

## KLASIFIKASI PENGENALAN BUAH MENGGUNAKAN ALGORITMA NAIVE BAIYES

Arif Saputra

Universitas Nahdlatul Ulama Sidoarjo  
Jl. Monginsidi Dalam, Kav DPR, Sidoklumpuk, Sidoarjo, Sidoklumpuk, Sidokumpul, Kec. Sidoarjo,  
Kabupaten Sidoarjo, Jawa Timur 61218-Indonesia

e-mail: [Ariefsaputro.si@unusida.ac.id](mailto:Ariefsaputro.si@unusida.ac.id)

|                        |                         |                          |
|------------------------|-------------------------|--------------------------|
| Received: August, 2019 | Accepted: October, 2019 | Published: October, 2019 |
|------------------------|-------------------------|--------------------------|

### Abstract

*Manually sorting varieties of apples result in high costs, subjectivity, boredom, and inconsistencies associated with humans. A means is needed to distinguish between types of apples and, therefore, some reliable techniques are necessary to identify varieties quickly and without damage. The purpose of conducting research is to investigate the application and performance for Naive Bayes algorithm for apple varieties. This software methodology involves image acquisition, preprocessing, segmentation and analysis classification varieties for apple. The prototype of Apple's classification system was built using the MATLAB R2017 development platform environment. The results in this study indicate that the estimated average accuracy, sensitivity, precision, and specificity are 81%, 73%, 100%, and 70%, respectively. MLP-Neural shows that performance of the Naive Bayes technique is consistent with Principal, Fuzzy Logic, and Neural analysis with 89%, 91%, 87%, and 82% respectively in terms of accuracy. This study shows that Naif Bayes has excellent potential for identifying nondestructive and accurate apple varieties.*

**Keywords:** Apples, Image processing, Naive Bayes, Pattern Recognition, Classification

### Abstrak

*Penyortiran varietas buah apel secara manual menghasilkan biaya tinggi, subjektivitas, kebosanan, dan ketidakkonsistenan yang terkait dengan manusia. Dibutuhkan sarana untuk membedakan varietas apel dan oleh karena itu, beberapa teknik yang andal diperlukan untuk membedakan varietas dengan cepat dan tidak merusak. Tujuan melakukan penelitian adalah untuk menyelidiki penerapan dan kinerja algoritma Naive Bayes dalam klasifikasi varietas buah apel. Metodologi perangkat lunak ini melibatkan akuisisi gambar, pra-pemrosesan dan segmentasi, analisis dan klasifikasi varietas apel. Prototipe sistem klasifikasi Apple dibangun menggunakan lingkungan platform pengembangan MATLAB R2017. Hasil dalam penelitian menunjukkan bahwa nilai rata-rata estimasi akurasi, sensitivitas, presisi dan spesifisitas masing-masing adalah 81%, 73%, 100% dan 70%. MLP-Neural menunjukkan bahwa kinerja teknik Naive Bayes konsisten dengan analisis Komponen Utama, Fuzzy Logic dan MLP-Neural dengan 89%, 91%, 87%, dan 82% masing-masing dalam hal akurasi. Studi ini menunjukkan bahwa Naif Bayes memiliki potensi yang baik untuk mengidentifikasi varietas apel nondestruktif dan akurat.*

**Kata kunci:** Buah apel, Teknik pemrosesan gambar, Naive Bayes, Pengenalan Pola, Klasifikasi.

### 1. PENDAHULUAN

Klasifikasi menggunakan fitur dimensi tinggi sering muncul di banyak kontemporer studi

statistik seperti klasifikasi tumor menggunakan *microarray* atau *throughput* tinggi lainnya pada sebuah data. Dalam penelitian ini kami

