

ANALISIS PERBANDINGAN ALGORITMA NAIVE BAYES DAN C4.5 DALAM KLASIFIKASI PRODUK MAKANAN OLAHAN TERLARIS PADA Y3 FROZEN FOOD

Nazwa Fitri Gufary¹, Muhammad Sadikin²

1,2) Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Potensi Utama

Article Info	ABSTRACT
Article history: Received: 21 Mei 2026 Revised: 25 Juli 2026 Accepted: 12 Juli 2026	Abstrak Y3 Frozen Food adalah UMKM yang menjual berbagai produk makanan olahan seperti nugget, dimsum, lumpia dan lainnya. Pemilik usaha mengalami kesulitan dalam menentukan produk makanan olahan terlaris dikarenakan banyaknya variasi produk makanan olahan yang dijual. kesulitan yang terjadi berdampak pada pengelolaan produk Y3 Frozen Food menjadi tidak optimal karena berdasarkan intuisi atau perkiraan untuk promosi suatu produk tanpa dilakukan analisis data historis penjualan terlebih dahulu. Guna mengatasi permasalahan yang terjadi, penelitian menerapkan <i>data mining</i> melalui klasifikasi produk makanan olahan kedalam 3 kelas yaitu laris, kurang laris, dan tidak laris. Penelitian ini memilih metode klasifikasi yang akan digunakan adalah algoritma Naive Bayes dengan metode tambahan <i>laplace smoothing</i> dan algoritma C4.5. Tujuan penelitian dilakukan untuk menentukan algoritma klasifikasi yang terbaik menggunakan tools rapidminer versi 10.3.1 berdasarkan akurasi <i>confusion matrix</i> melalui K-Fold Cross Validation pengujian K=3, K=5, K=7. Hasil akurasi <i>confusion matrix</i> dari algoritma naive bayes paling tinggi berkisar 91,27% pengujian K=5 dan algoritma C4.5 berkisar 98,90% untuk setiap pengujian. Sehingga algoritma C4.5 dipilih menjadi yang algoritma yang terbaik berdasarkan hasil evaluasi <i>confusion matrix</i>. Kata Kunci: Naive Bayes, C4.5, Klasifikasi, Data Mining, Makanan Olahan Abstract Y3 Frozen Food is an MSMEs that sells various processed food products such as nuggets, dimsum, spring rolls and others. Business owners experience has difficulties in determining the best-selling processed food products due to the large variety of processed food products sold. The difficulties that occur has an impact on the management of Y3 Frozen Food products being suboptimal because it is based on intuition or estimates for product promotion without analyzing historical sales data first. To overcome the problems that occur, the study is using data mining through the classification of processed food products into 3 classes which is best-selling, less popular, and not popular. This study chose the classification method to be used is the Naive Bayes algorithm with additional methods of Laplace smoothing and the C4.5 algorithm. The purpose of the study was to determine the best classification algorithm by using rapidminer version 10.3.1 tools based on the accuracy of the confusion matrix through k-fold cross validation testing K = 3, K = 5, K = 7. The results of the confusion matrix accuracy of the naive bayes algorithm were highest at around 91.27% for testing K = 5 and the C4.5 algorithm was around 98.90% for each test. the conclusion is C4.5 algorithm was chosen as the best algorithm based on the results of the confusion matrix evaluation. Keywords: Naive Bayes, C4.5, Classification, Data Mining, Process Food Djtechno: Jurnal Teknologi Informasi oleh Universitas Dharmawangsa Artikel ini bersifat open access yang didistribusikan di bawah syarat dan ketentuan dengan Lisensi Internasional Creative Commons Attribution NonCommercialL ShareAlike 4.0 (CC-BY-NC-SA). 

1. PENDAHULUAN

Makanan olahan merupakan suatu proses yang mengubah produk pangan dari kondisi aslinya ke dalam bentuk siap untuk dimakan [1]. Tujuan dari semua proses dilakukan, guna membuat makanan menjadi lebih praktis dikonsumsi, memperpanjang usia simpan produk, dan bertahan lama saat didistribusikan ke berbagai tempat [2]. Salah satunya Y3 *Frozen Food* yang menawarkan 12 jenis produk makanan olahan dari 13 merk, yang dikelompokkan ke dalam 5 kategori olahan, memiliki 7 variasi berat, serta 18 variasi harga.

UMKM dengan berbagai sektor yang mencakup perdagangan, perindustrian, pertanian, jasa, hingga menjadi bagian dari rantai pasok lokal memiliki peran penting dalam mendorong pertumbuhan perekonomian sehingga perlu didukung agar mampu bersaing di tengah perkembangan usaha [3]. Seperti UMKM Y3 *Frozen Food*, dari 3362 data historis penjualan mencatat bahwa produk Nugget Goreng memiliki jumlah penjualan tertinggi, sebanyak 2108 unit selama periode September 2024 sampai November 2025. Tetapi, produk tersebut dipasarkan oleh beberapa merk membuat pemilik usaha masih mengalami kesulitan dalam menentukan produk yang paling laris berdasarkan merk, harga, dan berat sebagai dasar strategi promosi.

