

Dwi Vernanda<sup>1</sup>, Zahrani Nurizati<sup>2</sup>, Arif Hidayat<sup>3</sup>

<sup>1</sup>Politeknik Negeri Subang, nanda@polsub.ac.id

<sup>2</sup>Politeknik Negeri Subang, zahra209912@gmail.com

<sup>3</sup>Politeknik Negeri Subang, arf31371@gmail.com

\*)Korespondensi: nanda@polsub.ac.id

### Abstrak

Buah nanas merupakan salah satu komoditas buah tropis yang dihasilkan di negara Indonesia, salah satu daerah penghasil buah nanas yaitu Kabupaten Subang, Jawa Barat. Saat ini nanas yang dihasilkan untuk memenuhi berbagai macam kebutuhan yaitu pasar tradisional, supermarket, pabrik pengolahan buah nanas, hingga dijadikan oleh-oleh Kabupaten Subang. Penelitian ini bertujuan untuk mengklasifikasikan nanas layak jual menggunakan metode Naïve Bayes, yang dikenal sederhana dan efektif dalam pengolahan data probabilistik. Penelitian ini menggunakan data primer yang mencakup karakteristik fisik nanas seperti warna kulit dan tingkat kemanisan (Brix). Data tersebut diproses melalui beberapa tahap, termasuk praproses dan klasifikasi dengan Naïve Bayes. Hasil penelitian menunjukkan tingkat akurasi klasifikasi sebesar 94%, dengan performa yang memuaskan pada *recall* dan *precision*. Keterbatasan metode ini adalah asumsi independensi antar fitur, yang memengaruhi akurasi dalam kasus tertentu. Penelitian ini memberikan kontribusi signifikan pada penerapan kecerdasan buatan dalam sektor pertanian dan membuka peluang untuk dikembangkan pada komoditas buah lain.

**Kata Kunci:** Akurasi, Klasifikasi Nanas, Naïve Bayes.

### Abstract

*Pineapple is one of the tropical fruit commodities produced in Indonesia, one of the pineapple producing areas is Subang Regency, West Java. Currently, pineapples are produced to meet various needs, namely traditional markets, supermarkets, pineapple processing factories, and even souvenirs from Subang Regency. This study aims to classify pineapples that are suitable for sale using the Naïve Bayes method, which is known to be simple and effective in processing probabilistic data. This study uses primary data that includes the physical characteristics of pineapples such as skin color and sweetness level (Brix). The data is processed through several stages, including preprocessing and classification with Naïve Bayes. The results of the study showed a classification accuracy level of 94%, with satisfactory performance on recall and precision. The limitation of this method is the assumption of independence between features, which affects accuracy in certain cases. This study makes a significant contribution to the application of artificial intelligence in the agricultural sector and opens up opportunities for development in other fruit commodities.*

**Keywords:** Accuracy, Pineapple Classification, Naïve Bayes.

## I. PENDAHULUAN

Buah nanas merupakan salah satu komoditas buah tropis yang memiliki potensi ekonomi tinggi di berbagai negara, termasuk Indonesia. Indonesia sebagai negara agraris memiliki tanah yang subur dan iklim yang mendukung untuk pertumbuhan berbagai jenis tanaman, termasuk nanas. Salah satu daerah penghasil buah nanas yaitu Kabupaten Subang, nanas yang dihasilkan untuk memenuhi berbagai kebutuhan seperti pasar tradisional, supermarket, pabrik pengolahan buah nanas, hingga sebagai ikon oleh-oleh Kabupaten Subang. Buah nanas sendiri memiliki berbagai manfaat kesehatan dan nilai komersial yang membuatnya populer di pasar lokal dan internasional. Namun, dalam praktiknya, penjualan nanas di pasaran masih sering menghadapi tantangan terkait kualitas buah yang layak

jual. Hal ini disebabkan oleh variasi dalam kualitas buah nanas yang dipengaruhi oleh berbagai faktor seperti ukuran, warna, tingkat kematangan, dan kebersihan buah [1]. Untuk itu, diperlukan metode yang efektif dan efisien dalam menentukan kelayakan nanas untuk dijual di pasaran [2]

Penilaian kelayakan buah nanas secara manual oleh para petani atau pengepul sering kali memerlukan waktu dan ketelitian yang tinggi. Proses manual ini sangat bergantung pada keahlian subjektif penilai, yang tentu saja bisa berbeda antara satu individu dengan individu lainnya. Oleh karena itu, pendekatan berbasis teknologi menjadi solusi yang relevan untuk menyelesaikan masalah ini. Salah satu teknologi yang dapat digunakan adalah metode klasifikasi berbasis kecerdasan buatan, seperti *Naïve Bayes* [3].

Penerapan model klasifikasi pada sebuah robot berbasis *machine learning* untuk membantu pemanenan pohon palem, model klasifikasi menggunakan *convolutional neural network* (CNN) dan memanfaatkan perangkat keras yaitu kamera, akurasi pengujian model klasifikasi mencapai 98.89%. Akurasi pengujian didapatkan dari model AlexNet, RestNet-50, *Efficient Net*, dan *Inception-v3* [4].

Penelitian sebelumnya pernah dilakukan untuk memilah buah papaya, hal ini dikarenakan selama proses pemilahan buah papaya yang dilakukan secara manual memerlukan waktu yang lama, hal ini juga menjadi tidak akurat karena hanya berdasarkan persepsi visual manusia. Penelitian ini menggunakan dua pendekatan yaitu *machine learning* dan *transfer learning*. Dari 300 gambar buah papaya yang dijadikan data latih didapatkan tingkat akurasi 100% dengan waktu latih 0.0995 detik [5].

*Naïve Bayes* adalah salah satu metode klasifikasi yang sering digunakan dalam pengolahan data, khususnya dalam pemodelan probabilistik. Metode ini bekerja berdasarkan *Teorema Bayes* dengan asumsi adanya independensi antara fitur-fitur yang digunakan. *Naïve Bayes* dikenal sebagai metode yang sederhana namun cukup kuat dan cepat dalam melakukan klasifikasi, terutama ketika berhadapan dengan data yang bersifat diskret. Oleh karena itu, penerapan metode ini dalam mengklasifikasikan nanas yang layak jual atau tidak sangatlah relevan dan menarik untuk diteliti [6].