melakukan analisis kinerja dua linier yang ketat metode untuk klasifikasi dimensi tinggi, Independence Rule (atau Naif Bayes) dan Fisher diskriminan baik secara teori maupun simulasi. Kita tahu bahwa, untuk model populasi normal, ketika semua parameter diketahui Fisher optimal dan Naif Bayes suboptimal. Tapi dalam disertasi ini kami memberikan kondisi di mana Naif Bayes optimal. Melalui teori dan simulasi, kami selanjutnya, menunjukkan bahwa Naif Bayes berkinerja lebih baik daripada Fisher di bawah kondisi yang lebih luas. Kami juga mempelajari metode pemilihan fitur terkait. Dua sampel test adalah metode pemilihan fitur yang populer. Tapi itu sangat tergantung pada normalitasnya asumsi jadi kami mengusulkan algoritma pemilihan fitur umum yang berfungsi terlepas dari distribusi algorithm[1]. Dalam ilmu pertanian, gambar adalah sumber penting data dan informasi. Dubey dan Jalal, bereksperimen dan menemukan bahwa sulit untuk memproses atau mengukur data fotografi secara matematis. Analisis gambar digital dan teknologi pemrosesan gambar menghindari masalah-masalah ini berdasarkan pada kemajuan komputer dan mikroelektronika yang terkait dengan fotografi tradisional. Analisis gambar digital dan teknologi pemrosesan gambar membantu meningkatkan gambar dari kisaran visual mikroskopis ke teleskopik dan menawarkan ruang lingkup untuk analisis mereka. Beberapa aplikasi teknologi pemrosesan gambar telah dikembangkan untuk operasi pertanian. Aplikasi ini melibatkan implementasi sistem perangkat keras berbasis kamera atau pemindai warna untuk memasukkan gambar[2]. Para peneliti telah berusaha untuk memperluas pengolahan gambar dan teknologi analisis ke spektrum masalah yang luas di bidang pertanian. Pemrosesan gambar berbasis komputer sedang mengalami evolusi cepat dengan sistem komputasi yang terus berubah. Sistem pencitraan khusus yang tersedia di pasar, di mana pengguna dapat menekan beberapa tombol dan mendapatkan hasilnya, tidak

sangat fleksibel dan yang lebih penting, mereka memiliki label harga tinggi pada mereka. Penambahan pada beberapa data menggunakan pembelajaran mesin (ML) dapat digambarkan sebagai kombinasi dari berbagai metode untuk secara otomatis mendeteksi pola set data yang diberikan. Ini dapat dilakukan dengan dua cara: pembelajaran terawasi dan pembelajaran tanpa pengawasan. Seperti yang dijelaskan oleh Alpaydin, pembelajaran terawasi dapat digunakan ketika dataset yang diberikan diberi label terlebih dahulu. Selama fase pelatihan, algoritma membuat prediksi tentang setiap label titik data dengan melihat data dan mengoreksi dirinya dengan melihat label. Fase pelatihan berakhir ketika algoritma mencapai tingkat kinerja yang dapat diterima. Masalah pembelajaran yang diawasi dapat dikategorikan sebagai masalah klasifikasi dan regresi. Masalah klasifikasi adalah ketika variabel yang dihasilkan adalah kategori, seperti "oranye", "kebencian-bicara", atau "kendaraan". Masalah regresi adalah ketika variabel output adalah nilai riil, misalnya "panjang", "berat", atau "biaya". Dalam pembelajaran tanpa pengawasan label setiap titik data tidak diketahui. Tujuannya adalah untuk menemukan keteraturan dan pola dalam input data. Masalah belajar yang tidak diawasi dapat dikategorikan sebagai masalah asosiasi dan pengelompokan. Masalah asosiasi adalah di mana Anda ingin menemukan aturan yang menggambarkan sebagian besar data, seperti orang yang membeli X juga cenderung membeli Y. Masalah pengelompokan adalah di mana Anda ingin menemukan pengelompokan independen dalam data, seperti sebagai pengelompokan pelanggan dengan perilaku pembelian. Jika beberapa data pelatihan diberi label dan ada yang tidak, ada kemungkinan untuk menggunakan kombinasi pembelajaran terawasi dan tidak terawasi yang disebut pembelajaran semi-terawasi. Ini berguna karena proses pelabelan sejumlah besar data untuk pembelajaran yang diawasi seringkali sangat memakan waktu. Pra-pemrosesan data yang masuk adalah ketika data diubah dengan