Klasifikasi adalah salah satu teknik *data mining* dengan tujuan menemukan hubungan atau pola tersembunyi diantara beberapa (fitur/atribut/variabel/kriteria) yang saling berkaitan pada data[4]. Dalam proses klasifikasi, sistem akan dilatih menggunakan data yang sudah memiliki "jawaban benar" (disebut *training data* yang sudah dilabeli), sehingga model bisa menentukan suatu data baru kedalam label yang sesuai. Pada penelitian ini, klasifikasi digunakan agar pengambilan keputusan lebih tepat sasaran karena tidak berdasarkan intuisi atau perkiraan lagi tetapi sudah berdasarkan hasil klasifikasi produknya.

Dengan adanya hal tersebut, penelitian ini menganalisis perbandingan kinerja dari algoritma naive bayes dan algoritma C4.5 dalam mengklasifikasikan produk olahan terlaris pada Y3 *Frozen Food*. Data yang digunakan pada analisis, periode september 2024 sampai november 2025 berkisar 3362 data historis penjualan yang kemudian diolah menggunakan *tools* rapidminer versi 10.3.1. Evaluasi kinerja algoritma menggunakan *confusion matrix* dengan menerapkan *K-Fold Cross Validation*, K=3, K=5, K=7. Hasil

evaluasi dengan algoritma yang memiliki kinerja paling baik, diambil untuk menjadi acuan dalam pengelolaan produk yang lebih baik pada Y3 *Frozen Food*.

Penelitian terkait oleh M. Sadikin et al, dengan judul penelitian “Perbandingan Tingkat Akurasi Klasifikasi Penerimaan Dosen Tetap Menggunakan Metode Naive Bayes Classifier dan C4.5”. Studi penelitian ini menyimpulkan bahwa metode C4.5 lebih unggul dengan akurasi 91,89% dibandingkan metode naive bayes yang memperoleh akurasi 83,78% [5].

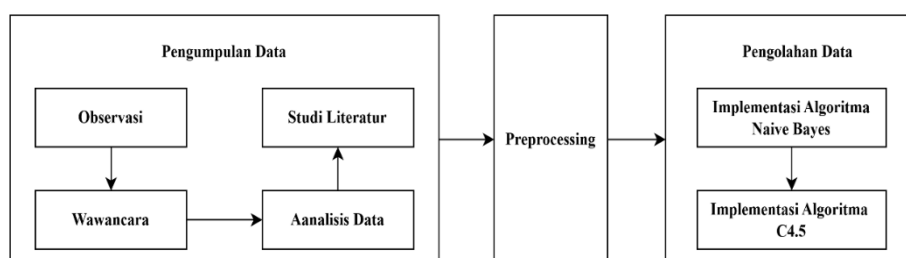
Penelitian terkait lainnya yang dilakukan oleh R. Pratama et al, dengan judul “Perbandingan Algoritma C4.5 dan Naive Bayes dalam Menentukan Persediaan Stok”. Studi penelitian ini menyimpulkan bahwa metode C4.5 lebih unggul dengan akurasi sebesar 96,80% sedangkan naive bayes akurasinya berkisar 91, 20% [6].

Penelitian terkait yang dilakukan oleh M. K. A. Fatah dan A. Wibowo dengan judul penelitian “Perbandingan Metode Algoritma C4.5 dan Naive Bayes Untuk Memprediksi Penjualan Kosmetik Pada Toko Jelita”. Studi penelitian ini menyimpulkan bahwa algoritma C4.5 lebih unggul dengan Tingkat akurasi sebesar 100% sedangkan naive bayes memperoleh akurasi sebesar 98% [7].

2. METODE PENELITIAN

2.1. Alur Analisis Penelitian

Terdapat alur analisis penelitian dalam bentuk diagram *Knowledge Discovery System* (KDD) disajikan pada Gambar1.



Gambar 1. Alur Analisis Penelitian

Alur analisis penelitian diawali dengan mengumpulkan data melalui 4 proses berupa observasi pada Y3 *Frozen Food*, wawancara kepada pemilik dari subjek penelitian, analisis data terhadap data historis penjualan, dan studi literatur guna mendapatkan referensi dan teori yang relevan terkait penelitian. alur berikutnya adalah *preprocessing* atau mempersiapkan data menjadi lebih terstruktur dan disesuaikan untuk kebutuhan

pengelahan data. Kemudian tahapan terakhir melakukan pengolahan 3362 data historis penjualan dengan variabel merk, olahan, berat, harga, dan terjual menggunakan algoritma Naive Bayes dan algoritma C4.5. Pengujian dilakukan menggunakan *K-Fold Cross Validation* dengan iterasi K=3, K=5, dan K=7 untuk melihat kestabilan kinerja algoritma pada pembagian data latih dan data uji yang berbeda.

