

## KLASIFIKASI NANAS LAYAK JUAL DENGAN METODE NAÏVE BAYES CLASSIFIER DAN K-NEAREST NEIGHBOR

Tri Sandhika Jaya<sup>1</sup>

<sup>1</sup> Program Studi Manajemen Informatika, Politeknik Negeri Lampung  
e-mail: sandi@polinela.ac.id

### Abstrak

Meningkatnya permintaan jajanan nanas menawarkan prospek bisnis yang menjanjikan bagi para petani buah Indonesia. Banyak petani lokal mengambil kesempatan untuk menanam buah berbentuk unik ini. Produk pertanian dan monokultur yang ditanam secara luas dapat menyebabkan hama dan penyakit. Beberapa hama dan penyakit berbahaya telah dilaporkan di beberapa negara penghasil nanas yang akan mengancam produksi pohon buah nanas. Untuk mengatasi masalah ini, digunakan metode naive bayes classifier (NBC) dan k-nearest neighbor (kNN). NBC memiliki keunggulan bahwa masing-masing variabel yang terlibat tidak saling terkait atau independen dalam penilaian. kNN memiliki keunggulan pelatihan yang sangat cepat dan ketahanan terhadap data pelatihan suara. Penelitian ini menyimpulkan bahwa metode NBC memiliki akurasi sebesar 73,3%, sedangkan metode kNN memiliki akurasi sebesar 53,3%.

**Kata kunci:** klasifikasi, nanas, naive bayes classifier, k-nearest neighbor

### Abstract

The increasing demand for pineapple snacks provides very promising business prospects for Indonesian fruit farmers. Many local farmers take this opportunity to grow this uniquely shaped fruit. Agricultural products that are widely grown and monocultures can cause pests and diseases. Several dangerous pests and diseases have been reported in some pineapple producing countries which will threaten the production of pineapple fruit trees. To solve this problem, we used naive bayes classifier (NBC) and k-nearest neighbor (kNN). NBC has the advantage that each of the variables involved are not mutually involved or independent in scoring. kNN has the advantage of very fast training and resistance to noise training data. This study concludes that the NBC method has an accuracy of 73,3% while kNN method has an accuracy of 53,3%.

**Keywords:** classification, pineapple, naive bayes classifier, k-nearest neighbor

## 1. PENDAHULUAN

Nanas (*Ananas comosus*) adalah tanaman tropis dengan buah yang dapat dimakan dan tanaman yang paling signifikan secara ekonomi dalam keluarga Bromeliaceae [1]. Nanas berasal dari Amerika Selatan, di mana ia telah dibudidayakan selama berabad-abad. Sejak tahun 1820, nanas telah ditanam secara komersial di rumah kaca dan banyak perkebunan tropis. Selanjutnya, itu adalah buah tropis terpenting ketiga dalam produksi dunia [2].

Kandungan dalam buah nanas adalah 86% air, 13% karbohidrat, 0,5% protein, dan mengandung lemak yang dapat diabaikan [3]. Dalam jumlah referensi 100 gram, nanas mentah memasok 209 kilojoule (50 kilokalori) energi makanan, dan merupakan sumber yang kaya mangan (44% Nilai Harian, DV) dan vitamin C (58% DV), tetapi selain itu tidak mengandung mikronutrien dalam jumlah yang signifikan[4].

Daging dan jus nanas digunakan dalam masakan di seluruh dunia. Di banyak negara tropis, nanas disiapkan dan dijual di pinggir jalan sebagai camilan. Potongan nanas digunakan dalam makanan penutup seperti salad buah, serta di beberapa hidangan gurih. Negara yang paling banyak mengkonsumsi jus nanas pada tahun 2017 adalah Thailand, Indonesia dan Filipina, dengan konsumsi gabungan sebesar 47% dari total dunia [1]. Di beberapa negara penghasil nanas telah dilaporkan hama dan penyakit tertentu yang dapat mengancam produksi pertanian nanas.