cara yang membantu algoritma selama fase pelatihan. Ini mungkin termasuk penghapusan elemen-elemen berisik dalam teks, memperbaiki kata-kata yang salah eja dan memilih fitur-fitur penting yang dapat berkontribusi pada klasifikasi yang akurat. Algoritma ML dan metode pra-pemrosesan dapat digunakan untuk klasifikasi konten media sosial, analisis sentimen, klasifikasi usia, perencanaan kota, dan banyak lagi[3]. Bagian ini akan menjelaskan beberapa algoritma yang paling umum digunakan untuk pembelajaran terawasi bersama dengan metode pra-pemrosesan yang sering digunakan dengan algoritma ini, membahas tantangan dan memberikan gambaran umum tentang pekerjaan terkait yang dilakukan. Kualitas adalah salah satu faktor penting dalam pemasaran produk pertanian. Grading Apple adalah salah satu proses penting dalam penanganan dan pemasaran pasca panen. Proses penanganan (pasca panen) buah diselesaikan dalam beberapa langkah: mencuci, menyortir, menilai, mengemas, mengangkut dan menyimpan. Penyortiran dan penilaian buah dianggap sebagai langkah penanganan yang paling penting. Kualitas produk dan evaluasi kualitas adalah aspek penting dari produksi buah dan sayuran. Penyortiran dan penilaian adalah tugas pemrosesan utama yang terkait dengan produksi jenis buah pasar segar. Usaha dan waktu yang cukup banyak telah diinvestasikan di bidang otomasi. Sebagian besar mesin penilaian saat ini beroperasi berdasarkan sistem visi mesin untuk mendeteksi cacat dan cacat produk, di mana satu gambar atau lebih diambil untuk setiap objek individu dan hasil pemrosesan akan menentukan kualitas objek[4]. Perdagangan dan penyortiran produk pertanian menggunakan visi mesin bersama dengan teknik pengenalan pola, menawarkan banyak keuntungan dibandingkan perangkat penyortiran optik atau mekanis konvensional. Komputer teknik penglihatan efektif biaya, konsisten, memiliki kecepatan superior dan akurasi tinggi dalam penyortiran. Pemrosesan

gambar digital, sebagai teknik berbasis komputer, telah sangat digunakan oleh para ilmuwan untuk memecahkan masalah di bidang pertanian. Fernando et al membangun sistem untuk mendiagnosis enam jenis cacat permukaan yang berbeda pada buah jeruk menggunakan strategi analisis gambar multivariat. Gambar dibuka dan diproyeksikan ke ruang eigen referensi untuk sampai pada matriks skor yang digunakan untuk menghitung peta yang rusak dan akurasi 94,2% dilaporkan[5]. Cho et al. menggunakan pencitraan fluoresensi hiperspektral untuk mendeteksi cacat retak pada tomat ceri sementara Omid et al. menggunakan fitur bentuk, tekstur dan warna untuk mengurutkan buah tomat sesuai dengan bentuk, ukuran, kematangan, dan cacatnya. Mereka mencapai keakuratan 84,4% untuk pendeteksian cacat menggunakan pengklasifikasi jaringan saraf probabilistik (PNN). Made Sukafona et al menggunakan content based image retrieval dengan metode color moment untuk retrieve, sekaligus meningkatkan nilai recall dan menurunkan nilai presisi. Pada sebuah data[6].