2.2. Algoritma Naive Bayes

Naive bayes adalah salah satu algoritma klasifikasi guna memprediksi label atau kategori suatu data berdasarkan atribut yang dimiliki [8]. Prinsip naive bayes bertumpu pada perhitungan probabilitas dan statistik yang didasarkan pada teorema bayes yang memiliki asumsi bahwa semua atribut bersifat independen atau tidak saling tergantung pada nilai atribut lainnya [9]. Teorema bayes memungkinkan penghitungan probabilitas suatu kelas (posterior) dengan memanfaatkan probabilitas awal kelas tersebut (prior), dan probabilitas kemunculan atribut yang ada jika kelasnya telah diketahui (*likelihood*) [10]. Adapun rumus persamaan teorema bayes sebagai berikut.

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

Keterangan :

X : Data dengan kelas yang tidak diketahui

H : Data hipotesis kelas tertentu

P(H|X): Probabilitas hipotesis H berdasarkan kondisi X (*posteriori probability*)

P(H) : Probabilitas hipotesis H (*prior probability*)

P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) : Probabilitas X

2.3. Laplace Smoothing

Laplace smoothing atau dikenal dengan *add one smoothing* adalah teknik tambahan algoritma naive bayes untuk mengatasi nilai 0 dengan menambah nilai konstanta 1 pada setiap frekuensi data [11]. Teknik ini memastikan bahwa setiap atribut tetap memiliki nilai probabilitas sehingga model tangguh dan bisa melakukan klasifikasi secara optimal. Adapun rumus persamaan *laplace smoothing* sebagai berikut.

$$P(H|X) = \frac{P(X|H) \cdot P(H) + 1}{P(X) + |V|} \quad (2)$$

Keterangan :

X : Data dengan kelas yang tidak diketahui

H : Data hipotesis kelas tertentu

$P(H|X)$: Probabilitas hipotesis H berdasarkan kondisi X (*posteriori probability*)

$P(H)$: Probabilitas hipotesis H (*prior probability*)

$P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$: Probabilitas X

+1 = *Laplace smoothing* (penghalusan laplace)

$|V|$ = Jumlah kemungkinan nilai fitur

2.4. Algoritma C4.5

Algoritma C4.5 adalah salah satu algoritma klasifikasi berbentuk pohon keputusan (*decision tree*) dengan memprediksi nilai atribut target berdasarkan atribut input. C4.5 memilih atribut terbaik sebagai *gain ratio* (rasio keuntungan) dan hasil pohon keputusan membentuk aturan IF – Then yang mudah dipahami [12]. Adapun rumus persamaan menghitung *entropy* sebagai berikut.

$$\text{Entropi (S)} = \sum_{j=1}^k - p_j \log_2 p_j \quad (3)$$

Dimana :

1. P_j = Proporsi jumlah kejadian nilai atribut terhadap jumlah total kejadian
2. k = Jenis Kasus

Adapun rumus untuk menghitung *Gain Ratio* :

$$\text{Gain(A)} = \text{Entropi (S)} - \sum_{i=1}^k \frac{|S_i|}{|S|} \times \text{Entropi (S}_i) \quad (4)$$

Dimana :

1. S_i = Jumlah kejadian nilai atribut i
2. S = Jumlah kejadian total
3. k = Jumlah nilai atribut.

3. HASIL DAN PEMBAHASAN

3.1 Implementasi Algoritma Naive Bayes

Algoritma naive bayes memiliki asumsi bahwa setiap atribut saling independen ketika nilai *output* telah didapatkan, sehingga nilai probabilitas muncul diperoleh dari hasil perkalian masing – masing atribut [13]. Keunggulan algoritma naive bayes terletak

pada bisa memproses data yang bertipe kuantitatif dan kualitatif tanpa memerlukan jumlah data latih (*training data*) yang besar untuk mengenali pola dan penaksiran parameter yang dibutuhkan dalam proses klasifikasi. Adapun implementasi algoritma naive bayes yaitu:

3.1.1 Data Latih (*Training Data*)

Proses klasifikasi memiliki data latih sebagai berikut.

