



Penerapan Data Mining Untuk Memprediksi Penjualan Kue Brownis Menggunakan Algoritma C4.5 Pada Toko Amanda Brownis

Andhika Bayu Prasetyo¹, Ines Heidiani Ikasari²

^{1,2}Ilmu Komputer, Teknik Infomatika, Universitas Pamulang, Tangerang Selatan, Indonesia
Email: andhikabayup585@gmail.com

Abstrak – Algoritma C4.5 merupakan algoritma yang digunakan dalam membangun pohon keputusan. Algoritma C4.5 termasuk metode dalam data mining yang merupakan proses pencarian pola dengan memilih dari sejumlah besar data menggunakan teknologi pengenalan pola. Penambangan data adalah istilah yang digunakan untuk menggambarkan temuan dalam basis data. Penambangan data adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengidentifikasi informasi. Penelitian ini membahas tentang penerapan data mining untuk memprediksi penjualan brownies dengan menggunakan algoritma C4.5. Tujuan dari penelitian ini agar stok penjualan brownies dapat sesuai dengan permintaan dari konsumen. Pada penelitian ini dapat disimpulkan bahwa dengan menggunakan algoritma C4.5 penjualan brownies dapat diprediksi dengan hasil yang relatif tinggi. Hasil perhitungan Accuracy yang didapat dari data training menggunakan Confusion Matrix mendapatkan hasil 78,67%, sedangkan Precision dan Recall mendapatkan hasil 65,22%. jika dilihat menggunakan kurva ROC hasilnya adalah Excellent Classification yaitu sebesar 0,901.

Kata Kunci: Algoritma C4.5, Data Mining, Prediksi

Abstract – Algorithm C4.5 is an algorithm used in building decision trees. The C4.5 algorithm includes a method in data mining which is the process of finding patterns by selecting from a large amount of data using pattern recognition technology. Data mining is a term used to describe findings in databases. Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to identify information. This study discusses the application of data mining to predict brownie sales using the C4.5 algorithm. The purpose of this research is that the stock for selling brownies can match the demand from consumers. In this study, it can be concluded that by using the C4.5 algorithm brownie sales can be predicted with relatively high results. Accuracy calculation results obtained from training data using the Confusion Matrix get 78.67% results, while Precision and Recall get 65.22% results. when viewed using the ROC curve the result is Excellent Classification which is equal to 0.901

.Keywords: C4.5 Algorithm, Data Mining, Prediction

1. PENDAHULUAN

Perkembangan ilmu teknologi yang semakin pesat, seiring dengan perkembangan zaman membuat teknologi semakin canggih dan berperan penting dalam setiap aspek kehidupan manusia. Mampu menyediakan segala kebutuhan informasi dan mempermudah segala pekerjaan yang dilakukan oleh setiap manusia, mampu menyelesaikan masalah setiap permasalahan yang ada terutama yang berada di perusahaan – perusahaan yang sedang berkembang pada saat ini.

Di zaman sekarang hampir semua toko *online* maupun *offline* menggunakan sistem komputerisasi dalam penyimpanan data penjualan, sehingga banyak data transaksi yang akan dihasilkan. Data transaksi yang ada dalam *database* penjualan barang menyimpan jumlah *record* transaksi penjualan yang memiliki volume yang sangat besar sehingga menyebabkan jumlah data terus bertambah di setiap harinya. Dari penumpukan data yang terjadi dapat digali untuk menentukan pola – pola penjualan barang yang dapat digunakan untuk menganalisis pasar dan meramalkan penjualan pada waktu yang akan datang.

Prediksi penjualan merupakan faktor penting untuk menentukan kelancaran usaha suatu perusahaan. Prediksi ini sangat berguna untuk menentukan berapa banyak yang akan dipesan pada bulan berikutnya. Permasalahan yang umum dihadapi oleh suatu perusahaan adalah bagaimana memprediksi atau meramalkan penjualan barang di masa mendatang berdasarkan data penjualan



sebelumnya. Prediksi tersebut sangat berpengaruh untuk menentukan target penjualan yang harus dicapai.

Amanda Brownis adalah usaha yang didirikan oleh Bapak Gatot yang bergerak di bidang makanan, khususnya kue – kue kering salah satunya adalah kue brownis. Dengan banyaknya permintaan kue brownis dari konsumen, oleh karena itu toko Amanda Brownis sangat membutuhkan prediksi penjualan yang bertujuan untuk menentukan kue brownis mana yang paling diminati oleh konsumen atau masyarakat.