Nanas rentan terhadap berbagai penyakit, yang paling serius adalah penyakit layu yang ditularkan oleh kutu putih yang biasanya ditemukan di permukaan nanas, tetapi mungkin juga

di kelopak bunga yang tertutup. Penyakit lain termasuk penyakit jeruk merah muda, busuk jantung bakteri, antraknosa, jamur jantung busuk, busuk akar, busuk hitam, busuk pantat, busuk inti buah, dan virus bercak kuning [2]. Penyakit nanas merah muda (bukan penyakit merah jambu jeruk) ditandai dengan buah yang berubah warna menjadi kecoklatan hingga hitam saat dipanaskan selama proses pengalengan [3].

Untuk mempermudah proses analisis kualitas nanas yang dihasilkan dan menentukan apakah nanas layak untuk dijual, diperlukan *Naive Bayes Classifier* (NBC) dan *k-Nearest Neighbor* (kNN). Metode ini digunakan untuk perhitungan dan analisis dari kelayakan buah nanas yang dapat dijual dengan hasil yang lebih akurat dibandingkan dengan pengecekan manual.

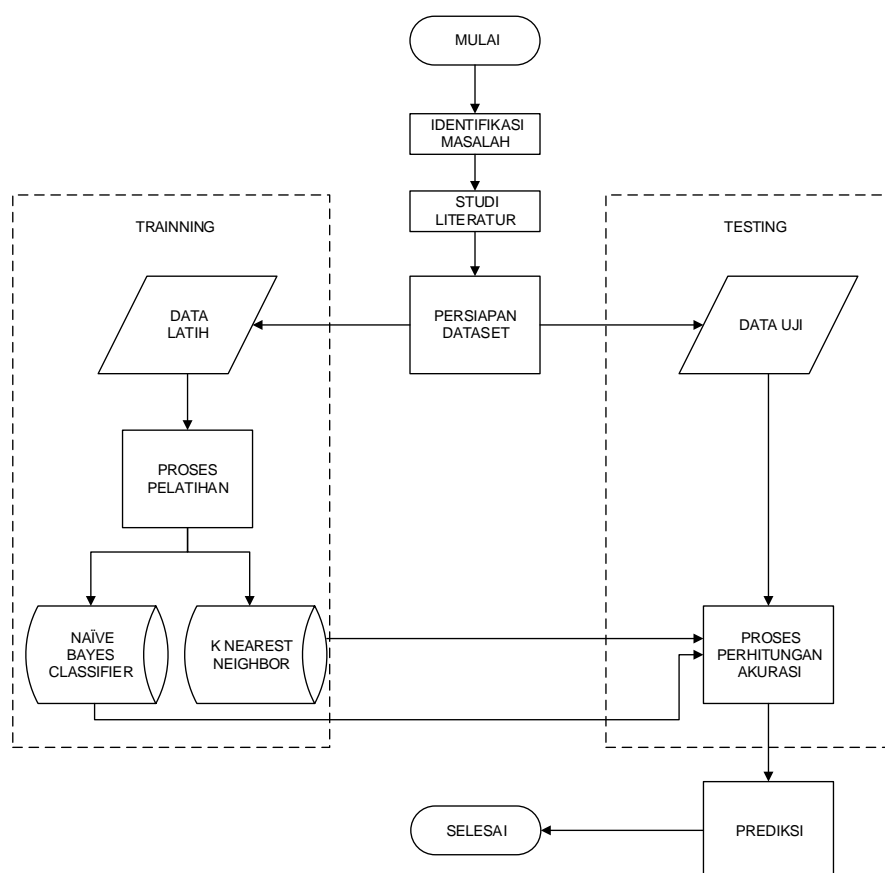
NBC adalah teknik sederhana untuk membangun model *classifier* yang memberikan label kelas ke *instance* masalah, direpresentasikan sebagai vektor nilai fitur, di mana label kelas diambil dari beberapa set [5]. Tidak ada algoritma tunggal untuk melatih pengklasifikasi seperti itu, tetapi sekelompok algoritma berdasarkan prinsip-prinsip umum. NBC mengasumsikan bahwa nilai fitur tertentu tidak bergantung pada nilai fitur lainnya, dengan mempertimbangkan variabel kelas [6]. Misalnya, buah dapat dianggap sebagai apel jika berwarna merah, bulat, dan berdiameter sekitar 10 cm. NBC menganggap masing-masing fitur ini secara independen berkontribusi pada kemungkinan bahwa buah ini adalah apel, terlepas dari kemungkinan korelasi antara fitur warna, kebulatan, dan diameter. Kelebihan dari Naive Bayes adalah hanya membutuhkan sedikit data latih untuk mengestimasi parameter yang dibutuhkan untuk klasifikasi [7].

