



## **Crude Palm Oil Quality Classification Using the K-Nearest Neighbor Method**

**Andiny<sup>1</sup>, Muhammad Rizka<sup>2\*</sup>, Musta'inul Abdi<sup>3</sup>**

<sup>123</sup>Jurusan Teknologi Informasi dan Komputer, Politeknik Negeri Lhokseumawe, Kota Lhokseumawe, 24301  
INDONESIA

\*Penulis Korespondensi : rizka@pnl.ac.id

### INFORMASI ARTIKEL

*Riwayat artikel:*

Diajukan pada 15 November 2025

Direvisi pada 29 November 2025

Publikasi pada 20 Desember 2025

*Kata kunci:*

CPO

Klasifikasi

K-Nearest Neighbor

Efisiensi Produk

Machine Learning

*Keywords:*

CPO

Classification

K-Nearest Neighbor

Product Efficiency

Machine Learning

### ABSTRAK

Kualitas Crude Palm Oil (CPO) merupakan faktor penting dalam industri pengolahan kelapa sawit, khususnya di PT Perkebunan Nusantara IV Regional VI yang dituntut untuk menjaga standar mutu secara konsisten. Namun, proses klasifikasi kualitas CPO yang dilakukan secara manual berpotensi menimbulkan ketidaktepatan, yang dapat berdampak pada efisiensi produksi dan distribusi. Penelitian ini bertujuan untuk membentuk model klasifikasi kualitas CPO menggunakan algoritma K-Nearest Neighbor (KNN) berdasarkan data hasil uji laboratorium. Proses yang dilakukan meliputi tahap preprocessing data dan pelatihan model klasifikasi. Berdasarkan pengujian menggunakan confusion matrix, model menghasilkan akurasi sebesar 96,35%, dengan F1-score sebesar 97,18% untuk kelas Grade 1 dan 94,81% untuk kelas Grade 2. Nilai precision dan recall yang tinggi di kedua kelas menunjukkan bahwa metode KNN mampu mengklasifikasikan kualitas CPO secara konsisten dan akurat. Model ini diharapkan dapat menjadi dasar untuk penelitian selanjutnya dalam penerapan klasifikasi mutu CPO berbasis data.

### ABSTRACT

*The quality of Crude Palm Oil (CPO) is a critical factor in the palm oil processing industry, particularly for PT Perkebunan Nusantara IV Regional VI, which is required to maintain consistent quality standards. However, the manually conducted CPO quality classification process is potentially prone to inaccuracies, which can impact production and distribution efficiency. This research aims to develop a CPO quality classification model using the K-Nearest Neighbor (KNN) algorithm based on laboratory test data. The process involved data preprocessing and classification model training. Based on evaluation using a confusion matrix, the model achieved an accuracy of 96.35%, with an F1-score of 97.18% for Grade 1 and 94.81% for Grade 2. The high precision and recall values for both classes indicate that the KNN method is capable of classifying CPO quality consistently and accurately. This model is expected to serve as a foundation for further research in implementing data-driven CPO quality classification.*

## **1. Pendahuluan**

Kelapa sawit merupakan salah satu sumber daya alam yang sangat dikenal di seluruh dunia, dan Indonesia menjadi negara dengan luas perkebunan kelapa sawit terbesar. PT Perkebunan Nusantara IV Regional VI PKS Cot Girek merupakan salah satu perusahaan yang bergerak di sektor perkebunan kelapa sawit di Indonesia. Sebagai bagian dari BUMN, perusahaan ini memiliki peran penting dalam mendukung

perekonomian nasional melalui produksi minyak kelapa sawit berkualitas tinggi. Minyak kelapa sawit mentah atau Crude Palm Oil (CPO) merupakan komoditas strategis yang memberikan kontribusi besar terhadap perekonomian nasional. Agar dapat digunakan dalam berbagai produk pangan maupun nonpangan, CPO harus memiliki kualitas tinggi sesuai dengan standar yang telah ditetapkan.

Kualitas CPO sangat dipengaruhi oleh beberapa parameter utama, yaitu kadar asam lemak bebas (ALB), kadar air, dan kadar kotoran. Tingginya kadar ALB mencerminkan penurunan mutu akibat proses pengolahan yang tidak optimal, sedangkan kadar air yang berlebihan dapat menyebabkan kerusakan dan menurunkan nilai jual produk. Hasil pengujian laboratorium terhadap parameter-parameter tersebut dilakukan secara berkala dan digunakan untuk menentukan mutu minyak yang dihasilkan setiap periode produksi. Jika hasil pengujian memenuhi standar mutu yang ditetapkan, minyak dapat langsung dipasarkan, namun jika tidak, dilakukan proses blending dengan minyak berkualitas lebih baik agar sesuai dengan spesifikasi pasar.