Pada dasarnya, proses klasifikasi menggunakan *Naïve Bayes* melibatkan dua tahap utama, yaitu pelatihan dan pengujian. Pada tahap pelatihan, algoritma akan membangun model klasifikasi berdasarkan data pelatihan yang telah dilabeli, di mana setiap data berisi fitur-fitur yang menggambarkan karakteristik nanas, seperti ukuran, berat, warna kulit, tingkat kematangan, serta kondisi fisik lainnya [7]. Setelah model terbentuk, pada tahap pengujian, model akan digunakan untuk mengklasifikasikan nanas baru berdasarkan fitur-fitur yang ada, sehingga dapat menentukan apakah nanas tersebut layak dijual atau tidak.

Pemanfaatan metode *Naïve Bayes* dalam klasifikasi nanas layak jual memiliki beberapa kelebihan. Pertama, metode ini sederhana dan mudah diimplementasikan, sehingga cocok untuk diterapkan dalam industri pertanian yang umumnya masih menggunakan teknologi sederhana. Kedua, *Naïve Bayes* dapat menangani data dalam jumlah besar dengan cukup efisien. Ketiga, metode ini juga relatif cepat dalam memproses data, sehingga memungkinkan untuk diterapkan dalam sistem yang memerlukan kecepatan dalam pengambilan keputusan, misalnya dalam proses penyortiran buah nanas di tempat produksi sebelum distribusi ke pasar [8].

Namun demikian, metode *Naïve Bayes* juga memiliki beberapa keterbatasan yang perlu

dipertimbangkan. Salah satu kelemahan utama dari metode ini adalah asumsi independensi antar fitur. Dalam kasus nyata, sering kali fitur-fitur yang ada memiliki ketergantungan satu sama lain. Misalnya, warna kulit nanas mungkin berkorelasi dengan tingkat kematangan buah [9]. Dalam situasi seperti ini, asumsi independensi yang diadopsi oleh *Naïve Bayes* dapat mengurangi akurasi klasifikasi. Selain itu, metode ini juga cenderung kurang optimal ketika berhadapan dengan data yang sangat kompleks atau data dengan distribusi yang tidak normal.

Untuk mengatasi kelemahan tersebut, beberapa pendekatan dapat dilakukan. Salah satunya adalah dengan melakukan pra-proses data secara cermat sebelum diterapkan pada algoritma *Naïve Bayes*. Pra-proses data ini bisa meliputi normalisasi fitur, pemilihan fitur yang paling relevan, serta pengelompokan data ke dalam kategori yang sesuai. Selain itu, *Naïve Bayes* juga bisa dikombinasikan dengan metode lain, seperti *decision tree* atau *support vector machine* (SVM), untuk meningkatkan akurasi hasil klasifikasi. Dengan kombinasi metode yang tepat, performa model klasifikasi dapat ditingkatkan, sehingga mampu menghasilkan keputusan yang lebih akurat dan dapat diandalkan.

Penerapan metode *Naïve Bayes* dalam klasifikasi nanas layak jual juga memberikan manfaat jangka panjang bagi para pelaku industri pertanian, khususnya petani dan pengepul nanas. Dengan adanya sistem otomatis yang mampu mengklasifikasikan buah nanas secara cepat dan akurat, proses distribusi nanas ke pasar dapat berjalan lebih efisien. Hal ini pada akhirnya dapat meningkatkan keuntungan bagi para petani dan pengepul, sekaligus memastikan konsumen mendapatkan buah nanas yang berkualitas tinggi. Di sisi lain, penggunaan teknologi ini juga mendorong modernisasi dalam sektor pertanian, yang selama ini dikenal masih cukup tradisional, terutama di negara-negara berkembang seperti Indonesia [10].

Selain itu, penelitian mengenai klasifikasi nanas layak jual menggunakan metode *Naïve Bayes* juga dapat menjadi referensi bagi pengembangan sistem serupa pada komoditas buah lainnya. Mengingat setiap jenis buah memiliki karakteristik fisik yang berbeda, penelitian lanjutan dapat dilakukan untuk mengadaptasi metode ini agar sesuai dengan kebutuhan komoditas lain. Misalnya, klasifikasi kelayakan buah mangga, jeruk, atau apel. Dengan begitu, potensi penggunaan kecerdasan buatan dalam sektor pertanian akan semakin luas dan bermanfaat bagi berbagai komoditas penting lainnya [11].

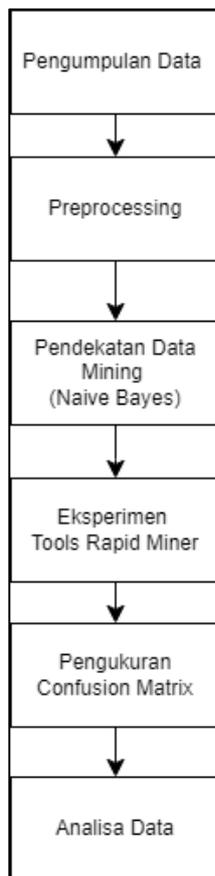
Seperti halnya pada penelitian mengenai klasifikasi kualitas buah jeruk berdasarkan ada tidaknya penyakit pada proses pascapanen buah jeruk. Menggunakan *machine learning* dan metode klasifikasi didapatkan model untuk klasifikasi sehingga dapat menjaga kualitas buah jeruk [12]. Pada proses pemanenan buah stroberi dibutuhkan suatu model klasifikasi yang dapat

menentukan kualitas buah stroberi yaitu dari segi warna, tingkat kemanisan, dan bentuk buahnya. Hasil penelitian ini berupa model klasifikasi tingkat kematangan buah stroberi dan dapat ditanamkan pada sistem penglihatan robot [13]

Secara keseluruhan, penerapan metode *Naïve Bayes* dalam klasifikasi nanas layak jual merupakan langkah penting dalam upaya memanfaatkan teknologi untuk mendukung sektor pertanian. Dengan kemampuannya yang sederhana namun efektif, *Naïve Bayes* menjadi alat yang berharga untuk membantu para petani dan pengepul dalam mengoptimalkan proses penjualan nanas. Meskipun memiliki beberapa kelemahan, metode ini tetap menawarkan solusi yang menarik dan dapat dikembangkan lebih lanjut agar lebih akurat dan efisien di masa mendatang. Penelitian lebih lanjut di bidang ini juga dapat membuka peluang untuk mengaplikasikan teknologi serupa pada komoditas pertanian lainnya, sehingga meningkatkan daya saing produk pertanian di pasar global [14].

