, RapidMiner, Classification.*

1. Pendahuluan

Makan adalah hal yang mendasar yang pasti setiap manusia butuhkan dalam bertahan hidup, sehingga akan terus muncul berbagai bisnis yang menawarkan jenis makanan. Namun, saat ini kita rasakan makanan telah berubah dari sekedar “kebutuhan bertahan hidup” berubah menjadi “gaya hidup”. Hal ini bisa kita lihat banyak pemilik bisnis makanan yang menjual menu yang biasa-biasa saja lalu oleh pelaku bisnis makanan dirubah menjadi makanan yang “kekinian” dan akhirnya menjadi kesukaan bagi semua kalangan usia. Bahkan gaya hidup manusia zaman sekarang yang selalu menginginkan kepraktisan dan cepat, sehingga banyak muncul usaha penjualan makanan cepat saji atau yang sering biasa kita sebut fast food. Perkembangan teknologi dan informasi pada zaman sekarang banyak melahirkan inovasi-inovasi cerdas dalam berbisnis makanan, yang dapat kita sebut kecerdasan bisnis atau business intelligence. Salah satu yang dapat kita manfaatkan adalah teknologi Data Mining dalam membuat klasifikasi Penjualan Makanan.

Ayam penyet nabila merupakan suatu usaha bisnis dalam bidang penjualan produk makanan yang baru dirintis sejak akhir tahun 2018. Masyarakat pasti setiap hari akan butuh makan, sehingga peluang bisnis makanan ini akan menjanjikan. Mayoritas orang Indonesia pasti suka pedas, dan suka daging ayam, sehingga bisnis ayam penyet ini pasti akan disukai oleh semua orang, apalagi kalau rasanya enak sesuai dengan selera konsumen. Untuk Memenuhi keinginan dan kebutuhan konsumen perlu menentukan strategi penjualan yang tepat. Ketersediaan data penjualan di gerai Ayam Penyet Nabila tidak digunakan semaksimal mungkin, sehingga data penjualan tersebut tidak dimanfaatkan secara maksimal dan belum adanya sistem pendukung keputusan dan metode yang dapat digunakan untuk merancang sebuah strategi bisnis dalam meningkatkan penjualan. Penjualan di gerai Ayam Penyet Nabila setiap harinya bisa terjual kurang lebih sekitar 200 porsi, itu terbagi kedalam 15 menu ada 10 jenis makanan ada 5 jenis minuman. Dalam seminggu bisa sekitar 1300 porsi terjual, tapi data penjualan tersebut hanya dicatat atau diinput saja setiap harinya tidak dimanfaatkan untuk mencari tahu kira-kira menu apa yang paling disukai konsumen, sehingga apabila kita tahu menu apa yang paling disukai konsumen, informasi ini bisa membantu dalam membuat strategi penjualan kedepannya agar bisa meningkatkan omset penjualan. Mungkin penjual bisa melakukan inovasi lagi terhadap menu-menu yang paling disukai konsumen atau bisa juga mendirikan beberapa cabang ayam penyet nabila

dengan mengutamakan menu yang paling disukai konsumen, sehingga gerai ayam penyet nabila bisa lebih berkembang lagi. Dari beberapa uraian di atas timbul beberapa masalah ketika pemilik Ayam Penyet Nabila sulit untuk menentukan jenis makanan yang paling banyak digemari oleh konsumen. Untuk memecahkan masalah tersebut maka peneliti melakukan penelitian dengan menggunakan penerapan teknik data mining *Decision tree Algoritma C4.5* pada bisnis rumah makan cepat saji yaitu Ayam Penyet Nabila dan diharapkan bisa memberikan suatu pengetahuan berupa suatu klasifikasi menu makanan yang disukai banyak pelanggan dari dataset yang ada (Laris dan Kurang Laris). Sehingga kedepannya pemilik bisnis harus melakukan analisa klasifikasi atau prediksi menu makanan mengikuti *trend* dan kegemaran pelanggannya. Untuk penelitian tersebut saya beri judul “**Klasifikasi Penjualan Makanan Cepat Saji Menggunakan Metode Algoritma C4.5 (Studi Kasus : Ayam Penyet Nabila).**”