k-NN adalah jenis klasifikasi di mana fungsinya hanya didekati secara lokal dan semua perhitungan ditunda sampai evaluasi fungsi [8]. Karena algoritma ini bergantung pada jarak untuk klasifikasi, jika fitur mewakili unit fisik yang berbeda atau datang dalam skala yang sangat berbeda maka menormalkan data pelatihan dapat meningkatkan akurasi secara dramatis [9]. Keunikan dari algoritma k-NN adalah sensitif terhadap struktur lokal data [10]. Karakteristik kumpulan data yang akan diuji dalam penelitian ini memiliki 2 kategori numerik numerik, yaitu “bisa dijual dan tidak bisa dijual”. Dengan kategori ini sangat relevan dengan model NBC dan kNN. Berdasarkan permasalahan yang sudah dijelaskan sebelumnya, penelitian ini akan menerapkan model NBC dan kNN dalam menentukan kualitas buah nanas yang layak dijual atau tidak.

Terdapat beberapa penelitian yang menggunakan metode NBC dan kNN. Firdaus, et al (2019) melakukan penelitian tentang penyelesaian masalah pengelolaan lumbung desa dengan metode kNN. Penelitian ini menghasilkan akurasi 92% dari hasil rekomendasi yang dihasilkan. Afrianti, et al (2020) melakukan penelitian untuk klasifikasi laporan durasi recovery time gangguan listrik PT. PLN (Persero) WS2JB area Palembang dengan menggunakan metode NBC. Penelitian ini menghasilkan kesimpulan bahwa dari 2487 laporan, 64,4% masuk kedalam klasifikasi recovery time dibawah 180 menit, 34,8% masuk kedalam klasifikasi recovery time dibawah satu hari, dan 0,08% masuk kedalam klasifikasi recovery time diatas satu hari. Soepriyanto (2011) melakukan penelitian tentang prediksi *stock price* dengan menggunakan NBC dan kNN. Penelitian ini menghasilkan akurasi sebesar 68,38% untuk model NBC dan 67,25% untuk model kNN.

## 2. METODE PENELITIAN

Pada penelitian ini, metode yang digunakan dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian

### 1. Identifikasi Masalah

Identifikasi masalah merupakan tahapan menginventarisir permasalahan kualitas buah nenas yang layak jual dari petani nenas tradisional di Lampung Tengah. Tahap ini menggali variabel – variabel yang dapat menentukan kualitas buah nenas yang layak jual.

### 2. Studi Literatur

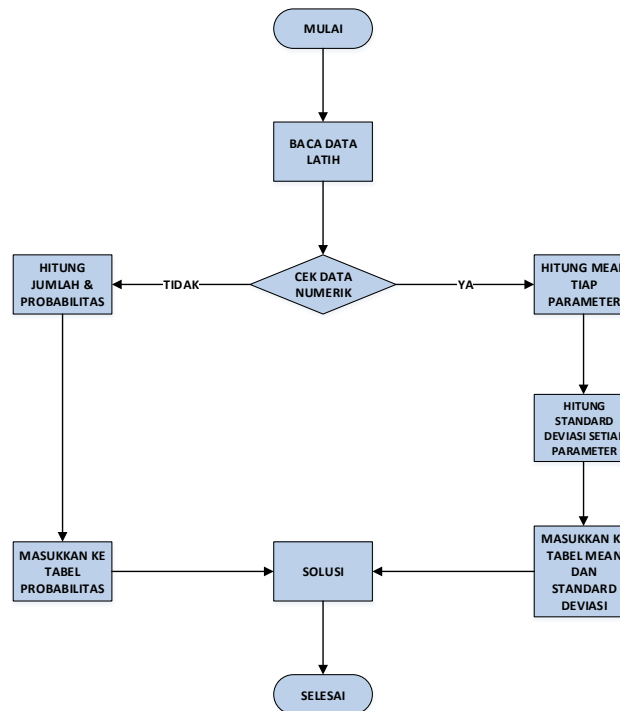
Setelah permasalahan teridentifikasi maka proses selanjutnya dilakukan proses studi literatur mengenai *naïve bayes classifier* dan *k-nearest neighbor* yang akan digunakan sebagai metode untuk membantu menentukan kualitas buah nenas layak jual berdasarkan variabel yang ditentukan. Variabel yang digunakan adalah jenis penyakit, warna buah, ukuran, kepadatan, dan kelayakan.