## II. METODE

Penelitian ini menggunakan metode *Naïve Bayes* dengan melakukan tahapan penelitian seperti pada gambar 1 sebagai berikut.



Gambar 1 Metode penelitian

### 2.1 Tahapan Pengumpulan Data

Tahapan ini melakukan pengumpulan data awal yang berupa data primer untuk mendapatkan informasi awal tentang data nanas untuk mendapatkan kelayakan penjualan nanas yang dimana terdapat beberapa faktor diantaranya tingkat kematangan, harga, kemanisan dsb.

Beberapa faktor utama yang menjadi fokus dalam pengumpulan data tersebut antara lain tingkat kematangan buah nanas, harga jual di pasaran, tingkat kemanisan yang dicapai, serta aspek-aspek lain yang memengaruhi kualitas dan preferensi konsumen.

Tingkat kematangan menjadi penting karena buah yang terlalu matang atau belum matang dapat mempengaruhi kepuasan pembeli. Sementara itu, harga jual nanas harus dianalisis dalam konteks persaingan pasar dan daya beli masyarakat, sehingga penjual bisa menentukan strategi harga yang tepat. Faktor kemanisan juga menjadi indikator kualitas yang dipertimbangkan konsumen, di mana kadar manis yang ideal akan meningkatkan daya tarik nanas di pasaran. Melalui pengumpulan data primer yang komprehensif ini, berbagai faktor tersebut dapat dievaluasi dengan lebih mendetail, sehingga menghasilkan analisis kelayakan penjualan yang akurat.

### 2.2 Tahapan *Preprocessing* Data

Pada tahap ini, data yang akan digunakan diproses terlebih dahulu. Langkah-langkah utama meliputi pembersihan data (mengatasi data yang hilang atau salah), transformasi data (normalisasi atau encoding variabel), serta reduksi data (menghapus fitur yang tidak relevan) agar siap digunakan oleh model.

### 2.3 Tahapan Pendekatan Data Mining (*Naïve Bayes*)

*Naïve Bayes* merupakan algoritma berbasis probabilitas yang digunakan untuk klasifikasi. Pada tahapan ini, data dilatih menggunakan teorema Bayes, di mana prediksi dilakukan berdasarkan probabilitas tertinggi dari tiap kelas berdasarkan fitur yang ada [15]

Proses pelatihan ini dilakukan dengan cara menghitung probabilitas dari setiap kelas yang tersedia, berdasarkan fitur-fitur atau variabel yang ada dalam data. *Naïve Bayes* mengasumsikan bahwa setiap fitur bersifat independen satu sama lain, meskipun dalam kenyataannya bisa terdapat hubungan di antara fitur-fitur tersebut. Meskipun asumsi ini sering kali tidak sepenuhnya akurat, pendekatan ini terbukti sangat efektif dalam banyak kasus klasifikasi sederhana [16].

Selama pelatihan, algoritma ini akan mengidentifikasi pola pada data untuk menentukan kelas mana yang memiliki probabilitas tertinggi berdasarkan kombinasi fitur-fitur yang ada. Prediksi dibuat dengan memilih kelas dengan probabilitas tertinggi untuk setiap sampel, yang memberikan hasil klasifikasi secara cepat dan efisien. Tahapan ini memungkinkan praktisi data untuk membangun model

prediktif yang mampu mengelompokkan data baru ke dalam kategori-kategori yang telah ditentukan, menjadikan Naïve Bayes pilihan yang optimal dalam berbagai aplikasi klasifikasi, terutama ketika waktu komputasi dan kemudahan implementasi menjadi prioritas.

## 2.4 Tahapan Eksperimen Tools Rapid Miner

Pada tahap ini, RapidMiner digunakan untuk membangun, menjalankan, dan mengevaluasi model data *mining*. Data diimpor, proses *preprocessing* dan klasifikasi dibuat, eksperimen dijalankan, dan hasil dievaluasi berdasarkan metrik performa seperti akurasi dan *precision* [17]

## 2.5 Tahapan Pengukuran Confusion Matrix

Tahap ini melibatkan evaluasi model klasifikasi dengan menggunakan confusion matrix. Matriks ini menghitung nilai *True Positive*, *False Positive*, *True Negative*, dan *False Negative*, yang kemudian digunakan untuk mengukur akurasi, *precision*, *recall*, dan *F1-score* dari model [18]

Tahap ini mencakup proses evaluasi model klasifikasi melalui penggunaan *confusion matrix*, yaitu matriks yang berfungsi untuk memberikan gambaran menyeluruh mengenai performa model. *Confusion matrix* ini menghitung empat kategori utama dalam hasil prediksi, yaitu *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN). Setiap kategori memiliki arti penting yang membantu dalam menilai seberapa baik model dalam memprediksi hasil secara benar maupun salah.

Nilai *True Positive* (TP) menunjukkan jumlah kasus di mana model secara benar mengidentifikasi sampel sebagai positif ketika sampel tersebut memang positif. *False Positive* (FP), di sisi lain, adalah jumlah kasus di mana model keliru mengklasifikasikan sampel negatif sebagai positif. *True Negative* (TN) menghitung jumlah kasus di mana model benar dalam mengidentifikasi sampel negatif, sedangkan *False Negative* (FN) menghitung jumlah kesalahan model dalam mengklasifikasikan sampel positif sebagai negatif.

