

KLASIFIKASI PENJUALAN TEMPE DENGAN DATA MINING MENGUNAKAN ALGORITMA C4.5

Suhartini¹, Veren Nita Permatasari², Imam Santoso³

Program Studi Sarjana Sistem Informasi^{1,2}, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika Jakarta^{1,2}, Dosen Universitas Teknologi Muhammadiyah Jakarta³
suhartini@stikomcki.ac.id¹, verenita@stikomcki.ac.id², imam.santoso@utmj.ac.id³

ABSTRAK

Tempe adalah makanan yang terbuat dari fermentasi kacang kedelai yang memiliki kandungan gizi sangat tinggi. Produksi tempe adalah salah satu industry makanan penting dalam populasi umum, kualitas suatu produk sangat penting untuk menjamin kepuasan konsumen. Mengklasifikasikan produk tempe dapat membantu produsen meningkatkan kualitas produk mereka. Di era digitalisasi, ada banyak peluang bisnis di berbagai bidang. Salah satunya di industri makanan, bahkan di Indonesia banyak sekali UKM makanan yang terlibat. Mengingat banyaknya minat masyarakat dalam bisnis dan pekerjaan yang mungkin di kejar di bidang ini, tidak dapat dihindari bahwa kita akan merasakan dorongan untuk memasuki dunia bisnis untuk menuai imbalan finansial atau bahkan lebih besar. Salah satu ide bisnis yang banyak ditemukan adalah usaha tempe. Penjualan merupakan suatu kegiatan transaksi jual – beli barang atau jasa, penjualan tempe di pasar-pasar salah satu contohnya. Bapak Suswanto salah satu orang yang memanfaatkan kondisi dalam memproduksi olahan kacang kedelai menjadi tempe. Implementasi dari masalah ini, yang cocok untuk menyelesaikan masalah klasifikasi data mining, adalah salah satu algoritmanya, yaitu C4.5. Algoritma C4.5 adalah salah satu teknik yang dapat membantu dalam menentukan angka atau, dalam kasus masalah ini, keduanya. Klasifikasi penjualan tempe pada usaha Bapak Suswanto terbukti dengan mendapatkan hasil akurasi sebesar 69.72%, Classification precision sebesar 71.20% dan AUC sebesar 0,586.

Kata Kunci : Penjualan, Tempe, Data Mining, Klasifikasi, Algoritma C4.5, Rapid Miner

ABSTRACT

Tempe is a food made from the fermentation of soybeans that has a very high nutritional content. Tempe production is one of the important food industries in the general population, the quality of a product is very important to ensure consumer satisfaction. Classification of tempered products can help manufacturers improve the quality of their products. In the era of digitalization, there are many business opportunities in various fields. One of them is in the food industry, even in Indonesia a lot of food SMEs are involved. Given the many public interests in business and jobs that may be pursued in this field, it is inevitable that we will feel the urge to enter the world of business to reap financial or even greater rewards. One of the business ideas that many find is tempe effort. Sales is an activity of sale transactions – purchase of goods or services, sales tempe in markets one of the examples. Mr. Suswanto was one of the people who took advantage of the conditions in the processing of soybeans into tempe. The implementation of this problem, which is suitable for solving the problem of data mining classification, is one of its algorithms, namely C4.5. The C4.5 algorithm is one of the techniques that can help in determining numbers or, in the case of this problem, both. Classification of sales tempe on Mr. Suswanto's enterprise was demonstrated by obtaining accuracy of 69.72%, classification precision of 71.20% and AUC of 0,586.

Key word : Sales, Tempeh, Data Mining, Classification, C4.5 Algoritm and Rapid Miner

1. PENDAHULUAN

Besarnya persaingan dalam dunia usaha memungkinkan para pelaku usaha untuk menyusun strategi demi kelangsungan usaha yang ditekuni. Prediksi dalam penjualan dilakukan demi untuk terus berjalannya suatu usaha agar tidak tertinggal (Indah Sari, n.d.). Di era digitalisasi ini banyak peluang bisnis dalam berbagai bidang. Salah satunya bidang

makanan, bahkan di Indonesia banyak sekali UMKM di bidang makanan sudah digeluti. Makanan menjadi hal pokok dalam kehidupan masyarakat, seperti halnya tempe. Sudah banyak olahan tempe yang digemari oleh masyarakat Indonesia bahkan semua kalangan menyukainya.