Tabel 1. Data Latih

No	Merk	Olahan	Berat	Harga	Terjual	Kelas
1	Popularitas Sedang	Daging	Besar	Murah	Sedang	Kurang Laris
2	Popularitas Tinggi	Sayur	Besar	Sedang	Sedang	Kurang Laris
3	Popularitas Sedang	Daging	Sedang	Murah	Sedikit	Tidak Laris
4	Popularitas Sedang	Daging	Besar	Murah	Sedikit	Tidak Laris
5	Popularitas Tinggi	Daging	Besar	Sedang	Sedang	Laris
6	Popularitas Rendah	Daging	Besar	Sedang	Sedikit	Tidak Laris
7	Popularitas Tinggi	Daging	Besar	Sedang	Sedang	Laris
8	Popularitas Tinggi	Daging	Besar	Sedang	Sedikit	Tidak Laris
9	Popularitas Tinggi	Daging	Sedang	Mahal	Sedang	Laris
10	Popularitas Sedang	Daging	Besar	Sedang	Banyak	Laris
11	Popularitas Tinggi	Seafood	Besar	Murah	Sedang	Tidak Laris
12	Popularitas Sedang	Daging	Besar	Murah	Sedang	Kurang Laris
13	Popularitas Tinggi	Daging	Besar	Sedang	Banyak	Laris
14	Popularitas Tinggi	Daging	Besar	Mahal	Sedikit	Kurang Laris
15	Popularitas Rendah	Daging	Besar	Sedang	Sedikit	Tidak Laris
16	Popularitas Tinggi	Seafood	Besar	Sedang	Sedang	Kurang Laris
17	Popularitas Tinggi	Daging	Sedang	Murah	Sedang	Laris
18	Popularitas Sedang	Daging	Sedang	Murah	Sedikit	Tidak Laris
19	Popularitas Sedang	Seafood	Besar	Sedang	Sedang	Tidak Laris
20	Popularitas Tinggi	Daging	Besar	Mahal	Sedang	Laris
21	Popularitas Rendah	Daging	Besar	Sedang	Sedikit	Tidak Laris
22	Popularitas Sedang	Daging	Besar	Murah	Sedang	Kurang Laris
23	Popularitas Tinggi	Daging	Besar	Sedang	Sedang	Laris
24	Popularitas Tinggi	Daging	Kecil	Murah	Banyak	Laris
25	Popularitas Tinggi	Daging	Besar	Mahal	Sedang	Laris
26	Popularitas Sedang	Daging	Besar	Murah	Sedang	Kurang Laris
27	Popularitas Tinggi	Seafood	Besar	Sedang	Sedikit	Tidak Laris
28	Popularitas Rendah	Daging	Besar	Sedang	Sedang	Tidak Laris
29	Popularitas Tinggi	Daging	Besar	Sedang	Sedang	Laris
30	Popularitas Sedang	Daging	Besar	Murah	Sedikit	Tidak Laris
...	...	...	...	...	...	...
3362	Popularitas Sedang	Daging	Besar	Murah	Sedikit	Tidak Laris

3.1.2. Variabel dan Perhitungan Probabilitas

Adapun nilai probabilitas pada tiap variabel diperoleh dari proses perhitungan berdasarkan tabel 1. Data Latih sehingga nilai probabilitas sebagai berikut.

3.1.2.1 Probabilitas Merk

Pada tabel 1. Data Latih, terdapat 571 merk popularitas tinggi yang laris, 623 merk popularitas tinggi tapi kurang laris, 688 merk popularitas tinggi tidak laris, 39 merk popularitas sedang yang laris, 260 merk popularitas sedang yang kurang laris, 935 merk popularitas sedang yang tidak laris, 4 merk popularitas rendah yang laris, 0 merk popularitas rendah yang kurang laris, dan 242 merk popularitas rendah yang tidak laris. Probabilitas terhadap variabel merk disajikan pada tabel 2.

Tabel 2. Probabilitas Merk

MERK	Kelas			Probabilitas		
	Laris	Kurang Laris	Tidak Laris	Laris	Kurang Laris	Tidak Laris
Popularitas Tinggi	571	623	688	0,930	0,706	0,369
Popularitas Sedang	39	260	935	0,064	0,294	0,501
Popularitas Rendah	4	0	242	0,007	0,000	0,130
Jumlah	614	883	1865	1,000	1,000	1,000

Adapun sampel perhitungan nilai untuk sub variabel populer :

$$P(\text{Popularitas Tinggi}|\text{Laris}) = \frac{571}{614} = 0,930$$

$$P(\text{Popularitas Tinggi}|\text{Kurang Laris}) = \frac{623}{883} = 0,706$$

$$P(\text{Popularitas Tinggi}|\text{Tidak Laris}) = \frac{688}{1865} = 0,369$$

3.1.2.2 Probabilitas Merk

Pada tabel 1. Data Latih, terdapat 584 olahan daging yang laris, 752 olahan daging tapi kurang laris, 1482 olahan daging tidak laris, 21 olahan *seafood* yang laris, 90 olahan *seafood* yang kurang laris, 301 olahan *seafood* yang tidak laris, 9 olahan sayur yang laris, 41 olahan sayur yang kurang laris, dan 82 olahan sayur yang tidak laris. Probabilitas terhadap variabel olahan disajikan pada tabel 3.

Tabel 3. Probabilitas Olahan

OLAHAN	Kelas			Probabilitas		
	Laris	Kurang Laris	Tidak Laris	Laris	Kurang Laris	Tidak Laris
Daging	584	752	1482	0,951	0,852	0,795
Seafood	21	90	301	0,034	0,102	0,161
Sayur	9	41	82	0,015	0,046	0,044

Jumlah	614	883	1865	1,000	1,000	1,000
--------	-----	-----	------	-------	-------	-------

Adapun sampel perhitungan nilai untuk sub variabel daging:

$$P(\text{Daging}|\text{Laris}) = \frac{584}{614} = 0,951$$

$$P(\text{Daging}|\text{Kurang Laris}) = \frac{752}{883} = 0,852$$

$$P(\text{Daging}|\text{Tidak Laris}) = \frac{1482}{1865} = 0,795$$

3.1.3.3 Probabilitas Berat

Pada tabel 1. Data Latih, terdapat 341 berat besar yang laris, 478 berat besar tapi kurang laris, 1578 berat besar tidak laris, 263 berat sedang yang laris, 405 berat sedang yang kurang laris, 205 berat sedang yang tidak laris, 10 berat kecil yang laris, 0 berat kecil yang kurang laris, dan 82 berat kecil yang tidak laris. Probabilitas terhadap variabel olahan disajikan pada tabel 4.