## 2.6 Tahapan Analisa Data

Tahapan analisis data adalah langkah penting yang dilakukan setelah memperoleh hasil dari model, di mana fokus utama diarahkan pada proses interpretasi dan analisis mendalam terhadap data dan keluaran model. Pada tahap ini, pola-pola yang muncul dalam data dieksplorasi dan diperiksa secara menyeluruh untuk memahami karakteristik utama yang mungkin memengaruhi hasil model. Analisis ini melibatkan pengamatan terhadap tren, anomali, dan hubungan antarvariabel yang mungkin terlewatkan dalam tahap pemodelan awal.

Selain itu, performa model juga dievaluasi dengan menggunakan berbagai metrik dan alat evaluasi untuk menilai tingkat keakuratannya serta relevansi prediksi yang dihasilkan. Kinerja model dinilai secara detail

untuk memastikan bahwa hasilnya sesuai dengan tujuan analisis dan bahwa model mampu bekerja dengan baik pada data yang ada. Dalam proses ini, baik kekuatan maupun kelemahan model diidentifikasi secara menyeluruh. Kekuatan model mencakup area di mana model berhasil memberikan hasil yang akurat, sementara kelemahannya mencakup kesalahan atau ketidakakuratan dalam prediksi, serta keterbatasan dalam menangani variasi data.

## Metode Perhitungan Naïve Bayes

*Naive Bayes* adalah sebuah algoritma klasifikasi dalam machine learning yang didasarkan pada teorema Bayes, dengan asumsi sederhana bahwa setiap fitur atau variabel dalam data bersifat independen satu sama lain. Algoritma ini sering digunakan untuk memprediksi kategori atau kelas dari suatu data berdasarkan nilai-nilai fitur atau atribut yang dimiliki oleh data tersebut [19]. Pada intinya, *Naive Bayes* mencoba menghitung probabilitas suatu kelas tertentu berdasarkan fitur-fitur yang diamati, dan memilih kelas dengan probabilitas tertinggi sebagai hasil prediksi. Algoritma ini bekerja dengan menghitung probabilitas bersyarat dari setiap fitur terhadap kelas tertentu, dan kemudian mengalikan semua probabilitas tersebut untuk mendapatkan prediksi yang paling mungkin.

Dalam prosesnya rumus probabilitas dijelaskan pada persamaan 1, sebagai berikut:

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

Keterangan:

- X : Data dengan class yang belum diketahui
- H : Hipotesis data merupakan suatu class spesifik
- P(H|X) : Probabilitas hipotesis terhadap H berdasar kondisi X (posteriori probabilitas)
- P(H) : Probabilitas hipotesis terhadap H (prior probabilitas)
- P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H
- P(X) : Probabilitas X

Proses klasifikasi dalam menentukan kelas apa yang cocok terhadap sample yang akan dianalisis memerlukan sebuah langkah yang dapat digambarkan pada persamaan 2 berikut:

$$P(C|F1 \dots Fn) = \frac{P(C)P(F1 \dots Fn|C)}{P(F1 \dots Fn)}$$

Variabel C menggambarkan kelas, sementara variabel F1 ... Fn menjelaskan karakteristik yang digunakan untuk proses klasifikasi. Rumus ini merepresentasikan karakteristik tertentu dalam kelas C (Posterior), yaitu probabilitas terjadinya kelas C yang dikalikan dengan probabilitas kemunculan karakteristik sampel pada kelas C (disebut juga

likelihood). Oleh karena itu, rumus tersebut dapat disederhanakan menjadi persamaan 3.

$$\text{Posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

Multinomial *Naive Bayes*, atau yang dikenal sebagai *Naive Bayes Classifier*, adalah model yang merupakan penyederhanaan dari Teorema *Bayes*. Algoritma *Naive Bayes* mengasumsikan bahwa pengaruh nilai suatu variabel dalam sebuah kelas tidak saling berkaitan dengan variabel lain. Proses ini dapat dijelaskan lebih rinci dengan menguraikan nilai  $P(C|F_1, \dots, F_n)$  menggunakan aturan perkalian, seperti sebagai berikut:

$$\begin{aligned} P(C|F_1, \dots, F_n) &= P(C)P(F_1, \dots, F_n|C) \\ &= P(C)P(F_1|C)P(F_2, \dots, F_n|C, F_1) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3, \dots, F_n|C, F_1, F_2) \\ &= (C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2)P(F_4, \dots, F_n|C, F_1, F_2, F_3) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1}) \end{aligned}$$

Dapat disimpulkan bahwa hasil dari penjelasan rumus Bayes di atas menunjukkan bahwa semakin tinggi kompleksitas serta syarat-syarat yang mempengaruhi nilai probabilitas, semakin sulit untuk menganalisis setiap data satu per satu. Hal ini menyebabkan perhitungan probabilitas menjadi semakin rumit [20]. Oleh karena itu, persamaan berikut dapat diterapkan:

$$P(F_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i)$$

Dimana  $i \neq j$ , sehingga di dapatkan:

$$P(F_i|C, F_j) = P(F_i|C)$$

Persamaan tersebut adalah model yang akan diterapkan dalam pemrosesan pengklasifikasian. Untuk klasifikasi dengan data yang berlanjut kontinyu dapat menggunakan rumus Densitas Gaus:

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

Keterangan:

- P : Peluang
- $X_i$  : Atribut terhadap  $i$
- $X_i$  : Nilai atribut berdasarkan  $i$
- Y : Kelas yang akan dicari
- $Y_i$  : Sub kelas Y yang akan dicari
- $\mu$  : mean, mendefenisikan rata – rata terhadap seluruh atribut
- $\sigma$  : Deviasi standar, menyatakan varian terhadap seluruh atribut.