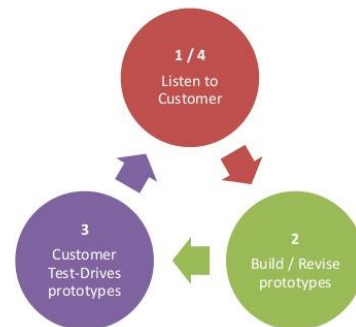
Rocha et al. menyajikan pendekatan terpadu yang dapat menggabungkan banyak fitur dan pengklasifikasi. Para penulis mendekati masalah klasifikasi multi-kelas sebagai satu set masalah klasifikasi biner sedemikian rupa sehingga orang dapat mengumpulkan bersama-sama beragam fitur dan pendekatan classifier 4 yang dirancang khusus[7] untuk bagian-bagian dari masalah. Mereka telah mencapai akurasi klasifikasi hingga 99% untuk beberapa buah, tetapi mereka menggabungkan tiga fitur, yaitu Border-interior klasifikasi (BIC), Color coherence vector (CCV), dan fitur Unser dan menggunakan dua respons teratas untuk mencapainya. Metode mereka menunjukkan hasil yang buruk untuk beberapa jenis buah dan sayuran seperti Fuji Apple[8]. Arivazhagan et al. menggabungkan fitur warna dan tekstur untuk mengklasifikasikan buah dan sayuran. Mereka menggunakan classifier jarak minimum dan mencapai akurasi 86% dibandingkan dataset

yang memiliki 15 jenis buah dan sayuran. Lebih lanjut, Faria et al. disajikan kerangka kerja untuk fusi pengklasifikasi untuk pengenalan otomatis buah dan sayuran dalam alingkungan supermarket. Mereka menggabungkan pengklasifikasi berbiaya rendah yang dilatih untuk kelas minat tertentu untuk meningkatkan tingkat pengakuan. Chowdhury et al. telah mengenali 10 sayuran berbeda menggunakan histogram warna dan fitur tekstur statistik. Mereka telah memperoleh akurasi klasifikasi hingga 96,55% menggunakan jaringan saraf sebagai penggolong[9]. Danti et al. mengklasifikasikan 10 jenis sayuran berdaun menggunakan klasifikasi BPNN dengan tingkat keberhasilan 96,40%. Mereka pertama kali memotong dan mengubah ukuran gambar dan kemudian mengekstraksi rerata dan rentang rona dan saluran saturasi gambar HSV untuk membentuk vektor fitur. Suresha et al. telah mencapai akurasi klasifikasi 95% dari dataset berisi 8 jenis sayuran yang berbeda menggunakan ukuran tekstur dalam ruang warna RGB[10]. Mereka telah menggunakan segmentasi DAS untuk mengekstraksi wilayah yang diminati sebagai pra-pemrosesan dan keputusan classifier pohon untuk tujuan pelatihan dan klasifikasi.

## 2. METODE PENELITIAN

Bagian metodologi sistem perangkat lunak atau metodologi rekayasa perangkat adalah kerangka yang akan digunakan untuk penyusunan, perencanaan, dan pengendalian proses pengembangan suatu sistem. Penelitian ini mengikuti metodologi perangkat lunak prototyping seperti yang ditunjukkan pada gambar. 2.. Dikatakan pula bahwa Prototyping sangat baik untuk mendesain antarmuka manusia-komputer yang baik. Overmyer menyatakan bahwa “Salah satu penggunaan prototyping cepat yang paling produktif hingga saat ini adalah sebagai alat untuk rekayasa kebutuhan pengguna yang berulang dan desain antarmuka manusia-komputer”. Prototyping memungkinkan pengembang dan pengusaha memiliki bukti konsep untuk dipresentasikan

sebelum penyandang dana atau dewan akademik. Prototipe adalah model fisik yang berfungsi dari suatu sistem atau subsistem, prototipe berfungsi sebagai versi awal dari sistem atau komponen dari mana persyaratan diekstraksi dan yang versi selanjutnya didasarkan. Langkah-langkah berbeda yang terlibat dalam pembuatan prototipe.