Meskipun pengukuran parameter kualitas minyak telah dilakukan menggunakan alat laboratorium, proses klasifikasi mutu CPO masih bersifat manual dan bergantung pada analisis petugas, sehingga berpotensi menimbulkan ketidakkonsistenan akibat perbedaan interpretasi. Selain itu, proses manual juga membutuhkan waktu yang cukup lama, terutama dengan bertambahnya volume data pengujian. Oleh karena itu, dibutuhkan sistem berbasis data yang mampu melakukan klasifikasi mutu CPO secara akurat dan konsisten. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi kualitas minyak kelapa sawit mentah menggunakan metode K-Nearest Neighbor (KNN), yang bekerja dengan mengelompokkan data baru berdasarkan kedekatannya dengan data yang telah memiliki label mutu sebelumnya.

## 2. Metode

Pada bagian metode, penelitian ini dijelaskan secara sistematis mulai dari desain penelitian hingga evaluasi model. Setiap tahapan dirancang agar dapat direplikasi pada penelitian berikutnya.

### 2.1 Data dan Preprocessing Data

Data penelitian ini berasal dari data sekunder yang diperoleh dari PT Perkebunan Nusantara IV Regional VI, berupa catatan hasil pengujian mutu minyak kelapa sawit mentah (Crude Palm Oil/CPO) di laboratorium perusahaan. Dataset berisi parameter utama yang dijadikan acuan penilaian mutu, yaitu:

- Asam Lemak Bebas (ALB/FFA), yang mencerminkan tingkat degradasi minyak,
- Kadar Air, yang berpengaruh pada stabilitas penyimpanan, dan
- Kadar Kotoran, yang menunjukkan tingkat kemurnian minyak.

Ketiga parameter tersebut dipilih karena merupakan standar industri yang umum digunakan untuk menentukan kualitas dan grade minyak sawit. Standar mutu CPO yang digunakan dalam penelitian ini mengacu pada spesifikasi perusahaan, yaitu  $ALB \leq 3,50\%$ , kadar air  $\leq 0,20\%$ , dan kadar kotoran  $\leq 0,020\%$  untuk Grade 1, sedangkan data yang tidak memenuhi kriteria tersebut dikategorikan sebagai Grade 2.

Proses pengumpulan data dilakukan secara sistematis melalui pencatatan uji laboratorium yang berlangsung berkala selama proses produksi. Pengambilan sampel dilakukan setiap 2 jam dalam periode 24 jam, kemudian diuji di laboratorium untuk memperoleh nilai rata-rata harian dari parameter ALB, air, dan

kotoran. Data yang terkumpul kemudian diverifikasi dan dibersihkan dari nilai kosong atau duplikasi agar konsisten dan dapat diandalkan.

Tahap selanjutnya adalah preprocessing, yaitu mengubah data mentah hasil uji laboratorium ke dalam format terstruktur. Proses ini mencakup pembersihan data, normalisasi, serta perhitungan rata-rata harian sebelum digunakan untuk analisis. Hasil preprocessing disimpan dalam format digital terstruktur (Excel/JSON) agar siap digunakan pada tahap pelabelan dan klasifikasi. Gambar 1 merupakan tahapan rancangan preprocessing pada data.



**Gambar 1.** Rancangan *Preprocessing*

Proses labelling dilakukan secara otomatis dengan aturan berbasis ambang batas standar mutu. Jika minimal dua dari tiga parameter memenuhi standar (ALB, air, kotoran), maka data diberi label Grade 1. Jika hanya satu atau tidak ada parameter yang memenuhi, maka data dikategorikan sebagai Grade 2. Data berlabel ini menjadi ground truth yang digunakan pada tahap pelatihan dan pengujian metode K-Nearest Neighbor (KNN). Gambar 2 adalah proses alur kerja labelling pada data.



**Gambar 2.** Rancangan *Labeling*

## 2.2 Metode KNN

Tahap Metode klasifikasi yang digunakan dalam penelitian ini adalah K-Nearest Neighbor (KNN), yang bekerja dengan menentukan kelas suatu data uji berdasarkan kedekatannya dengan data latih. Kedekatan tersebut dihitung menggunakan rumus Euclidean Distance sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

dengan  $x$  adalah data uji,  $y$  data latih, dan  $n$  jumlah fitur.