### III. HASIL DAN PEMBAHASAN

#### 3.1 Pengumpulan Data

Data yang diperoleh merupakan data primer dari buah nanas yang terdiri dari 100 data. Buah nanas yang

digunakan berasal dari satu wilayah yaitu Kec. Cujambe, Kabupaten Subang, sehingga 100 buah nanas yang dijadikan data penelitian merupakan buah yang memiliki jenis yang sama serta penanaman maupun pemanen dalam periode waktu yang sama.

Pengumpulan data dilakukan dengan cara mengambil nilai *red*, *green*, and *blue* dari masing-masing kulit buah nanas. Kulit buah nanas difoto menggunakan kamera Sony A6400, seperti pada Gambar 2. Foto Buah Nanas.



Gambar 2. Foto Buah Nanas

Model klasifikasi buah nanas yang akan dibentuk berdasarkan 2 variabel utama yaitu warna kulit buah nanas dan tingkat kemanisan buah nanas, untuk itu dari 100 buah nanas yang telah diambil nilai RGB selanjutnya proses penentuan tingkat kemanisan masing-masing buah nanas.

Pengukuran tingkat kemanisan buah nanas yaitu dengan cara mengambil sari buah nanas yang berada 5 cm dari pangkal buah nanas, selanjutnya sari buah nanas tersebut dites menggunakan alat *refractometer*. Berikut gambar pengambil sari buah nanas dan pengukuran tingkat kemanisan buah nanas.



Gambar 3. Pengambilan Sari Buah Nanas



Gambar 4. Pengukuran Tingkat Kemanisan Nanas

### 3.2 Preprocessing Data

Berikut dataset nanas yang telah diambil berupa nilai *red*, *green*, dan *blue*, serta nilai tingkat kemanisan buah nanas (*brix*).

	A	B	C	D	E	F
1	Nanas	Red	Green	Blue	Brix	Label
2	1	136	126	129	15,3	A
3	2	120	150	154	13,2	B
4	3	102	118	125	14,7	A
5	4	85	118	119	13,7	B
6	5	110	139	136	5,3	C
7	6	74	105	109	14,1	A
8	7	109	136	146	5,6	C
9	8	113	145	146	6,4	C
10	9	112	90	95	9,8	C
11	10	106	112	98	12	B
12	11	129	148	150	16,1	A
13	12	86	128	144	7,9	C
14	13	97	112	156	9,2	C
15	14	111	107	125	11,6	B
16	15	142	151	114	16,8	A
17	16	126	134	118	14,5	A
18	17	140	137	120	17	A
19	18	96	104	117	12,3	B
20	19	112	129	95	10,4	B

Gambar 2 Dataset

Dataset ini berisi data buah nanas dengan karakteristik fisik dan kimiawi. Setiap baris mewakili satu sampel nanas, dan kolom-kolomnya mencakup:

- 1 *Red*, *Green*, *Blue* (RGB): Kolom-kolom ini mewakili intensitas warna dalam skala RGB (*Red*, *Green*, *Blue*) yang mengindikasikan karakteristik warna pada setiap nanas. Nilai numerik di kolom "*Red*" menunjukkan intensitas warna merah, kolom "*Green*" menunjukkan intensitas warna hijau, dan kolom "*Blue*" menunjukkan intensitas warna biru. Informasi RGB ini dapat digunakan untuk mengklasifikasikan atau mengelompokkan nanas berdasarkan karakteristik warnanya, yang mungkin berhubungan dengan tingkat kematangan atau kualitas buah.

- 2 *Brix*: Kolom ini mencantumkan nilai *Brix* untuk setiap sampel nanas, yaitu ukuran kandungan gula dalam buah nanas yang dinyatakan dalam persentase. Nilai *Brix* memberikan informasi tentang tingkat kemanisan buah; semakin tinggi nilai *Brix*, semakin manis buah tersebut. Pengukuran ini sangat penting dalam industri buah, karena nilai *Brix* yang tinggi biasanya diinginkan oleh konsumen dan dianggap sebagai indikator kualitas buah yang baik.
- 3 *Label*: Kolom terakhir, yaitu kolom "*Label*," menunjukkan kategori atau kelas untuk setiap nanas. Label ini tampaknya berupa kategori kualitatif yang mungkin menandakan klasifikasi buah berdasarkan karakteristik tertentu, yaitu tingkat kelayakan (A, B, atau C).
  - Label A memiliki ciri fisik yaitu berat buah nanas 1-2 kg dan nilai *brix* 16-17.
  - Label B memiliki ciri fisik yaitu berat kurang dari 1 kg atau lebih dari 2 kg dan nilai *brix* ada pada 5-7.
  - Sedangkan label C memiliki ciri fisik yaitu berat kurang dari 500 gr dan nilai *brix* yang tidak menentu.

Label ini dapat berfungsi sebagai variabel target dalam analisis klasifikasi, sehingga model pembelajaran mesin dapat digunakan untuk memprediksi kategori atau label nanas baru berdasarkan data fitur yang ada.

- 4 Secara keseluruhan, dataset ini menyajikan informasi penting mengenai fitur-fitur warna (*Red*, *Green*, *Blue*), tingkat kemanisan (*Brix*), dan kategori kelas (*Label*) untuk berbagai sampel nanas. Dataset ini berpotensi digunakan dalam analisis data atau pembelajaran mesin untuk mengklasifikasikan nanas berdasarkan warna dan tingkat kemanisan. Misalnya, model *Naïve Bayes* atau model lain dapat dilatih untuk mengenali pola dalam data, sehingga mampu mengelompokkan nanas baru ke dalam kategori yang relevan berdasarkan atribut-atribut yang tersedia.