Tabel 4. Probabilitas Berat

BERAT	Kelas			Probabilitas		
	Laris	Kurang Laris	Tidak Laris	Laris	Kurang Laris	Tidak Laris
Besar	341	478	1578	0,555	0,541	0,846
Sedang	263	405	205	0,428	0,459	0,110
Kecil	10	0	82	0,016	0,000	0,044
Jumlah	614	883	1865	1,000	1,000	1,000

Adapun sampel perhitungan nilai untuk sub variabel besar:

$$P(\text{Besar}|\text{Laris}) = \frac{341}{614} = 0,555$$

$$P(\text{Besar}|\text{Kurang Laris}) = \frac{478}{883} = 0,541$$

$$P(\text{Besar}|\text{Tidak Laris}) = \frac{1578}{1865} = 0,846$$

3.1.2.4 Probabilitas Harga

Pada tabel 1. Data Latih, terdapat 301 harga mahal yang laris, 578 harga mahal tapi kurang laris, 128 harga mahal tidak laris, 244 harga sedang yang laris, 217 harga sedang yang kurang laris, 983 harga sedang yang tidak laris, 69 harga murah yang laris, 88 harga murah yang kurang laris, dan 754 harga murah yang tidak laris. Probabilitas terhadap variabel olahan disajikan pada tabel 5.

Tabel 5. Probabilitas Harga

HARGA	Kelas			Probabilitas		
	Laris	Kurang Laris	Tidak Laris	Laris	Kurang Laris	Tidak Laris
Mahal	301	578	128	0,490	0,655	0,069

Sedang	244	217	983	0,397	0,246	0,527
Murah	69	88	754	0,112	0,100	0,404
Jumlah	614	883	1865	1,000	1,000	1,000

Adapun sampel perhitungan nilai untuk sub variabel mahal:

$$P(\text{Mahal}|\text{Laris}) = \frac{301}{614} = 0,490$$

$$P(\text{Mahal}|\text{Kurang Laris}) = \frac{578}{883} = 0,655$$

$$P(\text{Mahal}|\text{Tidak Laris}) = \frac{128}{1865} = 0,069$$

3.1.2.5 Probabilitas Terjual

Pada tabel 1. Data Latih, terdapat 131 terjual banyak yang laris, 0 terjual banyak tapi kurang laris 0 terjual banyak tidak laris, 484 terjual sedang yang laris, 351 terjual sedang yang kurang laris, 161 terjual sedang yang tidak laris, 0 terjual sedikit yang laris, 532 terjual sedikit yang kurang laris, dan 1704 terjual sedikit yang tidak laris. Probabilitas terhadap variabel olahan disajikan pada tabel 6.

Tabel 6. Probabilitas Terjual

TERJUAL	Kelas			Probabilitas		
	Laris	Kurang Laris	Tidak Laris	Laris	Kurang Laris	Tidak Laris
Banyak	130	0	0	0,212	0,000	0,000
Sedang	484	351	161	0,788	0,398	0,086
Sedikit	0	532	1704	0,000	0,602	0,914
Jumlah	614	883	1865	1,000	1,000	1,000

Adapun sampel perhitungan nilai untuk sub variabel sedang:

$$P(\text{Sedang}|\text{Laris}) = \frac{484}{614} = 0,788$$

$$P(\text{Sedang}|\text{Kurang Laris}) = \frac{351}{883} = 0,398$$

$$P(\text{Sedang}|\text{Tidak Laris}) = \frac{161}{1865} = 0,086$$

3.2 Perhitungan Algoritma Naive Bayes

Algoritma naive bayes pada penelitian ini dilaksanakan dengan bantuan Microsoft Excel dan RapidMiner agar proses pengolahan data menjadi lebih mudah dan cepat. Data penjualan produk makanan olahan diproses dan diklasifikasikan menggunakan algoritma Naive Bayes dan C4.5. Setiap data dianalisis guna melihat hasil klasifikasi serta mengetahui akurasi kinerja dari kedua algoritma. Proses perhitungan sesuai persamaan 1 dan 2 disajikan pada tabel 7 untuk sampel data baris 24 berikut.