Dengan menggunakan Data Mining dan menggunakan algoritma C4.5 untuk memprediksi penjualan kue brownis diharapkan dapat membantu toko Amanda Brownis untuk mengetahui kue mana yang paling diminati oleh konsumen dari data penjualan sebelumnya, sehingga penjualan lebih efektif.

Berdasarkan latar belakang yang telah diuraikan di atas, maka penulis mengambil judul yaitu **“Penerapan Data Mining Untuk Prediksi Penjualan Kue Brownis Menggunakan Algoritma C4.5 Pada Toko Amanda Brownis”**.

2. METODE

Alur penelitian dimulai dengan mengumpulkan data terlebih untuk mendapatkan data penjualan kue brownis, setelah itu input data training, kemudian dari atribut itu dihitung nilai *entropy* nya setelah itu hitung nilai *entropy* dan *gain* pada setiap atributnya. Berikut perhitungan nilai *entropy* dan *gain* :

Entropy :

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

Dimana :

- S = Himpunan Kasus
- n = Jumlah Partisi S
- p_i = Proporsi Himpunan Kasus ke – i

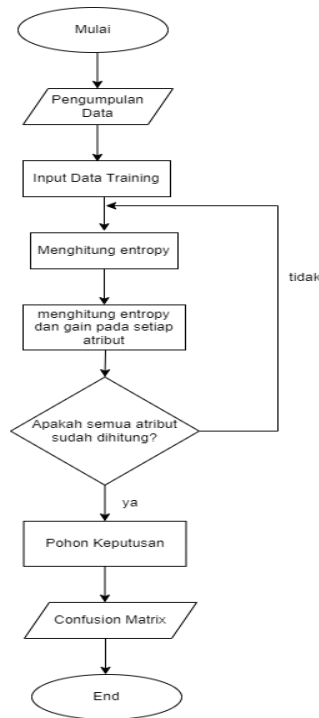
Gain :

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

Dimana :

- S = Himpunan Kasus
- n = Jumlah Partisi S
- A = Atribut
- S_i = Jumlah kasus atribut A yang memiliki partisi ke-i

Jika sudah dihitung semua nilai *entropy* dan *gain* pada setiap atribut lalu dibuatlah pohon keputusannya, namun jika belum ada atribut yang belum dihitung maka hitung kembali nilai *entropy* dan *gain* pada setiap atribut hingga selesai. Lalu didapatkan hasil *Confusion Matrix* berupa *Accuracy*, *Precision*, *Recall* dan juga AUC. Dapat dilihat pada gambar 1 berikut ini.



Gambar 1. Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

Pada bagian ini berisi hasil dari kegiatan penelitian yang sudah dilakukan

3.1 Hasil

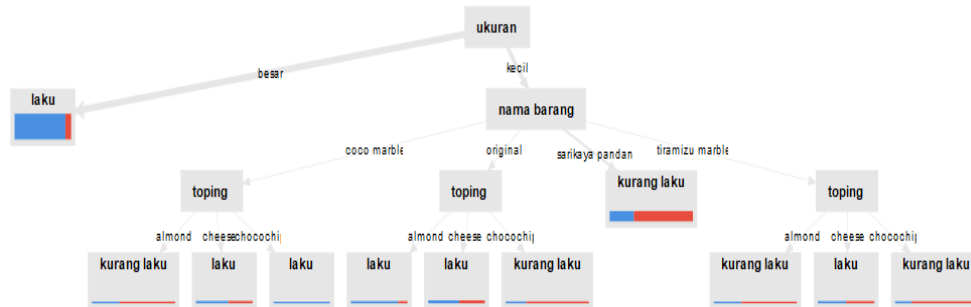
Untuk memenuhi permintaan dari konsumen yang tidak bisa ditebak maka dilakukan pengelompokan kue brownis yang sering dibeli oleh konsumen dengan melihat penjualan di tahun 2019. 250 data dengan atribut nama barang, ukuran, *topping*, warna, status.

Dari tabel data *training* yang sudah di-*input* berjumlah sebanyak 250 data, diperoleh atribut status dengan keterangan laku berjumlah 172 kue brownis sedangkan untuk keterangan tidak laku berjumlah 77 kue brownis. Dan didapatkan nilai *entropy* sebesar 0,8923. Sedangkan untuk atribut dengan nilai tertinggi yaitu atribut ukuran dengan jumlah total *gain* berjumlah 0.2158 . Perhatikan pada tabel 2 berikut :