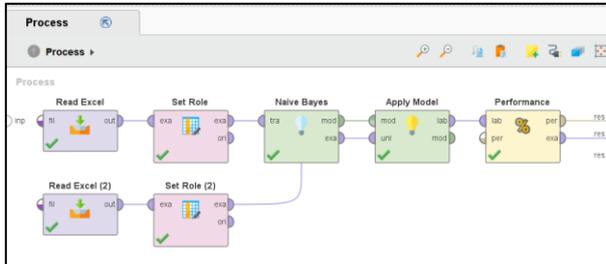
### 3.3 Tahap Modeling Naïve Bayes

Proses klasifikasi menggunakan sebuah perangkat lunak yaitu Rapid Miner. Platform ini dirancang untuk memudahkan pengguna dalam membuat dan menguji beragam model analisis data, salah satunya yaitu model *Naïve Bayes*. Berikut merupakan model *Naïve Bayes*

Hasil metode *Naïve Bayes* yang ditampilkan dari RapidMiner menunjukkan bagaimana model memprediksi kategori nanas berdasarkan fitur input seperti warna (*Red*, *Green*, *Blue*) dan tingkat kemanisan (*Brix*). Berikut penjelasan detailnya:

*Label dan Prediction (Label)*: Setiap baris dalam tabel mewakili satu data yang memiliki label asli serta prediksi dari model. Kolom "*Label*" menunjukkan

label sebenarnya dari data, sedangkan kolom "Predictions" adalah hasil prediksi yang diberikan oleh model Naive Bayes. Beberapa baris menunjukkan prediksi yang sesuai dengan label aslinya (misalnya, baris 1 dan 2), sementara baris lainnya menunjukkan prediksi yang berbeda (seperti pada baris 3 dan 4).



Gambar 5. Model Rapid Miner

Row No.	Label	prediction	confidence	confidence	confidence	Nanas	Red	Green	Blue	Brix
1	A	A	0.995	0.005	0.000	1	136	126	129	15.322
2	B	A	0.631	0.369	0.000	2	123	150	154	13.203
3	A	A	0.867	0.132	0.000	3	102	118	125	14.700
4	B	B	0.287	0.709	0.004	4	85	118	119	13.700
5	C	C	0	0.000	1.000	5	110	129	126	5.300
6	A	B	0.339	0.587	0.004	6	74	100	109	14.100
7	C	C	0	0.000	1.000	7	139	136	146	5.600
8	C	C	0.000	0.000	1.000	8	113	145	146	6.400
9	C	C	0.000	0.392	0.608	9	112	90	95	9.800
10	B	B	0.002	0.995	0.014	10	109	112	99	12

Gambar 6. Hasil dari metode Naive Bayes

**Confidence:** Terdapat tiga kolom Confidence untuk masing-masing kategori (A, B, dan C). Nilai confidence menunjukkan tingkat kepercayaan model terhadap prediksi untuk setiap kelas. Sebagai contoh, pada baris 1, model memiliki confidence 0.995 untuk kategori A (prediksi benar karena label asli juga A) dan sangat rendah untuk kategori B (0.005) dan C (0.000), sehingga model sangat yakin prediksinya adalah A.

Tabel ini juga menampilkan probabilitas atau tingkat kepercayaan model terhadap setiap prediksi untuk masing-masing kelas. Ada tiga kolom yang menunjukkan tingkat kepercayaan model terhadap kelas A, B, dan C. Sebagai contoh, pada baris 1, model memprediksi kelas A dengan tingkat kepercayaan sebesar 0.896, yang jauh lebih tinggi dibandingkan dengan kelas B dan C yang hanya memiliki tingkat kepercayaan sebesar 0.268 dan 0.000. Ini menunjukkan bahwa model sangat yakin bahwa data tersebut termasuk dalam kelas A.

Selain hasil prediksi dan tingkat kepercayaan, tabel ini juga mencantumkan fitur input yang digunakan oleh model untuk melakukan klasifikasi. Fitur tersebut mencakup nilai RGB (merah, hijau, dan biru) serta nilai Brix, yang merupakan parameter penting dalam analisis. Sebagai contoh, pada baris 1, nilai RGB yang diberikan adalah 136, 145, dan 149, serta nilai Brix sebesar 15.322. Fitur-fitur ini berfungsi sebagai masukan bagi model *Naive Bayes* untuk memprediksi kelas dari data.

**Evaluasi Hasil Prediksi:** Beberapa prediksi model benar, seperti pada baris 1, 4, 5, dan 8, di mana prediksi

sesuai dengan label sebenarnya. Namun, ada juga kesalahan, seperti di baris 2 dan 6, di mana model memprediksi kategori yang salah. Pada baris 2, misalnya, label sebenarnya adalah B, tetapi model memprediksi A dengan confidence 0.631.

Terdapat beberapa kasus di mana prediksi model tidak sesuai dengan label asli. Sebagai contoh, pada baris ke-3, label asli adalah B, namun model memprediksi C. Meski tingkat kepercayaan model terhadap kelas C lebih tinggi (0.719), prediksi ini salah. Hal ini mengindikasikan bahwa meskipun Naive Bayes memberikan prediksi dengan tingkat kepercayaan tertentu, ada kemungkinan kesalahan prediksi, terutama ketika fitur-fitur yang ada memiliki tumpang tindih antar kelas.

### 3.4 Confusion Matrix

Model klasifikasi yang telah didapatkan selanjutnya dievaluasi untuk menguji performa model, pengujian menggunakan *confusion matrix*. Hasil pengujian didapatkan tingkat akurasi mencapai 94% hal ini menunjukkan bahwa model klasifikasi kelayakan penjualan buah nanas akurat dan handal. Berikut hasil *confusion matrix* yang diuji menggunakan *platform Rapid Miner*.

## Performance Vector

Performance Vector:  
accuracy: 94.00%

Confusion Matrix:

True:	A	B	C
A:	26	3	0
B:	3	39	0
C:	0	0	29

Gambar 7. Deskripsi Performance

accuracy: 94.00%			
	true A	true B	true C
pred. A	26	3	0
pred. B	3	39	0
pred. C	0	0	29
class recall	89.66%	92.86%	100.00%

Gambar 3 Tabel Performance

### 3.5 Analisis Data

#### Confusion Matrix:

Memberikan rincian prediksi model terhadap masing-masing kelas (A, B, dan C). Tabel ini menunjukkan bagaimana model memprediksi kelas-kelas tersebut, dan membantu mengidentifikasi di mana kesalahan prediksi terjadi.