Gambar 1. Prototype model COBIT

Pengumpulan persyaratan melibatkan pertemuan pengembang dan pelanggan untuk menentukan tujuan keseluruhan untuk perangkat lunak; mengidentifikasi persyaratan yang diketahui dan menguraikan bidang-bidang di mana definisi lebih lanjut wajib. Perangkat lunak yang akan dirancang akan mengklasifikasikan buah apel sesuai dengan varietasnya. Sistem akan memiliki kamera untuk mengambil gambar buah apel, folder penyimpanan komputer untuk menyimpan gambar yang diambil, kotak pengambilan gambar untuk mengambil gambar dan menghindari cahaya yang akan merusak kualitas gambar.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Deskripsi Data

Bagian ini memuat data-data hasil penelitian yang dapat disajikan dalam bentuk deskripsi, tabel, grafik, maupun gambar. Ini adalah proses persiapan untuk mendapatkan gambar varietas apel. 150 gambar warna RGB apel Kasus penggunaan sistem klasifikasi apel menyarankan partisi skala besar dari domain masalah. Ini juga menyediakan penataan objek

analisis (yaitu aktor dan sub-sistem). Ini menjelaskan system dan tanggung jawab objek. Di bawah ini adalah penjelasan rinci dari setiap kasus penggunaan sistem: Ini adalah proses persiapan untuk mendapatkan gambar varietas apel. 150 gambar warna RGB apel varietas ditangkap menggunakan kamera ponsel dengan resolusi piksel 2048x1024 pada latar belakang putih. Gambar-gambar ini dipotong menjadi gambar yang lebih kecil dan disimpan dalam format JPG. Tugas segmentasi dan pra-pemrosesan adalah tahap awal sebelum gambar digunakan untuk proses selanjutnya. Tujuan utama dari proses ini adalah untuk mendapatkan citra biner dengan metode Otsu. Metode Otsu didasarkan pada pemilihan titik terendah antara dua kelas histogram mempertimbangkan varian antar kelas. Pra-pemrosesan pada gambar melibatkan menghilangkan noise latar belakang frekuensi rendah, menormalkan intensitas gambar partikel individu, menghapus refleksi dan menutupi bagian gambar. Ini adalah teknik

untuk meningkatkan gambar data sebelum pemrosesan komputasi. Diperlukan pra-pemrosesan untuk menghilangkan bayangan, koreksi gambar. Penghapusan bayangan sangat penting karena shadow dapat mengganggu segmentasi dan ekstraksi fitur.

### 3.2 Pembahasan

Bagian kinerja identifikasi kerangka klasifikasi selalu dievaluasi oleh empat metrik yang berbeda: sensitivitas, spesifisitas, akurasi dan presisi sensitivitas mengukur proporsi varietas apel yang benar yang diidentifikasi dengan benar dan diklasifikasikan ke jumlah total varietas apel. Spesifisitas mengukur proporsi varietas apel yang salah yang ditolak dengan benar jumlah total varietas apel. Akurasi mengukur proporsi dengan benar dan salah varietas apel yang diidentifikasi dengan benar untuk jumlah total apel. Untuk mengevaluasi kinerja analisis sistem statistik hasil eksperimen telah dilakukan.

Tabel 1. Hasil naiyes baiyes

| Apple Name       | Sensitivity (%) | Precision (%) | Specificity (%) | Accuracy (%) |
|------------------|-----------------|---------------|-----------------|--------------|
| Pink Lady        | 75              | 100           | 75              | 91           |
| Honey Crisp      | 67              | 100           | 72              | 92           |
| Golden Delicious | 90              | 100           | 0               | 90           |
| Rata -Rata       | 81              | 73            | 100             | 70           |