### 3. Persiapan Dataset

Setelah variabel dan metode sudah ditentukan, maka dilakukan proses menyiapkan dataset yang akan digunakan. Dataset yang digunakan merupakan data yang dikumpulkan dari petani nenas tradisional di Lampung Tengah. Kemudian dataset dibagi menjadi 2 jenis data, yaitu data latih dan data uji.

### 4. Pelatihan

Tahap pelatihan terdapat 3 proses besar yaitu yang pertama adalah menentukan data latih. Data latih yang digunakan adalah 200 data dari total 215 data keseluruhan. Setelah ditentukan data pelatihan. Mulai masuk proses pelatihan, dalam pelatihan ini menggunakan 2 metode yaitu metode NBC dan kNN. Alur metode NBC bisa dilihat pada Gambar 2.



Gambar 2. Alur Metode NBC

Tahapan NBC dimulai dari proses baca data untuk data latih, kemudian data tersebut dicek apakah data latih yang digunakan bertipe numerik atau bukan. Apabila bertipe numerik maka hitung nilai *mean* seperti pada persamaan 1 [11].

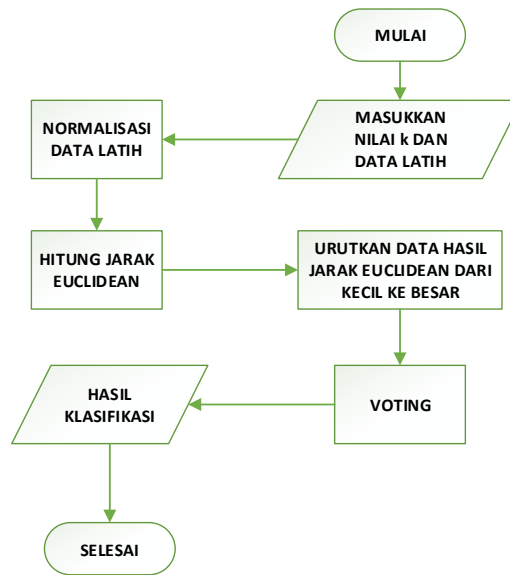
$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_i}{n} \quad (1)$$

Dimana  $\mu$  adalah *mean*,  $x_i$  adalah nilai ke- $i$ , dan  $n$  adalah jumlah data. Kemudian menghitung nilai standard deviasi seperti pada persamaan 2 [12].

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}} \quad (2)$$

Dimana  $\sigma$  adalah standard deviasi,  $x_i$  adalah nilai ke- $i$ ,  $\mu$  adalah *mean*, dan  $n$  adalah jumlah data, yang hasilnya kemudian masukkan ke dalam tabel. Apabila data bukan numerik maka hitung jumlah dan probabilitas kemudian masukkan ke tabel solusi [13].

Untuk metode kNN alurnya seperti pada Gambar 3.