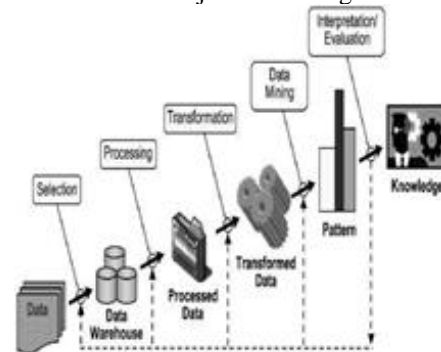
2. Tinjauan Pemikiran

2.1. Data Mining

Data Mining adalah proses pengambilan pola menggunakan teknik statistik, matematika, machine learning pada data yang nantinya akan kita proses lalu output yang muncul berupa informasi yang sangat penting. Jadi apa itu data? Data merupakan kumpulan fakta yang dapat memberikan suatu gambaran, jadi setiap kita melakukan sesuatu dalam pengambilan data maka data tersebut tersimpan dan pola-pola data itu akan diteliti secara manual sehingga kita dapat mengetahui hal-hal yang akan terjadi [8].

2.2. Tahapan Data Mining

Tahapan yang dilakukan dalam proses data mining diawali dari seleksi data dari data sumber ke data target, tahap Pra-processing untuk memperbaiki kualitas data, transformasi, data mining serta tahap interpretasi dan evaluasi yang menghasilkan keluaran berupa informasi yang diharapkan memberikan kontribusi yang lebih baik. Secara detail dijelaskan sebagai berikut [9]:



Gambar 1. Tahapan Data Mining [9]

2.3. Klasifikasi

klasifikasi adalah proses pencarian sekumpulan model yang menggambarkan dan membedakan kelas data dengan tujuan agar model tersebut dapat digunakan untuk memprediksi kelas dari kelasnya [10].

2.4. Decision Tree

Decision tree atau pohon keputusan adalah alat pendukung keputusan yang menggunakan model keputusan yang berbentuk seperti pohon. Decision tree memetakan berbagai alternatif yang mungkin untuk mengatasi suatu masalah, dan terdapat juga faktor-faktor kemungkinan yang dapat mempengaruhi alternatif tersebut beserta estimasi akhirnya jika memilih alternatif yang ada. Decision tree merupakan salah satu metode yang bisa digunakan untuk menampilkan algoritma dimana hanya berisi pernyataan kontrol bersyarat [11].

2.5. Algoritma C4.5

Algoritma C4.5 adalah algoritma klasifikasi data dengan teknik pohon keputusan yang memiliki kelebihan-kelebihan. Kelebihan ini misalnya dapat mengolah data numerik (continue) dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma yang lain [1].

2.6. Entropy

Secara istilah Entropy adalah perbedaan atau keberagaman. Dalam data mining, Entropy didefinisikan sebagai suatu parameter untuk mengukur heterogenitas (keberagaman) dalam suatu himpunan data. Semakin heterogen himpunan suatu data, semakin besar pula nilai Entropinya [13]. Secara matematis Entropy dirumuskan sebagai berikut :

Keterangan:

S : Himpunan kasus

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

A : Atribut

N : Jumlah partisi S

Pi : Proporsi dari Si terhadap S

2.7. 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:

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

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

3. Desain Penelitian/Metodologi

3.1 Objek Penelitian

Pada penelitian ini peneliti mengambil objek penelitian yang berlokasi di daerah Pasir Konci Gemal Apik, Kecamatan Cikarang Selatan, Kabupaten Bekasi dengan mengambil data penjualan bisnis makanan cepat saji yaitu rumah makan Ayam Penyet Nabila milik bapak Suryanto.

3.2 Tahapan Penelitian

Pada penelitian ini, tahapan yang akan digunakan untuk melakukan klasifikasi dan prediksi terhadap penjualan makanan dapat dilihat pada gambar 2. Tahap ini dilakukan untuk mempermudah penelitian sehingga penelitian bisa berjalan dengan baik dan sistematis, serta memenuhi tujuan yang diinginkan. Berikut adalah waktu dan langkah-langkah dalam tahapan yang dilakukan :

Tabel 1. Time Line Penelitian

No	Aktivitas	2019												2020							
		September				Oktober				November				Desember				Januari			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Konsultasi Dosen pembimbing	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
2	Studi Pendahuluan	█	█																		
3	Pengumpulan Data Di Gerai Ayam Penyet Nabila	█	█																		
4	Pengolahan Data Awal		█	█																	
5	Penentuan Metode Pemodelan			█	█																
6	Pengujian Menggunakan Metode				█																
7	Evaluasi Dan Validasi Hasil					█	█														
8	Penyusunan Skripsi	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
9	Sidang Skripsi																				█

3.3 Pengumpulan Data

a. Sumber Data

•Data Primer

Data-data yang peneliti cantumkan disini merupakan data-data yang terdapat dari berbagai macam media yaitu : jurnal, buku, survei, internet, dan lain-lain.