Tabel 1. Perhitungan Entropy dan Gain

Atribut	Value	Jml Kasus (S)	Laku (Si)	Kurang Laku (Si)	Entropy	Gain Total
Total		249	172	77	0.8923	
Nama Barang						0.0494
	original	54	42	12	0.7642	
	coco marble	52	43	11	0.7008	
	tiramizu marble	52	39	12	0.7995	
	sarikaya pandan	91	49	42	0.9957	
Ukuran						0.2158
	besar	147	132	15	0.4754	
	kecil	102	40	62	0.9662	
Topping						0.0003
	almond	108	75	33	0.8880	
	cheese	55	37	18	0.9121	
	chocochip	86	60	26	0.8841	
Warna						0.0362
	coklat	160	123	37	0.7802	
	hijau	89	49	40	0.9926	

3.2 Pembahasan



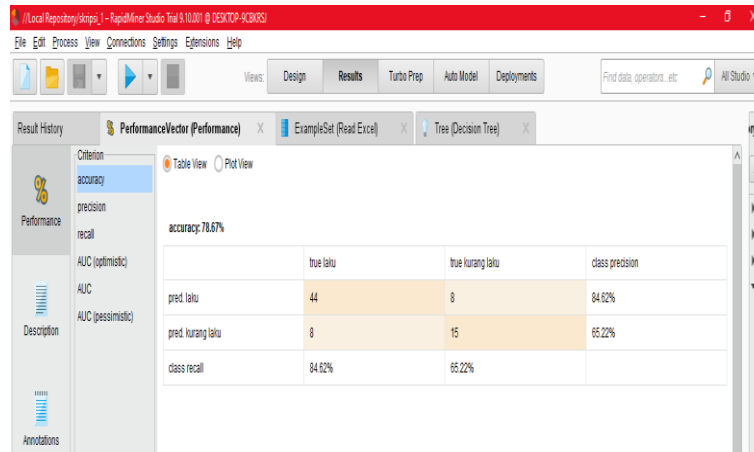
Gambar 2. *Decision Tree*

Berdasarkan hasil dari *decision tree* di atas, maka dapat disimpulkan dengan beberapa kondisi yang terjadi, yaitu sebagai berikut :

- Brownis dengan ukuran besar memiliki jumlah 132 laku dan 15 kurang laku, jadi brownis dengan ukuran yang besar masuk di kategori laku.
- Brownis bentuk kecil rasa *coco marble* dengan *topping* almond memiliki perbandingan laku dengan yang kurang laku berjumlah 1:2 dan dikategorikan ke dalam brownis yang kurang laku. Kemudian untuk yang memiliki *topping* *cheese*, memiliki perbandingan yang laku dengan yang kurang laku berjumlah 4:3 dan masuk ke kategori brownis yang laku. Sedangkan untuk yang memiliki *topping* *chocochip* ini memiliki jumlah perbandingan 2 yang laku dan 0 yang kurang laku, dan termasuk ke kategori brownis yang laku.
- Brownis bentuk kecil rasa original dengan top *topping* ing almond masuk ke dalam kategori brownis yang laku karena memiliki jumlah 5 untuk laku dan 1 untuk yang kurang laku. Kemudian untuk yang mempunyai *topping* *cheese* memiliki perbandingan 6:5 untuk kategori yang laku dan yang kurang laku, maka untuk brownis rasa original *topping* *cheese* masuk ke dalam kategori laku. Sedangkan untuk yang memiliki *topping* *chocochip* masuk ke kategori kurang laku, karena memiliki perbandingan 1:3 untuk laku dan yang kurang laku.
- Brownis bentuk kecil rasa sarikaya pandan memiliki perbandingan untuk semua rasa yaitu berjumlah 16:39 untuk yang laku dan yang kurang laku, maka untuk rasa sarikaya pandan masuk ke dalam kategori kurang laku.
- Brownis bentuk kecil rasa tiramizu marble dengan *topping* almond memiliki perbandingan laku dengan yang kurang laku berjumlah 1:2 dan masuk ke dalam kategori brownis yang kurang laku. Kemudian untuk yang memiliki *topping* *cheese*, memiliki perbandingan yang laku dengan yang kurang laku berjumlah 3:3 dan masuk ke dalam kategori brownis yang laku. Sedangkan untuk yang memiliki *topping* *chocochip* mempunyai jumlah perbandingan 1 untuk yang laku dan 4 untuk yang kurang laku, dan termasuk ke dalam kategori brownis yang kurang laku.