Perbandingan akurasi varietas apel untuk validasi dan pengujian data menunjukkan bahwa akurasi tertinggi di Naif Bayes diamati di Honey Crisp (92%), Pink lady (91%) dan yang terakhir adalah Golden delicious (90%). Untuk data validasi dan pengujian adalah 100%. Tingkat sensitivitas set data validasi adalah 75% untuk wanita pink, 67% untuk madu segar dan 90% untuk emas lezat. Untuk set data pengujian sensitivitasnya adalah 88% untuk wanita pink, 89% untuk madu segar dan 90% untuk emas lezat. Sensitivitas rata-rata untuk sistem adalah 77% untuk set data validasi dan 89% untuk set data pengujian. Tingkat

spesifisitas dari set data validasi adalah 75% untuk wanita pink, 72% untuk madu segar dan 0% untuk emas lezat. Untuk set data pengujian, spesifisitasnya adalah 80% untuk wanita pink, 80% untuk madu segar dan 0% untuk emas lezat. Golden delicious memiliki spesifisitas 0% karena selama validasi dan pengujian kami tidak menemukan nilai positif negatif dan palsu yang sebenarnya. Ini dapat dikaitkan dengan warna uniknya (hijau) yang mudah dibedakan dari dua varietas apel lainnya yang warnanya hampir serupa. Spesifisitas rata-rata untuk sistem adalah 80% untuk set data validasi dan 74% untuk set data pengujian.

#### 4. KESIMPULAN

Prototipe sistem klasifikasi buah apel menggunakan teknik pemrosesan gambar dan Naive Bayes algoritma dibangun menggunakan lingkungan platform pengembangan MATLAB R2017. Hasil yang terkait dengan tiga varietas apel: Madu renyah, emas lezat dan pink wanita menunjukkan bahwa nilai rata-rata estimasi akurasi, sensitivitas, presisi dan spesifisitas adalah 81%, 73%, 100% dan 70% masing-masing. Melalui karya penelitian sebelumnya, tinjauan literatur mengidentifikasi MLP-Neural, logika fuzzy, analisis komponen utama dan jaringan sebagai teknik lain yang telah digunakan sebelumnya. Teknik Benchmarking Naif Bayes terhadap analisis komponen utama, logika fuzzy dan teknik klasifikasi MLP-Neural menunjukkan bahwa kinerja teknik Naif Bayes konsisten dengan analisis komponen utama, logika fuzzy dan MLP-Neural dengan 91%, 90%, 89%, dan 83% masing-masing dalam hal akurasi.

#### DAFTAR PUSTAKA

- [1] P.Redman. *Good essay writing: a social sciences guide3rd ed.*, London: Open University in assoc. with Sage, 2006.
- [2] J. Bell, *Machine Learning: Hands-On for Developers and Technical Professionals*. Indianapolis : Wiley, 2015, p. 2.
- [3] A. Géron, *Hands-On Machine Learning with Scikit-Learn & TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Sebastopol : O'Reilly, 2017.
- [4] P. Willett, "The Porter stemming algorithm: then and now," *Program: Electronic Library and Information Systems*, vol. 40, no. 3, pp. 219-223,2006.
- [5] X. Lu, *Computational Methods for Corpus Annotation and Analysis*. New York: Springer, 2014.
- [6] L. Weitzel, R. A. Freire, P. Quaresma, T. Gonc,alves, and R. Prati. How does irony affect sentiment analysis tools? In *Progress in Artificial Intelligence*, pages 803–808, Cham, 2015. Springer International Publishing.
- [7] I Made Sukafona, Emmy Febriani Thalib Content Based Image Retrieval Dengan Metode Color Moment Dan K-Means, pages 73–78, 2018. Jurnal Resistor.
- [8] A. Mccallum and K. Nigam. A comparison of event models for naive bayes text classification. In *AAAI-98 Workshop on Learning for Text Categorization*, 1998.
- [9] L. Weitzel, R. A. Freire, P. Quaresma, T. Gonc,alves, and R. Prati. How does irony affect sentiment analysis tools? In *Progress in Artificial Intelligence*, pages 803–808, Cham, 2015. Springer International Publishing.
- [10] P. Yang and Y. Chen. A survey on sentiment analysis by using machine learning methods. In *2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, pages 117–121, Dec 2017.