Tempe ialah salah satu olahan makanan populer dari Indonesia yang berasal dari proses fermentasi kacang kedelai yang sangat tinggi akan gizi. Di Indonesia

olahan tempe dijadikan makanan lauk ataupun keripik serta kerupuk tempe oleh masyarakat. Tempe juga terkenal tidak hanya di Indonesia saja, di negara Eropa tempe terkenal dengan olahan yang mengandung zat antibakteri yang dapat mencegah penyakit jantung serta kolestrol. Selain manfaat dari gizi tempe yang tinggi, tempe termasuk olahan makanan dengan harga yang murah sehingga banyak masyarakat yang minat konsumsi olahan tempe. Tempe sendiri merupakan olahan yang bisa divariasikan dalam pengolahannya dan mudah dibuat. Besarnya minat masyarakat untuk konsumsi tempe menjadikan peluang dalam bisnis.

Salah satu yang memanfaatkan peluang bisnis penjualan tempe yaitu Bapak Suswanto. Usaha penjualan tempe sudah berdiri sejak 2003 dan berlokasi di jalan Wijaya 2 RT 002 RW 004, Jatirahayu Bekasi hingga sekarang penjualan tempe dikelola langsung oleh Bapak Suswanto. Bapak Suswanto dibantu oleh sang istri mengolah fermentasi kacang kedelai sampai menjadi tempe yang kemudian olahan kacang kedelai atau tempe dijual di pasar Pondok Gede Bekasi. Selain menjual Tempe di pasar, Bapak Suswanto juga memiliki satu pedagang perantara yang mendagangkan tempe keliling.

Adanya hasil penjualan yang bervariasi menjadikan usaha Bapak Suswanto harus memiliki strategi untuk menghindari kerugian yang tidak diinginkan melihat dari banyaknya pesaing diluar sana. Bapak Suswanto mengalami kesulitan dalam input data hasil produksi tempe. Belum adanya system untuk mengolah data serta kurangnya pengetahuan menyebabkan sulitnya penanganan masalah tersebut. Hal ini memerlukan klasifikasi dalam memecahkan masalah yang dihadapi. Analisis dari pola data penjualan untuk mendapatkan hasil klasifikasi dari penjualan tempe (Setiawan et al., n.d.). Sehingga menghasilkan sebuah pohon keputusan untuk melihat prediksi dari penjualan tempe Bapak Suswanto.

Implementasi dari masalah tersebut yang sangat cocok digunakan dalam penyelesaian masalahnya dengan penggunaan data mining klasifikasi salah satu algoritmanya yaitu C4.5 dapat dengan mudah mengetahui pola produksi tempe (Arif & Yulianto, 2022; Calvin1 & Rika Harman, 2022). Kelebihan algoritma C4.5 dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima, efisien dalam menangani atribut bertipe diskret dan dapat

menangani atribut bertipe diskret dan numerik (Pambudi et al., 2021)

2. METEDOLOGI

2.1. Pengumpulan Data

Pada penelitian ini dilakukan pengumpulan data dengan wawancara dan observasi langsung di tempat produksi Tempe Bapak Suswanto yang berada di jalan Wijaya 2 RT 002 RW 004, Jatirahayu Bekasi. Sumber data diperoleh dari wawancara kepada Bapak Suswanto dan penunjang data lainnya.

2.2. Algoritma C4.5

Berikut tahapan-tahapan dalam algoritma C4.5 sebagai berikut :

- a. Menghitung nilai entropy
- b. Menghitung nilai gain ratio masing-masing atribut
- c. Atribut yang memiliki gain ratio tertinggi dipilih menjadi akar dan atribut yang memiliki nilai gain ratio lebih rendah dijadikan cabang
- d. Menghitung nilai gain ratio dari tiap-tiap atribut kecuali akar
- e. Atribut yang memiliki gain ratio tertinggi dipilih menjadi akar
- f. Ulangi langkah 4 dan 5 sampai dihasilkan nilai gain ratio = 0.

Untuk menghitung nilai entropy dapat dihitung dengan persamaan berikut.