Gambar 3. Alur Metode kNN

Tahapan kNN dimulai dengan memasukkan nilai k dengan nilai minimal  $k = 1$  dan nilai maksimal k adalah jumlah dataset yang digunakan. Kemudian proses dilanjutkan dengan normalisasi data latih dengan tujuan agar memiliki rentang nilai yang sama yaitu antara 0 dan 1. Metode yang digunakan untuk normalisasi adalah metode *min – max* seperti pada persamaan 3 [14].

$$\text{Normalisasi} = \frac{\text{data}_i - \text{data}_{\min}}{\text{data}_{\max} - \text{data}_{\min}} \quad (3)$$

Dimana  $\text{data}_i$  adalah data yang akan dinormalisasi berdasarkan kolom data,  $\text{data}_{\min}$  adalah data terkecil pada kolom yang sama, dan  $\text{data}_{\max}$  adalah data terbesar pada kolom yang sama. Setelah proses normalisasi kemudian hitung jarak *euclidean* dengan persamaan 4 [15].

$$d_{\text{Euclidean}}(x,y) = \sqrt{\sum_i (x_i - y_i)^2} \quad (4)$$

Dimana i adalah jumlah dataset, x adalah data latih, dan y adalah data uji. Setelah menghitung jarak *euclidean* kemudian masuk ke tahap voting. Voting merupakan tahap pemilihan jarak *euclidean* terkecil yang masuk peringkat sejumlah k. Tahap akhir adalah pengelompokan atau klasifikasi berdasarkan tahap voting yang mempunyai hasil terbanyak [14].

## 5. Pengujian

Pada tahap pengujian dilakukan penentuan data yang akan digunakan sebagai data uji. Data yang akan digunakan sebanyak 15 sampel. Data ini nanti akan digunakan untuk pembandingan pada proses prediksi. Selain itu pada tahap ini juga melakukan proses perhitungan akurasi dari model yang terbentuk pada proses pelatihan.

## 6. Prediksi

Tahap ini merupakan tahap akhir dimana, model yang terbentuk pada proses pelatihan dan sudah dihitung akurasi akan dibandingkan langsung dengan data real. Pada proses ini akan dilihat tingkat kesalahan dari model yang terbentuk pada proses pelatihan.

### 3. HASIL DAN ANALISIS

Pengujian data buah nanas layak jual ini menggunakan dataset sebanyak 238 data. Dataset diambil dari salah satu kelompok Tani di Lampung Tengah. Dataset akan dianalisis menggunakan *software* Orange Data Mining. Tabel 1 menyajikan data uji yang digunakan.

Tabel 1. Data Uji

No	penyakit	warna	ukuran	kepadatan	kelayakan	bisa dijual
1	uret	kuning	besar	padat	layak	bisa
2	uret	kuning	normal	padat	layak	bisa
3	busuk pangkal	kuning	besar	empuk	tidak	bisa
4	uret	cokelat	kecil	empuk	tidak	tidak
5	antraknosa	kuning	besar	empuk	tidak	bisa
6	antraknosa	kuning	besar	lembek	layak	bisa
7	busuk bakteri	cokelat	normal	empuk	tidak	tidak
8	busuk bakteri	cokelat	kecil	padat	tidak	bisa
9	layu	cokelat	normal	padat	tidak	tidak
10	layu	kuning	kecil	empuk	layak	tidak
11	layu	kuning	besar	lembek	layak	tidak
12	uret	cokelat	kecil	lembek	tidak	bisa
13	busuk pangkal	kuning	normal	lembek	tidak	tidak
14	busuk pangkal	cokelat	besar	padat	layak	bisa
15	antraknosa	cokelat	besar	padat	layak	bisa

#### 3.1. Hasil Uji Metode NBC

Tabel 2 menyajikan hasil uji dengan metode NBC terhadap data uji nanas.

Tabel 2 Hasil Pengujian Metode NBC

Indikator	Hasil
Akurasi	73,3%
<i>Precision</i>	73,3%
<i>Recall</i>	73,3%
AUC	0,704

Dari Tabel 2 pengujian data uji dengan metode NBC memiliki akurasi 73,3% , *precision* 73,3%, *recall* 73,3% dan AUC sebesar 0,704.

		Predicted		$\Sigma$
		bisa	tidak	
Actual	bisa	77.8 %	22.2 %	9
	tidak	33.3 %	66.7 %	6
$\Sigma$		9	6	15

Gambar 4. *Confusion Matrix* Metode NBC

Dari Gambar 4 dijelaskan bahwa *confusion matrix* data uji yang diprediksi Bisa ternyata Bisa ada 77,8% dan yang diprediksi Bisa ternyata Tidak ada 33,3%. Sedangkan data uji yang diprediksi Tidak ternyata Tidak ada 66,7% dan yang diprediksi Tidak ternyata Ya ada 22,2%.