3.2.1. Confusion Matrix

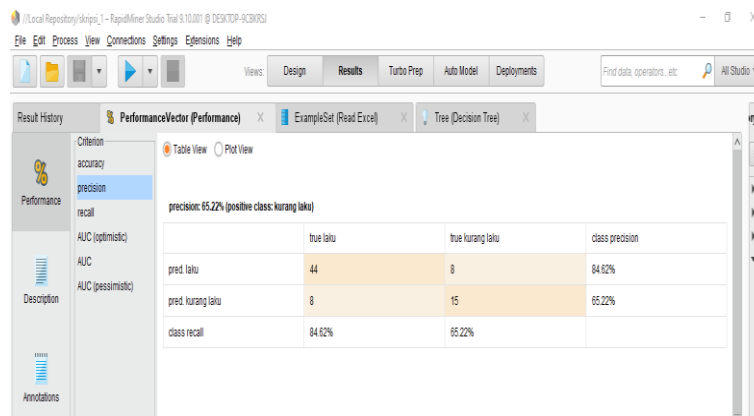
a. Accuracy



Gambar 3. Tampilan Accuracy

Hasil pengukuran akurasi data yang didapatkan dari data *training* yaitu berjumlah 78.67%. dari gambar di atas bisa dilihat bahwa untuk prediksi laku dengan *true* laku berjumlah 44 barang dan *true* yang kurang laku berjumlah 8 barang, dan dengan hasil pencapaian dari nilai presisi berjumlah 84.62%. sedangkan untuk dari prediksi yang kurang laku untuk yang *true* laku berjumlah 8 barang dan sedangkan yang *true* kurang laku berjumlah 15 barang, dan dengan hasil dari nilai presisi berjumlah 65.22%. untuk nilai dari *recall* tertinggi terdapat di *class recall true* laku yaitu mencapai 84.62%, sedangkan untuk nilai *true* kurang laku mencapai 65.22%.

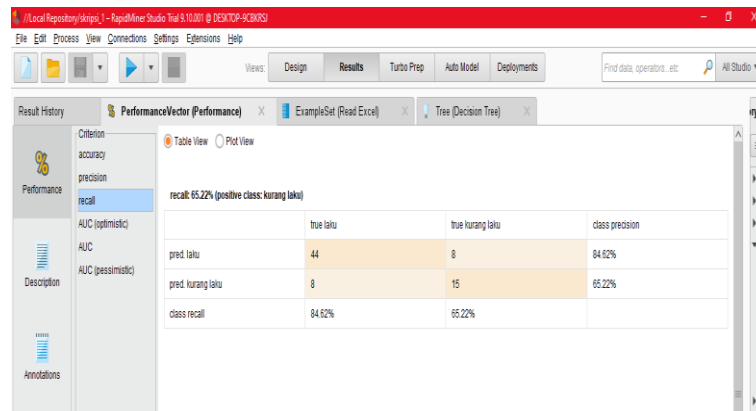
b. Precision



Gambar 4. Tampilan Precision

Hasil pengukuran *precision* data yang didapatkan dari data *training* yaitu berjumlah 65.2% (*positive class* : kurang laku). dari gambar tersebut bisa dilihat bahwa untuk prediksi laku dengan *true* laku berjumlah 44 barang dan *true* yang kurang laku berjumlah 8 barang, dan dengan hasil pencapaian dari nilai presisi berjumlah 84.62%. sedangkan untuk dari prediksi yang kurang laku untuk yang *true* laku berjumlah 8 barang dan sedangkan yang *true* kurang laku berjumlah 15 barang, dan dengan hasil dari nilai presisi berjumlah 65.22%. untuk nilai dari *recall* tertinggi terdapat di *class recall true* laku yaitu mencapai 84.62%, sedangkan untuk nilai *true* kurang laku mencapai 65.22%.