Rumus perhitungan entropy

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

Keterangan :

S berupa himpunan kasus

A berupa atribut yang digunakan

n berupa jumlah partisi yang dimiliki atribut

p_i berupa proporsi dari S_i kepada S

Sementara itu nilai informasi gain (Gain) dapat dihitung menggunakan persamaan :

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

Dimana :

S berupa himpunan kasus

A berupa atribut yang digunakan

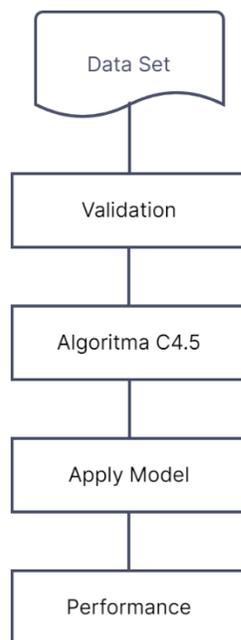
n berupa jumlah partisi yang dimiliki oleh atribut

$|S_i|$ berupa jumlah partisi yang dimiliki oleh partisi ke – i

2.3. Metode

Pada tahap penelitian kali ini akan menggunakan rapid miner studio versi

10.0, untuk menunjukkan bagaimana penggambaran serta penyelesaian masalah penelitian dengan banyaknya tahapan. Berikut disajikan rancangan dari tahapan-tahapan penelitian dibawah ini :



Gambar 1. Metode Penelitian

3. LANDASAN TEORI

3.1. Penjualan

Penjualan adalah proses atau inisiatif menjual barang atau jasa kepada pelanggan dengan tujuan menghasilkan pendapatan atau keuntungan finansial bagi perusahaan atau orang yang melakukan penjualan. Penjualan mengacu pada interaksi antara penjual dan pembeli di mana penjual

menawarkan produk atau layanan kepada pelanggan dan pembeli melakukan pembelian berdasarkan kebutuhan atau keinginan mereka (Arif & Yulianto, 2022)

3.2. Data Mining

Data Mining, sering juga disebut sebagai *knowledge discovery in database* (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data yang berukuran besar. Data Mining didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semiotomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Data yang dibutuhkan dalam jumlah besar (Santoso et al., 2016)

Menurut Turban dalam bukunya yang berjudul "*Decision Support Systems and Intelligent Systems*", data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam basis data.

3.3. Klasifikasi

Klasifikasi merupakan proses mengembangkan model atau sistem yang dapat digunakan untuk menentukan kategori atau tingkat entitas baru

berdasarkan pengetahuan yang diperoleh dari entitas yang ada (Rachman et al., 2019).

Klasifikasi adalah sebuah proses pengelompokan koleksi objek platform data berdasarkan kategori kesamaan dan klasifikasi yang telah ditentukan sebelumnya (Rahman et al., 2020)

3.4. Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma klasifikasi dimana dapat digunakan untuk pengambilan atau menentukan keputusan atau pohon keputusan yang biasa disebut oleh decision tree (Tri Romadloni et al., n.d.). Algoritma C4.5 merupakan salah satu algoritma yang dapat digunakan untuk membangun pohon keputusan

Metode pohon keputusan mengubah fakta-fakta besar menjadi pohon keputusan yang representative aturan (Rahman et al., 2020). Pohon keputusan dibentuk dengan tujuan untuk mempermudah dalam mengatasi penyelesaian masalah. Dalam penggunaan algoritma C4.5 ada beberapa tahapan umum yaitu mengubah bentuk data dalam table menjadi pohon keputusan, kemudian mengubah model pohon menjadi sebuah aturan (*rule*) dan menyederhanakan aturan (*rule*) (Basuki et al., n.d.)

4. HASIL DAN PEMBAHASAN

4.1. Analisa Data

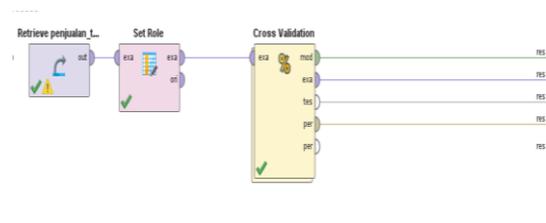
Data yang digunakan merupakan data penjualan tempe dalam kurun waktu 2021 – 2023 (29 bulan). Berikut data penjualan yang disajikan pada gambar dibawah ini :