Tabel 7. Perhitungan Algoritma Naive Bayes

Variabel	X C1	Kelas		
		Laris	Kurang Laris	Tidak Laris
Merk	Popularitas Tinggi	0,930000	0,706000	0,369000
Olahan	Daging	0,951000	0,852000	0,795000
Berat	Kecil	0,016000	0,000000	0,044000
Harga	Murah	0,112000	0,100000	0,404000
Terjual	Banyak	0,212000	0,000000	0,000000
P(X Ci) : P(X Kelas)		0,000336	0,000000	0,000000
P(X Ci)*P(Ci) = P(X Kelas) * P(kelas)		0,000061	0,000000	0,000000

Hasil rumus **P(X|Ci) : P(X|Kelas)** adalah hasil kali total dari seluruh nilai probabilitas variabel pada tiap-tiap data. Diketahui prediksi kelas untuk data pada baris 24 dengan produk **sosis**, variabel merk **popularitas tinggi**, olahan **daging**, berat produk **kecil**, harganya **murah**, dan terjual **banyak** :

$$\mathbf{P(X|Ci) : P(X|Kelas = Laris)}$$

$$= 0,930000 \times 0,951000 \times 0,016000 \times 0,112000 \times 0,212000 = 0,000336$$

$$\mathbf{P(X|Ci) : P(X|Kelas = Kurang Laris)}$$

$$= 0,706000 \times 0,852000 \times 0,000000 \times 0,100000 \times 0,000000 = 0,000000$$

$$\mathbf{P(X|Ci) : P(X|Kelas = Tidak Laris)}$$

$$= 0,369000 \times 0,795000 \times 0,044000 \times 0,404000 \times 0,000000 = 0,000000$$

Berdasarkan Tabel 7, terdapat beberapa probabilitas bernilai nol karena kombinasi data tertentu tidak ditemukan pada data latih. Hal ini dapat mempengaruhi hasil klasifikasi, sehingga digunakan metode *Laplace smoothing* dengan menambahkan nilai 1 pada setiap frekuensi agar probabilitas tetap dapat dihitung. Rumus yang digunakan sebagai berikut :

Sehingga diperoleh hasil yaitu:

$$\mathbf{P(X|Ci) : P(X|Kelas = Kurang Laris) \text{ sesudah Laplace Smoothing}}$$

$$= \frac{623+1}{883+5} \times \frac{752+1}{883+5} \times \frac{0+1}{883+5} \times \frac{88+1}{883+5} \times \frac{0+1}{883+5}$$

$$= 0,7027027027 \times 0,8479729730 \times 0,0011261261 \times 0,1002252252$$

$$\times 0,0011261261$$

$$= 0,00000000757$$

$$\mathbf{P(X|Ci) : P(X|Kelas = Tidak Laris) \text{ sesudah Laplace Smoothing}}$$

$$= \frac{688+1}{1865+5} \times \frac{1482+1}{1865+5} \times \frac{82+1}{1865+5} \times \frac{754+1}{1865+5} \times \frac{0+1}{1865+5}$$

$$\begin{aligned}
&= 0,3684491979 \times 0,7930481283 \times 0,0443850267 \times 0,4037433155 \\
&\quad \times 0,0005347594 \\
&= 0,00000280012
\end{aligned}$$

Kemudian hasil dikalikan dengan nilai **P(Kelas)** utama yakni,

$$P(X|Ci)*P(Ci) = P(X|Kelas = Laris) * P(Kelas = Laris)$$

$$= 0,0003359984 \times 0,183 = 0,00006148$$

$$P(X|Ci)*P(Ci) = P(X|Kelas = Kurang Laris) * P(Kelas = Kurang Laris)$$

$$= 0,00000000757 \times 0,263 = 0,0000000019384$$

$$P(X|Ci)*P(Ci) = P(X|Kelas = Tidak Laris) * P(Kelas = Tidak Laris)$$

$$= 0,00000280012 \times 0,555 = 0,0000015547$$

Hasil perhitungan membuktikan bahwa Nilai **P(X|Ci)*P(Ci) = P(X|Kelas = Laris) * P(Kelas = Laris)** lebih besar artinya output pada data 24 adalah **Laris**.

3.3 Perhitungan Algoritma C4.5

Langkah awal dari perhitungan algoritma C4.5 adalah mencari nilai *Gain* dan *Entropy* seperti pada berikut ini.

1. Menentukan nilai *entropy* dilakukan dengan menghitung jumlah data pada setiap kelas berdasarkan keseluruhan data latih. Dari total 3362 data latih, diketahui data dengan kategori Laris sebanyak 614, Kurang Laris sebanyak 883, dan Tidak Laris sebanyak 1865. Sehingga dilakukan proses perhitungan dengan persamaan 3.

$$\begin{aligned}
\text{Entropy Total} &= (-614/3362 * \log_2(614/3362)) + (-883/3362 * \log_2(883/3362)) + \\
&\quad (-1865/3362 * \log_2(1865/3362)) \\
&= 1,426
\end{aligned}$$

2. Perhitungan nilai *entropy* dan *gain* pada masing-masing variabel, yaitu merk, olahan, berat, harga, dan terjual. Perhitungan dilakukan untuk variabel mana yang berpengaruh dalam proses klasifikasi, dapat digunakan sebagai dasar pembentukan pohon keputusan pada algoritma C4.5. Dapat dilihat pada tabel 8 berikut.