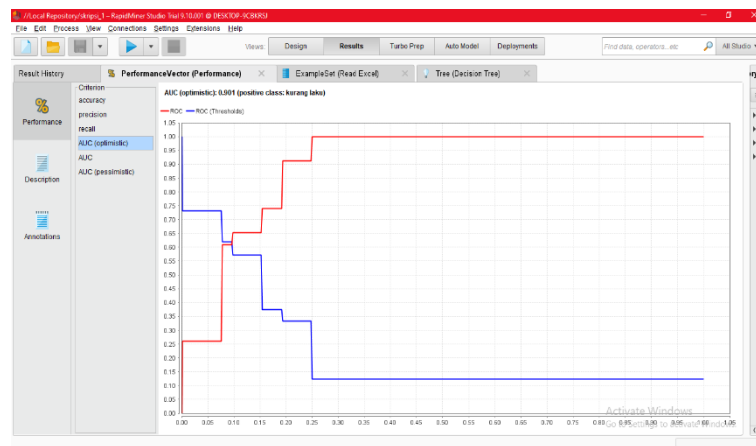
c. Recall



Gambar 5. Tampilan Recall

Hasil pengukuran *recall* data yang didapatkan dari data *training* yaitu berjumlah 65.22% (*positive class* : kurang laku). dari gambar di atas bisa dilihat bahwa untuk prediksi laku dengan *true* laku berjumlah 44 barang dan *true* yang kuran laku berjumlah 8 barang, dan dengan hasil pencapaian dari nilai presisi berjumlah 84.62%. sedangkan untuk dari prediksi yang kurang laku untuk yang *true* laku berjumlah 8 barang dan sedangkan yang *true* kurang laku berjumlah 15 barang, dan dengan hasil dari nilai presisi berjumlah 65.22%. untuk nilai dari *recall* tertinggi terdapat di *class recall true* laku yaitu mencapai 84.62%, sedangkan untuk nilai *true* kurang laku mencapai 65.22%.

d. AUC (Area Under Cover)



Gambar 6. Tampilan AUC

Performance tingkat keakurasian AUC (Gorunescu,2010) bisa diklasifikasikan menjadi lima kategori, yaitu:

- 0.90 – 1.00 = *Excellent Classification*
- 0.80 – 0.90 = *good Classification*
- 0.70 – 0.80 = *Fair Classification*
- 0.60 – 0.70 = *Poor Classification*
- 0.50 – 0.60 = *Failure*

Hasil yang diperoleh dari pengolahan AUC dengan metode algoritma C4.5 sebesar 0.901, dapat dilihat pada gambar di atas dan dikategorikan ke dalam *Excellent Classification*.



4. KESIMPULAN

Berdasarkan hasil penelitian tentang “Prediksi Penjualan Kue Brownis Menggunakan Algoritma C4.5 Pada Toko Amanda Brownis”. Peneliti dapat menyimpulkan dengan menggunakan algoritma C4.5 mendapatkan hasil dengan tingkat akurasi yang cukup tinggi yaitu sebesar 78.67%. jika dilihat dengan kurva ROC, didapatkan hasil *Excellent Clasification* yaitu sebesar 0.901. Dan dapat disimpulkan juga kue brownis dengan ukuran besar lebih diminati dibandingkan dengan brownis yang berukuran kecil serta brownis dengan rasa coco marble dan original lebih diminati dari pada brownis dengan rasa sarikaya pandan dan juga tiramizu marble. Kemudian untuk prediksi penjualan dan untuk meningkatkan penjualan kue brownis di Toko Amanda Brownis dapat dilihat dari hasil analisis penjualan yang sudah dihitung menggunakan algoritma C4.5, yaitu dapat ditambahkan lebih banyak stok brownis dengan ukuran besar dibandingkan dengan ukuran yang kecil. Dan juga ditambahkan lebih banyak kue brownis dengan rasa coco marble dan juga rasa original untuk meningkatkan penjualan dari Toko Amanda Brownis.

REFERENCES

- T. R. I. B. Tusrwenda, “Penerapan data mining dengan algoritma c4.5 dalam prediksi penjualan botol pada cv. seribukilo,” 2018.
- R. Y. Wulandari, “Analisa Data Mining Dengan Metode Klasifikasi Untuk Produk Cacat Pada Pt. Shuangying International Indonesia,” *Progr. Stud. Tek. Inform. Sekol. Tinggi Teknol. Pelita Bangsa*, vol. 14, no. 5, hal. 2–5, 2018.
- Y. Mardi, “Data Mining : Klasifikasi Menggunakan Algoritma C4.5,” *Edik Inform.*, vol. 2, no. 2, hal. 213–219, 2017.
- F. Harahap, “Penerapan Data Mining dalam Memprediksi PembHarahap, F. (2015). Penerapan Data Mining dalam Memprediksi Pembelian cat. Konferensi Nasional Sistem & Informatika, 1(2), 102–108. <https://doi.org/10.1017/CBO9781107415324.004>elian cat,” *Konf. Nas. Sist. Inform.*, vol. 1, no. 2, hal. 102–108, 2015.
- J. Eska, “Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5,” vol. 2, 2018.
- N. Azwanti, “Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada Pt. Capella Dinamik Nusantara Cabang Muka Kuning,” *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 13, no. 1, hal. 33, 2018.
- R. Harman, “Penerapan Data Mining Untuk Prediksi Penjualan Readymix Menggunakan Metode Algoritma C4.5 Pada Pt Remicon Widyaprima,” *J. Comasie*, vol. 3, no. 3, hal. 84–93, 2020.