| Tanggal | Jumlah Tempe (KG) | Harga / Kg | Penjualan | Bentuk | Kemasan | Jumlah Terjual |
|---------------|-------------------|------------|-------------------|---------------|---------|----------------|
| Januari 2021 | 1550 | Rp 10.000 | Produsen | Papan | Plastik | Sedikit |
| Januari 2021 | 1240 | Rp 10.000 | Penjual Perantara | Lonjor | Daun | Banyak |
| Januari 2021 | 620 | Rp 8.000 | Produsen | Bungkus Kecil | Daun | Banyak |
| Februari 2021 | 1400 | Rp 10.000 | Produsen | Papan | Plastik | Sedikit |
| Februari 2021 | 1120 | Rp 10.000 | Penjual Perantara | Lonjor | Daun | Sedikit |
| Februari 2021 | 560 | Rp 8.000 | Produsen | Bungkus Kecil | Daun | Banyak |
| Maret 2021 | 1550 | Rp 10.000 | Penjual Perantara | Papan | Plastik | Banyak |
| Maret 2021 | 1085 | Rp 10.000 | Produsen | Lonjor | Daun | Banyak |
| Maret 2021 | 620 | Rp 8.000 | Produsen | Bungkus Kecil | Daun | Banyak |
| April 2021 | 1350 | Rp 10.000 | Produsen | Papan | Plastik | Banyak |
| April 2021 | 1200 | Rp 10.000 | Penjual Perantara | Lonjor | Daun | Sedikit |
| April 2021 | 450 | Rp 8.000 | Produsen | Bungkus Kecil | Daun | Sedikit |
| Mei 2021 | 1705 | Rp 10.000 | Produsen | Papan | Plastik | Banyak |
| Mei 2021 | 1085 | Rp 10.000 | Produsen | Lonjor | Daun | Banyak |
| Mei 2021 | 620 | Rp 8.000 | Produsen | Bungkus Kecil | Daun | Banyak |
| Juni 2021 | 1800 | Rp 10.000 | Produsen | Papan | Plastik | Banyak |
| Juni 2021 | 1650 | Rp 10.000 | Penjual Perantara | Lonjor | Daun | Banyak |
| Juni 2021 | 600 | Rp 8.000 | Produsen | Bungkus Kecil | Daun | Banyak |
| Juli 2021 | 1860 | Rp 10.000 | Produsen | Papan | Plastik | Sedikit |
| Juli 2021 | 1550 | Rp 10.000 | Penjual Perantara | Lonjor | Daun | Sedikit |
| Juli 2021 | 620 | Rp 8.000 | Produsen | Bungkus Kecil | Daun | Banyak |

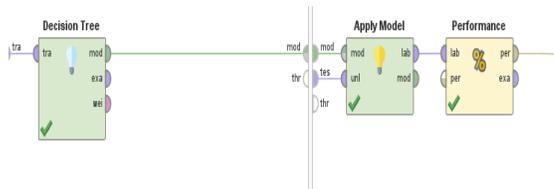
Gambar 2. Data Penjualan

4.2. Proses Data Mining

Proses pengolahan data mining penjualan tempe dengan algoritma C4.5 dilakukan melalui aplikasi Rapid Miner. Berikut desain model dalam pengolahan dataset :



Gambar 3. Model Algoritma C4.5



Gambar 4. Model Training & Testing

4.3. Hasil Akurasi

Berdasarkan hasil penerapan algoritma C4.5 pada data penjualan tempe, maka mendapatkan hasil akurasi yang baik yaitu sebesar 69,72%. Hal ini tertuang pada table berikut :

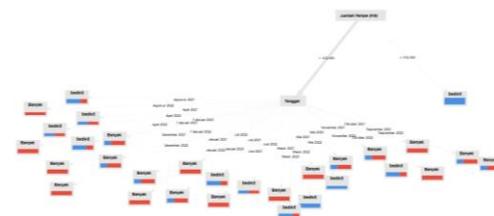
Tabel 1. Confusion Matrix

| | True Sedikit | True Banyak | Class Precision |
|---------------|--------------|-------------|-----------------|
| Pred. Sedikit | 12 | 7 | 63.16% |
| Pred. Banyak | 19 | 49 | 72.06% |
| Class Recall | 38.71% | 87.50% | |

Berdasarkan tabel di atas menjelaskan bahwa akurasi yang dicapai adalah 69.72% dengan rincian seperti berikut : Hasil prediksi Sedikit untuk True Sedikit sebanyak 12 dan Prediksi Sedikit untuk True Banyak 7 dengan Class Precision sebesar 63.16%. Hasil prediksi Banyak untuk True Sedikit sebanyak 19 dan Prediksi Banyak untuk True Banyak 49 dengan Class Precision sebesar 72.06%.