True A, True B, True C: Ini adalah kategori sebenarnya dari data (label asli). Kolom-kolom ini menunjukkan berapa banyak sampel yang benar-benar termasuk dalam kategori A, B, dan C.

Pred A, Pred B, Pred C: Ini adalah prediksi model. Baris-baris ini menunjukkan berapa banyak sampel yang diprediksi masuk ke dalam kategori A, B, atau C. Berikut contoh interpretasi:

- 26 sampel sebenarnya adalah A dan diprediksi benar sebagai A.
- 3 sampel sebenarnya A, tetapi diprediksi salah sebagai B.
- 39 sampel adalah B dan diprediksi benar sebagai B.
- 29 sampel adalah C dan diprediksi benar sebagai C.
- Tidak ada sampel kategori C yang salah diklasifikasikan.

#### Accuracy (Akurasi):

Tingkat akurasi yang didapat mencapai 94.00%, ini adalah akurasi keseluruhan model yang berarti 94% dari total prediksi model adalah benar. Akurasi dihitung dari total prediksi benar dibagi dengan jumlah total sampel.

#### Class Recall (Recall per Kelas):

*Recall* menunjukkan seberapa baik model mengenali setiap kelas:

- Untuk kelas A, recall-nya adalah 89.66%. Artinya, dari semua sampel A yang sebenarnya, 89.66% berhasil diprediksi dengan benar.
- Untuk kelas B, recall adalah 92.86%. Ini berarti 92.86% dari semua sampel B dikenali dengan benar oleh model.
- Untuk kelas C, recall adalah 100.00%, artinya model tidak pernah salah dalam memprediksi sampel dari kelas C.

#### Class Precision (Presisi per Kelas):

- Presisi menunjukkan seberapa akurat prediksi model untuk setiap kelas:
- Presisi A adalah 89.66%, artinya 89.66% dari semua sampel yang diprediksi sebagai A benar-benar merupakan A.
- Presisi B adalah 92.86%, artinya 92.86% dari semua sampel yang diprediksi sebagai B benar-benar merupakan B.
- Presisi C adalah 100.00%, artinya semua sampel yang diprediksi sebagai C memang benar-benar merupakan C.

Model ini bekerja dengan sangat baik dalam mengklasifikasikan sampel, terutama dalam mengenali kelas C, di mana tidak ada kesalahan prediksi yang terjadi. Meski demikian, ada beberapa kesalahan yang tercatat antara kelas A dan B, di mana sampel dari satu kelas terkadang diprediksi sebagai kelas lain. Dengan akurasi keseluruhan 94%, model ini dapat dianggap

cukup andal dalam melakukan tugas klasifikasinya, namun mungkin perlu peningkatan untuk mengurangi kesalahan antara kelas yang sedikit tumpang tindih, terutama antara kelas A dan B.

## IV. PENUTUP

### 4.1. Kesimpulan

Metode *Naïve Bayes* terbukti efektif dalam mengklasifikasikan nanas layak jual dengan tingkat akurasi sebesar 94%. Metode ini mampu mengolah data dengan cepat dan efisien, memudahkan petani dan pengepul dalam menentukan kualitas nanas yang layak dijual berdasarkan karakteristik fisik dan kimiawi seperti warna kulit dan tingkat kemanisan (Brix). Hasil penelitian juga menunjukkan bahwa metode ini memberikan hasil yang memuaskan pada pengukuran recall dan precision, khususnya untuk kategori tertentu.

Namun, kelemahan utama dari metode *Naïve Bayes* adalah asumsi independensi antar fitur, yang dalam kenyataannya sering kali tidak terpenuhi. Untuk itu, diperlukan praproses data yang lebih teliti serta pengembangan model lebih lanjut untuk meningkatkan akurasi, seperti kombinasi dengan metode klasifikasi lainnya. Penelitian ini berhasil membuktikan bahwa penerapan teknologi kecerdasan buatan dalam industri pertanian, khususnya pada proses klasifikasi buah nanas, dapat memberikan dampak positif pada efisiensi produksi dan distribusi.

Temuan ini juga membuka peluang untuk penerapan metode serupa pada komoditas buah lainnya, yang memerlukan teknologi klasifikasi dalam menentukan kualitas produk. Dengan demikian, penelitian ini memberikan kontribusi penting dalam pengembangan teknologi pertanian yang lebih modern dan efisien..

### 4.2. Saran

Adapun saran untuk penelitian ini yaitu diantaranya:

1. Praproses data lebih mendalam, lakukan praproses data seperti normalisasi dan pemilihan fitur yang relevan untuk meningkatkan akurasi klasifikasi.
2. Kombinasi metode klasifikasi, kombinasikan *Naïve Bayes* dengan metode lain seperti decision tree atau SVM untuk meningkatkan akurasi pada data kompleks.
3. Pengembangan sistem otomatisasi, terapkan sistem otomatisasi berbasis AI di sektor pertanian untuk mempercepat dan meningkatkan efisiensi proses klasifikasi.

## DAFTAR PUSTAKA

- [1] D. Vernanda, N. Nugraha Purnawan, T. Herdiawan Apandi, P. Negeri Subang, dan J. Brigjen Katamso No, "ANALISIS DATA