### 3.2. Hasil uji Metode kNN

Tabel 3 menyajikan hasil uji dengan metode kNN terhadap data uji.

Indikator	Hasil
Akurasi	53,3%
<i>Precision</i>	48,3%
<i>Recall</i>	53,3%
AUC	0,611

Dari Tabel 3 pengujian dataset telur dengan metode kNN memiliki akurasi 53,3% , *precision* 48,3%, *recall* 53,3% dan AUC sebesar 0,611.

		Predicted		$\Sigma$
		bisa	tidak	
Actual	bisa	77.8 %	22.2 %	9
	tidak	83.3 %	16.7 %	6
$\Sigma$		12	3	15

Gambar 5. *Confusion Matrix* Metode kNN

Dari Gambar 5 dijelaskan bahwa *confusion matrix* data uji yang diprediksi Bisa ternyata Bisa ada 77,8% dan yang diprediksi Bisa ternyata Tidak ada 83,3%. Sedangkan data uji yang diprediksi Tidak ternyata Bisa ada 22,2% dan yang diprediksi Tidak ternyata Tidak ada 16,7%.

### 3.3. Hasil Prediksi

Hasil prediksi data yang diujikan dengan menggunakan metode NBC dan kNN disajikan pada Gambar 6.



	bisa dijual	Naive Bayes	kNN	Naive Bayes (bisa)	Naive Bayes (tidak)	kNN (bisa)	kNN (tidak)
6	bisa	bisa	bisa	0.75985	0.24015	1	0
12	bisa	bisa	bisa	0.89404	0.10596	1	0
13	bisa	bisa	bisa	0.955424	0.0445763	1	0
10	bisa	bisa	bisa	0.887769	0.112231	0.808821	0.191179
15	tidak	bisa	bisa	0.781289	0.218711	0.808603	0.191397
11	tidak	bisa	bisa	0.703532	0.296468	0.801871	0.198129
14	tidak	tidak	bisa	0.471755	0.528245	0.8	0.2
2	bisa	bisa	bisa	0.587429	0.412571	0.664249	0.335751
7	bisa	bisa	bisa	0.542653	0.457347	0.651085	0.348915
9	tidak	tidak	bisa	0.345269	0.654731	0.6	0.4
5	bisa	tidak	bisa	0.296703	0.703297	0.557875	0.442125
4	tidak	tidak	bisa	0.174194	0.825806	0.55051	0.44949
1	bisa	bisa	tidak	0.56866	0.43134	0.476627	0.523373
3	tidak	tidak	tidak	0.226586	0.773414	0.408248	0.591752
8	bisa	tidak	tidak	0.191792	0.808208	0.359065	0.640935

Gambar 6. Hasil Prediksi Data Uji

#### 4. KESIMPULAN

Berdasarkan hasil penelitian menggunakan metode NBC dan kNN terhadap data uji dengan data aktual dengan 5 feaature dari metode NBC memiliki akurasi 73,3% , *precision* 73,3%, *recall* 73,3% dan AUC sebesar 0,704. Metode kNN menghasilkan akurasi 53,3% , *precision* 48,3%, *recall* 53,3% dan AUC sebesar 0,611. Dari semua atribut yang diuji terbukti mempengaruhi kualitas buah nanas yang dihasilkan. Oleh karena itu penggunaan metode NBC atau kNN dapat membantu petani dalam menentukan kualitas nanas yang diproduksi sehingga penjualan meningkat. Berdasarkan perbandingan dari kedua metode yang mempunyai akurasi terbsesar adalah metode NBC dengan akurasi 73,3% , *precision* 73,3%, *recall* 73,3% dan AUC sebesar 0,704.