4.4. Pembahasan

Pada penelitian penjualan tempe mendapatkan pembahasan dalam bentuk graph, AUC dan performance vector seperti gambar berikut :



Gambar 5. Graph Penjualan Tempe



Gambar 6. AUC

Pada pengolahan data set penjualan mendapatkan hasil AUC sebesar 0.586

PerformanceVector

```
PerformanceVector:
accuracy: 69.72% +/- 21.28% (micro average: 70.11%)
ConfusionMatrix:
True: Sedikit Banyak
Sedikit: 12 7
Banyak: 19 49
precision: 71.20% +/- 15.81% (micro average: 72.06%) (positive class: Banyak)
True: Sedikit Banyak
Sedikit: 12 7
Banyak: 19 49
recall: 87.00% +/- 28.30% (micro average: 87.50%) (positive class: Banyak)
ConfusionMatrix:
True: Sedikit Banyak
Sedikit: 12 7
Banyak: 19 49
AUC (optimistic): 0.737 +/- 0.294 (micro average: 0.737) (positive class: Banyak)
AUC: 0.586 +/- 0.256 (micro average: 0.586) (positive class: Banyak)
AUC (pessimistic): 0.459 +/- 0.310 (micro average: 0.459) (positive class: Banyak)
```

Gambar 7. Performance Vector

5. KESIMPULAN

Hasil penelitian yang dilakukan penulis menarik kesimpulan bahwa penelitian ini menggunakan 7 atribut yaitu Tanggal Produksi, Jumlah Tempe Produksi, Harga, Penjualan, Bentuk Tempe, Kemasan dan Jumlah Terjual. Metode klasifikasi dalam penjualan tempe menggunakan algoritma C4.5 sangat baik untuk menentukan kebenaran dari klasifikasi dalam data mining. Namun hasil klasifikasi dengan algoritma C4.5 untuk menganalisa data penjualan mendapatkan akurasi sebesar 69.72%, Classification precision sebesar 71.20% dan AUC sebesar 0,586.

DAFTAR PUSTAKA

- Arif, A., & Yulianto, A. (2022). Klasifikasi Penjualan Makanan Hewan Peliharaan di Muezza Petshop01 Bogor Menggunakan Metode Algoritma C4.5. *Remik*, 6(4), 840–854. <https://doi.org/10.33395/remik.v6i4.11859>
- Basuki, A., Syarif, I., Elektronika, P., & Surabaya, N. (n.d.). *Decision Tree Decision Tree Decision Tree*.
- Calvin1, & Rika Harman. (2022). PENERAPAN ALGORITMA C4.5 UNTUK MEMREDIKSI PROFIT PENJUALAN. *Comasie*, 3(3).
- Indah Sari, A. (n.d.). *OKTAL : Jurnal Ilmu Komputer dan Science Prediksi Penjualan Sepatu Dengan Algoritma C4.5 Di T&T Collection Di Tangerang*. <https://journal.mediapublikasi.id/index.php/oktal>
- Pambudi, F. A., Windarto, A. P., Fauzan, M., & ... (2021). Analisis Klasifikasi C4. 5 Pada Pola Pembayaran Sepeda Motor Adira Cabang Pematangsiantar. *KLIK: Kajian Ilmiah ...*, 1(4).
- Rachman, A., Furqon, M. T., & Ramdani, F. (2019). *Klasifikasi Varietas Unggul Padi menggunakan Algoritme C4.5* (Vol. 3, Issue 9). <http://j-ptiik.ub.ac.id>
- Rahman, F., Zahro', H. Z., & Ariwibisono, F. X. (2020). PENERAPAN ALGORITMA C4.5 DALAM MEMREDIKSI ASAL CALON MAHASISWA BERBASIS WEBSITE (STUDI KASUS : FAKULTAS HUKUM UNIVERSITAS MATARAM). In *Jurnal Mahasiswa Teknik Informatika* (Vol. 4, Issue 2).
- Santoso, H., Hariyadi, I. P., & Prayitno. (2016). Data Mining Analisa Pola Pembelian Produk. *Teknik Informatika*, 1.
- Setiawan, K., Rahmatullah, B., Budi Paryanti, A., Fauzi, F., Ilmu Komputer, M., Nusa Mandiri, S., & Suryadarma, U. (n.d.). *JISAMAR (Journal of Information System, Applied, Management, Accounting and Researh) KOMPARASI METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE MENGGUNAKAN PARTICLE*

*SWARM OPTIMIZATION UNTUK
ANALISIS SENTIMEN MOBIL
ESEMKA.*

<http://journal.stmikjayakarta.ac.id/index.php/jisamar>Telp.+62-21-3905050

Tri Romadloni, N., Santoso, I.,
Budilaksono, S., & Ilmu Komputer
STMIK Nusa Mandiri Jakarta, M.
(n.d.). *PERBANDINGAN METODE
NAIVE BAYES, KNN DAN
DECISION TREE TERHADAP
ANALISIS SENTIMEN
TRANSPORTASI KRL COMMUTER
LINE.*