- UNTUK KLASIFIKASI TINGKAT KEMATANGAN BUAH NANAS MENGGUNAKAN SVM,” *Jurnal Ilmiah Ilmu dan Teknologi Rekayasa* |, vol. 4, 2022, doi: 10.31962/jiitr.vvii.67.
- [2] T. Sandhika Jaya, “KLASIFIKASI NANAS LAYAK JUAL DENGAN METODE NAÏVE BAYES CLASSIFIER DAN K-NEAREST NEIGHBOR,” *JSI : Jurnal Sistem Informasi (E-Journal)*, vol. 13, no. 1, 2021, [Daring]. Tersedia pada: <http://ejournal.unsri.ac.id/index.php/jsi/index>
- [3] J. Y. Song, Z. S. Qin, C. W. Xue, L. F. Bian, dan C. Yang, “Fruit grading system by reconstructed 3D hyperspectral full-surface images,” *Postharvest Biol Technol*, vol. 212, hlm. 112898, Jun 2024, doi: 10.1016/J.POSTHARVBIO.2024.112898.
- [4] J. Yousaf dkk., “Autonomous smart palm tree harvesting with deep learning-enabled date fruit type and maturity stage classification,” *Eng Appl Artif Intell*, vol. 139, hlm. 109506, Jan 2025, doi: 10.1016/J.ENGAPPAI.2024.109506.
- [5] S. K. Behera, A. K. Rath, dan P. K. Sathy, “Maturity status classification of papaya fruits based on machine learning and transfer learning approach,” *Information Processing in Agriculture*, vol. 8, no. 2, hlm. 244–250, Jun 2021, doi: 10.1016/j.inpa.2020.05.003.
- [6] R. Retnosari, P. Studi, S. Informasi, S. Nusa, dan M. Jakarta, “ANALISIS KELAYAKAN KREDIT USAHA MIKRO BERJALAN PADA PERBANKAN DENGAN METODE NAIVE BAYES,” *Jurnal Teknologi Informasi dan Rekayasa Komputer*, 2021.
- [7] T. Khatun, M. A. S. Nirob, P. Bishshash, M. Akter, dan M. S. Uddin, “A comprehensive dragon fruit image dataset for detecting the maturity and quality grading of dragon fruit,” *Data Brief*, vol. 52, hlm. 109936, Feb 2024, doi: 10.1016/J.DIB.2023.109936.
- [8] I. Loelianto dkk., “IMPLEMENTASI TEORI NAÏVE BAYES DALAM KLASIFIKASI CALON MAHASISWA BARU STMIK KHARISMA MAKASSAR,” *SINTECH (Science and Information Technology) Journal*, 2020, [Daring]. Tersedia pada: <https://doi.org/10.31598>
- [9] S. N. Salsabila, B. N. Sari, dan R. Mayasari, “KLASIFIKASI ULASAN PENGGUNA APLIKASI DISCORD MENGGUNAKAN METODE INFORMATION GAIN DAN NAÏVE BAYES CLASSIFIER,” *INFOTECH journal*, 2023, [Daring]. Tersedia pada: <https://www.ejournal.unma.ac.id/index.php/infotech/article/view/6277>
- [10] M. Anjas Aprihartha, Z. Putrawan, D. Zulhan, dan F. A. Nurfaizal, “Klasifikasi Produktivitas Buah Nanas Menggunakan Algoritma Classification and Regression Tree (CART),” *Diophantine Journal of Mathematics and Its Applications*, 2024, doi: 10.33369/diophantine.v3i1.34193.
- [11] M. Rizzo, M. Marcuzzo, A. Zangari, A. Gasparetto, dan A. Albarelli, “Fruit ripeness classification: A survey,” *Artificial Intelligence in Agriculture*, vol. 7, hlm. 44–57, Mar 2023, doi: 10.1016/J.AIIA.2023.02.004.
- [12] S. Palei, S. K. Behera, dan P. K. Sathy, “A Systematic Review of Citrus Disease Perceptions and Fruit Grading Using Machine Vision,” *Procedia Comput Sci*, vol. 218, hlm. 2504–2519, Jan 2023, doi: 10.1016/J.PROCS.2023.01.225.
- [13] D. Wang, X. Wang, Y. Chen, Y. Wu, dan X. Zhang, “Strawberry ripeness classification method in facility environment based on red color ratio of fruit rind,” *Comput Electron Agric*, vol. 214, hlm. 108313, Nov 2023, doi: 10.1016/J.COMPAG.2023.108313.
- [14] I. Ramadan, D. Syauqy, dan R. Primananda, “Sistem Pendeteksi Kematangan Buah Apel menggunakan Metode Naive Bayes berbasis Embedded System,” 2021. [Daring]. Tersedia pada: <http://j-ptiik.ub.ac.id>
- [15] D. Rahmadani, W. Handoko, F. Sriutami, dan K. U. Br.Siagian, “PENERAPAN METODE NAÏVE BAYES DALAM PENENTUAN GIZI PADA IBU HAMIL DI KELURAHAN BUNGA TANJUNG,” *JUTSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 3, no. 1, hlm. 67–74, Feb 2023, doi: 10.33330/jutsi.v3i1.2064.
- [16] W. Yulianti, W. Wiyanto, dan N. Nurhidayanti, “APPLICATION OF DATA MINING USING NAIVE BAYES ALGORITHM FOR CLASSIFICATION OF EMPLOYEE CONTRACT EXTENSION AT PT TORINI JAYA ABADI,” *Jurnal Informatika Teknologi ...*, 2023, [Daring]. Tersedia pada: <http://www.jurnal.uts.ac.id/index.php/JINTEKS/article/view/3348>
- [17] M. Azhari, Z. Situmorang, dan R. Rosnelly, “Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, hlm. 640, Apr 2021, doi: 10.30865/mib.v5i2.2937.
- [18] M. K. Suryadewiansyah, T. Endra, dan E. Tju, “Jurnal Nasional Teknologi dan Sistem Informasi Naive Bayes dan Confusion Matrix untuk Efisiensi Analisa Intrusion Detection System Alert”, doi: 10.25077/TEKNOSI.v8i2.2022.081-088.

- [19] M. Sadikin, R. Rosnelly, dan T. Surya Gunawan, "Perbandingan Tingkat Akurasi Klasifikasi Penerimaan Dosen Tetap Menggunakan Metode Naive Bayes Classifier dan C4.5," *Jurnal Media Informatika Budidarma*, vol. 4, hlm. 1100–1109, 2020, doi: 10.30865/mib.v4i4.2434.
- [20] O. Peretz, M. Koren, dan O. Koren, "Naive Bayes classifier – An ensemble procedure for recall and precision enrichment," *Eng Appl Artif Intell*, vol. 136, hlm. 108972, Okt 2024, doi: 10.1016/J.ENGAPPAL.2024.108972.