Tabel 8. Perhitungan Nilai *Gain*

	Jumlah	Laris	Kurang Laris	Tidak Laris	Entropy	Gain
Merk						0,190
Popularitas Tinggi	1882	571	623	688	1,581	
Popularitas Sedang	1234	39	260	935	0,934	
Popularitas Rendah	246	4	0	242	0,120	

Olahan							0,023
	Daging	2818	584	752	1482	1,467	
	Seafood	412	21	90	301	1,029	
	Sayur	132	9	41	82	1,215	
Berat							0,117
	Besar	2397	341	478	1578	1,261	
	Sedang	873	263	405	205	1,526	
	Kecil	92	10	0	82	0,496	
Harga							0,268
	Mahal	1007	301	578	128	1,359	
	Sedang	1444	244	217	983	1,222	
	Murah	911	69	88	754	0,834	
Terjual							0,467
	Banyak	130	130	0	0	0,000	
	Sedang	996	484	351	161	1,461	
	Sedikit	2236	0	532	1704	0,792	

3.4 Pengujian Dengan Algoritma Naive Bayes

Berdasarkan hasil pengujian *Confusion Matrix* menggunakan *Cross Validation* pada algoritma Naive Bayes dengan *number of folds* 7, 5, dan 3, diperoleh tingkat akurasi yang tidak jauh berbeda, yaitu sebesar 90,90%, 91,23%, dan 91,17% pada gambar berikut.

accuracy: 91.23% +/- 1.88% (micro average: 91.23%)

	true Kurang Laris	true Tidak Laris	true Laris	class precision
pred. Kurang Laris	630	36	1	94.45%
pred. Tidak Laris	237	1824	0	88.50%
pred. Laris	16	5	613	96.69%
class recall	71.35%	97.80%	99.84%	

Gambar 1. Akurasi Algoritma Naive Bayes *Number Of Folds* 7

accuracy: 91.17% +/- 1.08% (micro average: 91.17%)

	true Kurang Laris	true Tidak Laris	true Laris	class precision
pred. Kurang Laris	634	36	1	94.49%
pred. Tidak Laris	237	1818	0	88.47%
pred. Laris	12	11	613	96.38%
class recall	71.80%	97.48%	99.84%	

Gambar 2. Akurasi Algoritma Naive Bayes *Number Of Folds* 5

accuracy: 90.90% +/- 1.34% (micro average: 90.90%)

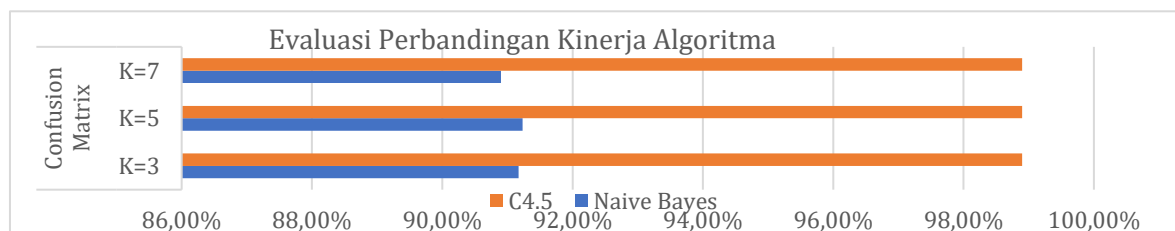
	true Kurang Laris	true Tidak Laris	true Laris	class precision
pred. Kurang Laris	624	36	1	94.40%
pred. Tidak Laris	237	1819	0	88.47%
pred. Laris	22	10	613	95.04%
class recall	70.67%	97.53%	99.84%	

Gambar 3. Akurasi Algoritma Naive Bayes *Number Of Folds* 3

Gambar 7. Pohon Keputusan Algoritma C4.5

3.6 Perbandingan Algoritma Naive Bayes Dan Algoritma C4.5

Berdasarkan hasil implementasi dari algoritma naive bayes dan C4.5 dengan menggunakan *tools* RapidMiner Versi 10.3.1 diperoleh sebagai berikut.



Gambar 7. Evaluasi Hasil Perbandingan Kedua Algoritma

Algoritma C4.5 memperoleh nilai *accuracy* yang lebih tinggi karena mampu membentuk pohon keputusan berdasarkan variabel yang paling berpengaruh. Sementara itu, algoritma Naive Bayes mengasumsikan setiap variabel bersifat independen, sehingga hubungan antar variabel tidak diperhitungkan menyebabkan akurasi lebih rendah dan memiliki nilai *accuracy* yang bervariasi pada setiap pengujian *K-Fold Cross Validation*.