•Data Sekunder

Data yang digunakan dalam penelitian ini adalah data sekunder, data yang peneliti dapatkan adalah data laporan hasil penjualan rumah makan ayam penyat nabila yang akan dijadikan sebagai bahan penelitian. Data yang akan digunakan dalam penelitian ini adalah data laporan hasil penjualan dari tanggal 1 Februari sampai tanggal 20 Maret 2019. Jumlah data yang digunakan adalah 715 data.

b. Sumber Data

Variabel yang terdapat pada data penelitian ini meliputi variabel input dan variabel output atau target.

- Variabel *input*

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 menu, nama menu, harga, jumlah terjual.

- Variabel Target

Variabel target 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".

3.4 Pengolahan Data Awal

a. 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 penelitian, sehingga diperoleh beberapa variabel yang akan digunakan untuk menjadi masukan variabel *input*.

b. 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.

c. Split Validation

Split Validation merupakan teknik validasi yang membagi data menjadi dua bagi secara acak, sebagian data *training* sebagian data *testing*. Hasil perhitungan untuk mengambil data *testing* adalah sebagai berikut :
Jumlah Data (N) = 715

(s)

= 1

Unsur Kedua = s + k

Unsur Ketiga = s + 2k

Unsur Keempat = s + 3k, 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 215 data, maka sisanya dijadikan data *training* yaitu 500 data.

3.5 Pemodelan

Pada penelitian ini akan dilakukan pemodelan menggunakan metode *Algoritma C4.5* untuk pengolahan data penjualan makanan. Berikut ini tahapan proses pemodelan dalam penelitian ini. Metode *Algoritma C4.5* dipilih karena salah satu kelebihanannya 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 akurasi. Ada beberapa tahap dalam membentuk pohon keputusan dengan *algoritma C4.5* antara lain :

1. Menyiapkan data training, dimana data tersebut akan diklasifikasikan.
2. Menentukan akar dari pohon, akar akan diperoleh dari atribut yang terpilih dengan cara menghitung nilai gain dari masing-masing atribut. Nilai gain tertinggi akan dijadikan akar pertama dalam pohon keputusan. Sebelum menghitung nilai gain, hitung dulu nilai *entropy* dengan persamaan

- (2.3)
- 3. Kemudian hitung nilai gain dengan persamaan (2.4)
- 4. Untuk langkah 2 hingga semua record terpartisi.
- 5. Proses partisi akan berhenti saat : Semua record pada simpul N mendapat kelas yang sama. Tidak ada atribut didalam record yang akan dipartisi lagi. Tidak ada record didalam cabang yang kosong.

3.6 Pengujian Data Metode

Pengujian metode dilakukan untuk mengetahui hasil 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.

3.7 Evaluasi Validasi Hasil

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

4. Hasil Dan Pembahasan

4.1 Select Data

Data yang sudah tersedia selanjutnya pemilihan terhadap parameter yang akan dianalisis. Parameter yang diambil adalah atribut dari data penjualan makanan yang telah didapatkan sebelumnya dari gerai Ayam Penyet Nabila, yang akan digunakan untuk menjadi masukan atau variable input.

Tabel 2. Select Data

No	Atribut	Indikator	Detail Pengguna
1	Date	NO	-
4	Jenis Menu	OK	ID
5	Nama Menu	OK	NILAI MODEL
6	Harga	OK	NILAI MODEL
7	Jumlah	OK	NILAI MODEL

Tabel diatas menerangkan atribut yang akan dipakai dalam penelitian ini. Indikator “OK” menandakan atribut akan digunakan, sedangkan

indikator “NO” menandakan atribut tersebut akan dieliminasi pada tahap pengolahan data awal.