#### DAFTAR PUSTAKA

- [1] K. Azhari and Supratman, "Klasifikasi Jenis-Jenis Buah Nanas Menggunakan Learning Vector Quantization ( LVQ )," *Konvergensi Teknol. dan Sist. Inf. Klasifikasi*, pp. 357–368, 2021.
- [2] V. P. Dewa, A. Pujiyanto, and M. H. Putra, "Sistem Pakar Diagnosa Penyakit Buah Nanas Menggunakan Algoritma Bayes Berbasis Web," *Semin. Nas. Teknol. Inf. dan Multimed. 2017*, pp. 43–48, 2017.
- [3] F. T. Desy, A. Surtono, A. Supriyanto, and J. Junaidi, "Rancang Bangun Purwarupa Pemilah Nanas Berdasarkan Tingkat Kematangan Menggunakan Mikrokontroler Blue Pill STM32F103C8T6," *J. Energy, Mater. Instrum. Technol.*, vol. 1, no. 3, pp. 81–90, 2020, doi: 10.23960/jemit.v1i3.27.
- [4] B. Yanto, A. Lubis, B. H. Hayadi, and E. A. Nst, "Klarifikasi Kematangan Buah Nanas Dengan Ruang Warna Hue Saturation Intensity," pp. 135–146, 2021.
- [5] R. Vijay, B. Vangara, K. Thirupathur, and S. P. Vangara, "Opinion Mining Classification using Naive Bayes Algorithm," *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 5, pp.



- 495–498, 2020, doi: 10.35940/ijitee.e2402.039520.
- [6] N. Hayatin, G. I. Marthasari, and L. Nuraini, “Optimization of Sentiment Analysis for Indonesian Presidential Election using Naïve Bayes and Particle Swarm Optimization,” *JOIN (Jurnal Online Inform.*, vol. 5, no. 1, pp. 81–88, 2020, doi: 10.15575/join.v5i1.558.
- [7] H. Zhang, L. Jiang, and L. Yu, “Attribute and instance weighted naive Bayes,” *Pattern Recognit.*, vol. 111, 2021, doi: 10.1016/j.patcog.2020.107674.
- [8] B. Soepriyanto, “Comparative Analysis of K-NN and Naïve Bayes Methods to Predict Stock Prices,” vol. 02, no. 02, pp. 49–53, 2021.
- [9] I.- Indriati, B. Rahayudi, and C. Dewi, “Analisis Sentimen Mengenai Moda Raya Terpadu (MRT) Jakarta dengan Metode BM25 dan K-Nearest Neighbor,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 2, p. 389, 2021, doi: 10.25126/jtiik.2021824508.
- [10] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavaya, T. Kiruthika, and J. Nisha, “Crop Recommendation System for Precision Agriculture,” in *International Conference on Advanced Computing (ICoAC)*, 2016, pp. 32–36.
- [11] H. Zhang, L. Jiang, and L. Yu, “Class-specific attribute value weighting for Naive Bayes,” *Inf. Sci. (Ny.)*, vol. 508, pp. 260–274, 2020, doi: 10.1016/j.ins.2019.08.071.
- [12] Z. M. Ali, N. H. Hassoon, W. S. Ahmed, and H. N. Abed, “The Application of Data Mining for Predicting Academic Performance Using K-means Clustering and Naïve Bayes Classification,” *Int. J. Psychosoc. Rehabil.*, vol. 24, no. 03, pp. 2143–2151, 2020, doi: 10.37200/ijpr/v24i3/pr200962.
- [13] F. Xu, Z. Pan, and R. Xia, “E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework,” *Inf. Process. Manag.*, vol. 57, no. 5, p. 102221, 2020, doi: 10.1016/j.ipm.2020.102221.
- [14] M. A. Firdaus, D. R. Indah, P. E. Sevtiyuni, and C. Qonitah, “Penyelesaian Masalah Pengelolaan Lumbung Pangan Desa Menggunakan Case-Based Reasoning dengan Algoritma K-Nearest Neighbor,” *JSI J. Sist. Inf.*, vol. 11, no. 1, pp. 1684–1694, 2019, doi: 10.36706/jsi.v11i1.7699.
- [15] M. Sholihin and M. G. Rohman, “Klasifikasi Mutu Telur Berdasarkan Fitur Warna dengan Menggunakan Metode K-Nearest Neighbor,” *Semin. Nas. Sist. Inf.*, pp. 1188–1193, 2018.