4. SIMPULAN

Berdasarkan penelitian yang dilakukan, klasifikasi produk makanan olahan pada Y3 *Frozen Food* menggunakan algoritma Naive Bayes dengan metode tambahan *Laplace Smoothing* dan C4.5 berhasil mengelompokkan 3362 data transaksi ke dalam tiga kategori yaitu laris, kurang laris, dan tidak laris, dengan unit terjual sebagai faktor penentu utama. Dari perbandingan kedua algoritma, C4.5 terbukti lebih unggul dengan akurasi stabil sebesar 98,90% dibandingkan Naive Bayes yang menghasilkan akurasi bervariasi di kisaran 90,90–91,23%, sehingga C4.5 lebih direkomendasikan sebagai algoritma klasifikasi produk terlaris di Y3 *Frozen Food*.

REFERENCES

- [1] R. Sundari, "Preferensi Konsumen Dalam Membeli Produk Makanan Beku / Frozen Food Di Kota Pekanbaru (Studi Kasus Comel Frozen Food)," *J. Bisnis Kompetif*, Vol. 2, No. 3, Hal. 203–209, 2023.
- [2] Uli, R. G. Y., & Moentamaria, D. (2024). Pengaruh Kadar Air Terhadap Masa Simpan Olahan Pangan Dengan Teknologi Sterilisasi Suhu Tinggi. *Distilat Jurnal Teknologi Separasi*, 10(9), 849–858.
- [3] Daulay, A., Zaki, M. I., Hidayat, M., & Febriansyah, A. (N.D.). *Derivatif: Jurnal Manajemen Ekonomi Dan Akuntansi Yayasan Salmiah Education Global International (Ysegi) Peran Umkm Dalam Menggerakkan Pembangunan Ekonomi Lokal Dan Meningkatkan Kesejahteraan Masyarakat*. 23–32.
- [4] Nursyahfitri, R., Maharadja, A. N., Farissa, R. A., & Umaidah, Y. (2021). Klasifikasi Penentuan Jenisobat Menggunakan Algoritma Decision Tree. *Jip (Jurnal Informatika Polinema)*, 7(3), 53–60.
- [5] M. Sadikin, R. Rosnelly, Roslina, T. S. Gunawan, Dan Wanayumini, "Perbandingan Tingkat Akurasi Klasifikasi Penerimaan Dosen Tetap Menggunakan Metode Naive Bayes Classifier Dan C4 . 5," *J. Media Inform. Budidarma*, Vol. 4, No. 4, Hal. 1100–1109, 2020, Doi: 10.30865/Mib.V4i4.2434.
- [6] R. Pratama, B. Huda, E. Novalia, Dan H. Kabir, "Perbandingan Algoritma C4 . 5 Dan Naïve Bayes Dalam

-
- Menentukan Persediaan Stok,” *Media Teknol. Inf. Dan Komput. J.*, Vol. 6, No. 2, Hal. 115–122, 2022, Doi: 10.47002/Metik.V6i2.379.
- [7] M. K. Al Fatah Dan A. Wibowo, “Perbandingan Metode Algoritma C4.5 Dan Naive Bayes Untuk Memprediksi Penjualan Kosmetik Pada Toko Jelita,” *J. Mnemon. (Jurnal Tek. Inform. Dari Inst. Teknol. Nas. Malang)*, Vol. 7, No. 2, Hal. 220–225, 2024.
- [8] Macfud, A. Z., Kusuma, A. P., & Puspitasari, W. D. (2023). *Analisis Algoritma Naive Bayes Classifier (Nbc) Pada Klasifikasi Tingkat Minat Barang Di Toko Violet Cell*. 7(1), 87–94.
- [9] N. S. Fauzia Dan R. D. Dana, “Implementasi Algoritma Naive Bayes Dalam Klasifikasi Status Kesejahteraan Masyarakatdesa Gunungsari,” *Blend Sains J. Tek.*, Vol. 1, No. 4, Hal. 295–305, 2023.
- [10] D. Irfan, P. Ramadani, I. N. Fitriyani, R. S. Damanik, Dan Y. I. Mahenda, “Prediksi Risiko Mahasiswa Mengulang Mata Kuliah Menggunakan Algoritma Naïve Bayes Jurnal Media Informatika [Jumin],” Vol. 5, Hal. 271–278, 2024.
- [11] I. P. Agus, E. Darma, Dan I. G. I. Sudipa, “Sentimen Analisis Inisiatif Vaksin Nasional Menggunakan Naïve Bayes Dan Laplacian Smoothing Pada Komentar Video Youtube,” *J. Resist. (Rekayasa Sist. Komput.)*, Vol. 5, No. 2, Hal. 116–126, 2022.
- [12] Barito, E. E., Beng, J. T., & Arisandi, D. (2022). Penerapan Algoritma C4.5 Untuk Klasifikasi Mahasiswa Penerima Bantuan Sosial Covid-19. *Jurnal Ilmu Komputer Dan Sistem Informasi*, 10, 1–8.
- [13] Nurhalimah Dan Nurhasanah, “Penerapan Metode Naïve Bayes Pada Pengaruh Penggunaan Gadget Terhadap Nilai Siswa Sekolah Dasar,” *Fakt. Exacta*, Vol. 18, No. 1, Hal. 72–82, 2025.