4.2 Proses Cleaning Data

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

4.3 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 Menu, Nama Menu, Harga, Jumlah Terjual, Status Penjualan), atribut numerik (Harga, Jumlah Terjual). Selanjutnya atribut data hasil proses cleaning data akan di transformasi ke dalam bentuk kordinal agar memudahkan proses pemodelan. Berikut adalah penjelasan proses transformasi atribut :

Tabel 3. Proses Transformasi Data

Jenis Menu	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”.
- Begitu pula dengan JUMLAH TERJUAL

4.4 Split Validation

Split Validation merupakan teknik validasi yang membagi data menjadi dua bagian, sebagian data *training* dan sebagian *data testing*. Data yang sudah disiapkan untuk klasifikasi dibagi menjadi dua menggunakan teknik sampling random untuk data training (70%) dan data

testing (30%).

4.5 Perhitungan Pada Performa

Pada perhitungan Performa *confusion matrix* harus ada 2 data yaitu data *training* dan data *testing*. Data *training* pada kali ini memakai 500 data dan data *testing* memakai *sample* 215 data. Sebelum proses *confusion matrix*, maka kita harus menentukan terlebih dahulu prediksi label. Jika sudah diketahui, maka langkah yang harus dilakukan adalah menentukan data aktual dengan data prediksi tersebut



Gambar 2. Perbandingan Antara Data Aktual Dan Data Prediksi

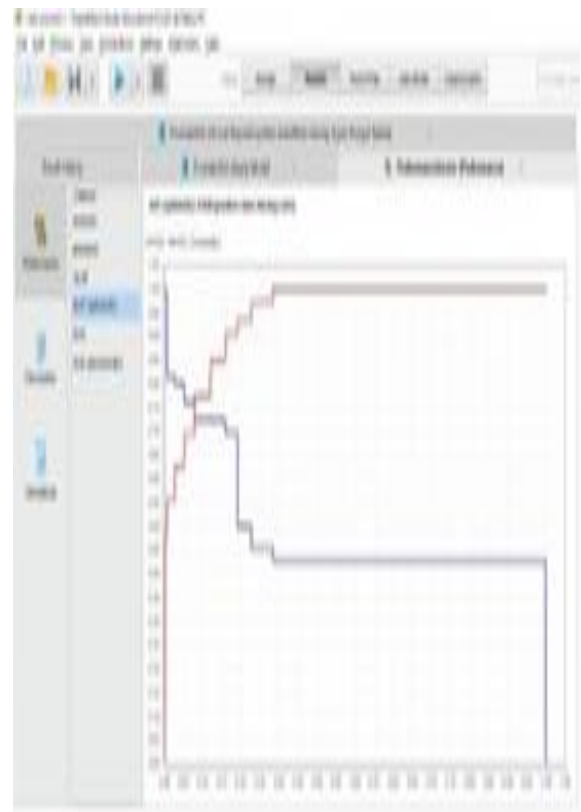
4.6 Perhitungan Confusion Matrix

Pada proses *confusion Matrix* untuk penelitian ini menampilkan nilai *accuracy*, *precision*, dan *recall*. Dari tabel tersebut sebanyak 500 data penjualan makanan, dengan hasil prediksi laris 201 data penjualan dan hasil prediksi kurang laris sebanyak 299 penjualan

Tabel 5. Perhitungan *Confusion Matrix*

Accuracy	100%		
	True Laris	True Kurang Laris	Class Precision
Pred. Laris	174	27	86.57%
Pred. Kurang Laris	41	258	86.29%
Class Recall	80.93%	90.53%	

4.7 AUC (Area Under Curve) Optimistic
 Hasil perhitungan divisualisasikan dengan kurva ROC (*Receiver Operating Characteristic*) atau AUC (*Area Under Curve*).



Gambar 3. Hasil AUC (Area Under Curve)

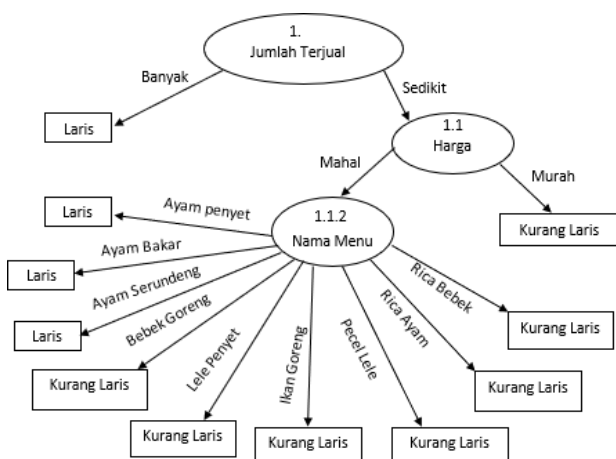
ROC memiliki tingkat diagnosa yaitu :

1. Akurasi bernilai 0.90 – 1.00 = *excellent classification*
2. Akurasi bernilai 0.80 – 0.90 = *fair classification*
3. Akurasi bernilai 0.70 – 0.80 = *poor classification*
4. Akurasi bernilai 0.50 – 0.60 = *failure*

Nilai dari kurva ROC (Receiver Operating Characteristic) diketahui hasil dari AUC (Area Under Curve) *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 diagnosa *excellent classification*.

4.8 Hasil Perhitungan Nilai Entrophy Dan Gain

Berikut adalah hasil dari perhitungan nilai entrophy dan Gain menggunakan bantuan dari Microsoft excel agar lebih cepat dan mudah



Gambar 4. 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

5. Kesimpulan

Berdasarkan hasil penelitian Klasifikasi penjualan makanan dengan Algoritma C4.5 pada gerai Ayam Penyet Nabila, peneliti dapat mengambil kesimpulan sebagai berikut :

1. Bahwa cara untuk mengklasifikasikan penjualan makanan yang laris dan kurang laris yaitu menggunakan data mining dengan metode Algoritma C4.5. Dari penelitian diatas dapat disimpulkan pula penjualan makanan terlaris pada gerai Ayam Penyet Nabila yaitu Ayam Bakar, Ayam Penyet, Ayam Serundeng.
2. Berdasarkan dari hasil pengolahan dataset sebanyak 500 data pada Rapid miner bahwa Algoritma C4.5 memiliki tingkat nilai accuracy sebesar 86.40% , precision 86.29% , recall sebesar 90.53%.

Daftar Pustaka

- [1] Kusrini & Emha Taufiq Luthfi. 2009. Algoritma Data Mining. CV ANDI OFFSET: Yogyakarta.
- [2] Kamadi, David Hartanto & Hangsun, Sen. 2014. Implementasi Data Mining Dengan Algoritma C.4.5 untuk Memprediksi Kelulusan Mahasiswa : Pelita Informatika
- [3] Sularno & Putri angraeni . 2017. Penerapan Algoritma C4.5 Untuk Klasifikasi Tingkat Keganasan hama pada tanaman Padi: Kerinci
- [4] Santoso, Teguh Budi. 2016. prediksi loyalitas pelanggan data seluler menggunakan metode klasifikasi dengan Algoritma C4.5.
- [5] Iskandar, Isa., Hiryanto, Leli., Hedryli, Janson. 2016. Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5 Dengan Teknik Prunning : Yogyakarta.
- [6] Nurhandayani, Nova. 2018. Penerapan Algoritma C4.5 Untuk Menentukan Data Stok dan Target Permintaan Material Yang Paling Dibutuhkan Gudang Logistik Pada PT. PLN (Persero). Jakarta.
- [7] Larose, Daniel T. 2005. Discovering Knowledge in Data : An Introduction to Data Mining. John Willey & Sons, Inc.
- [8] Turban, E. 2005. Decision Support Systems and Intelligent Systems Edisi Bahasa Indonesia Jilid 1. Andi: Yogyakarta.
- [9] Ayyad, Usama. 1996. Advances in Knowledge Discovery and Data Mining. MIT Press.
- [10] Herman, Aldino dan Naam, Julfriadi . 2012. Penerapan Kata Kunci Dokumen Teks Surat Kabar Singgalang Dengan Menggunakan Metoda Algoritma Naive Bayes. Singgalang.
- [11] Berry, Michael J., and Gordon Linoff. 1997. "Data Mining Techniques: For Marketing, Sales, and Customer Support." John Wiley & Sons, Inc.
- [12] Wibowo, Ari. 2011. Prediksi Nasabah

- Potensial Menggunakan Metode Klasifikasi Pohon Biner: Universitas Politeknik Negri Batam.
- [13] Suyanto. 2017. Data Mining Untuk Klasifikasi Dan Klasterisasi Data. Bandung: Informatika Bandung.
- [14] 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/>
- [15] Gorunescu, Florin. 2011. Data Mining: Concepts, Models, and Techniques. Verlag Berlin Heidelberg : Springer