**Gambar 3.** Flowchart metode K-NN

### 2.2.1 Arsitektur Umum

Proses klasifikasi KNN dalam penelitian ini diawali dengan input dataset hasil uji laboratorium yang mencakup kadar ALB, kadar air, dan kadar kotoran. Selanjutnya, data tersebut melalui tahap preprocessing yang meliputi pembersihan data, normalisasi, dan pelabelan sesuai standar mutu. Dataset kemudian dibagi menjadi data latih dan data uji. Data latih digunakan untuk membangun model KNN dengan menentukan parameter  $K$  yang optimal, sementara data uji diklasifikasikan berdasarkan mayoritas tetangga terdekat. Evaluasi performa model dilakukan menggunakan confusion matrix yang menghasilkan metrik akurasi, precision, recall, dan F1-score untuk mengukur efektivitas klasifikasi. Nilai  $K$  ditentukan melalui cross-validation, dengan mencoba beberapa kandidat (misalnya  $K = 3, 5, 7$ ). Pemilihan nilai  $K$  dilakukan untuk memperoleh hasil akurasi terbaik. Nilai  $K$  yang terlalu kecil berpotensi sensitif terhadap outlier, sedangkan nilai  $K$  yang terlalu besar dapat mengurangi kejelasan pemisahan antar kelas.

### 2.2.2 Rancangan Algoritma

Alur algoritma KNN yang diterapkan dimulai dengan inialisasi nilai  $K$ . Selanjutnya, dihitung jarak antara data uji dengan seluruh data latih menggunakan metode Euclidean Distance untuk mengukur kedekatan relatif antar data. Hasil perhitungan jarak ini kemudian diurutkan berdasarkan nilai terkecil hingga terbesar. Berdasarkan urutan tersebut, dipilih  $K$  tetangga terdekat dari data uji. Kelas dari data uji kemudian ditentukan melalui mekanisme mayoritas kelas dari  $K$  tetangga terdekat yang telah terpilih. Terakhir, hasil klasifikasi disimpan dan ditampilkan dalam bentuk kategori Grade 1 atau Grade 2. Dengan rancangan tersebut, sistem mampu mengklasifikasikan mutu CPO secara otomatis. Grade 1 diberikan pada

data dengan kualitas baik sesuai standar ( $ALB \leq 3,50\%$ ,  $Air \leq 0,20\%$ ,  $Kotoran \leq 0,020\%$ ), sedangkan Grade 2 untuk data di luar kriteria tersebut. Berikut Gambar 3 merupakan flowchart klasifikasi K-Nearest Neighbor.

## 2.3 Pengujian

Pengujian sistem klasifikasi kualitas minyak kelapa sawit dilakukan dengan tiga pendekatan, yaitu confusion matrix, black box testing, dan white box testing. Tujuan pengujian adalah memastikan bahwa sistem bekerja sesuai dengan rancangan, baik dari sisi logika internal maupun dari sisi fungsionalitas pengguna.

### 2.3.1 Confusion Matrix

Pengujian performa algoritma KNN dilakukan menggunakan confusion matrix, dengan membagi dataset menjadi 70% data latih dan 30% data uji. Evaluasi dilakukan untuk mengukur seberapa akurat model dalam mengklasifikasikan mutu CPO, melalui metrik accuracy, precision, recall, dan F1-score. Confusion matrix juga membantu mengidentifikasi kesalahan klasifikasi seperti false positive dan false negative.

### 2.3.2 Black Box Testing

Black box testing dilakukan untuk menguji fungsionalitas sistem dari sisi pengguna tanpa melihat struktur internal kode. Skenario uji mencakup halaman login, input data uji laboratorium, proses klasifikasi, pengelolaandata (tambah, edit, hapus), pembuatan laporan, serta logout. Misalnya, pada skenario login, jika admin memasukkan username dan password valid maka sistem mengarahkan ke halaman utama, sedangkan jika tidak valid sistem menampilkan pesan error. Dengan pendekatan ini, keandalan fungsi sistem dapat dipastikan sesuai spesifikasi.

### 2.3.3 White Box Testing

White box testing digunakan untuk menguji alur logika internal program, khususnya pada implementasi algoritma KNN. Fokus utama pengujian dalam penelitian ini mencakup beberapa aspek kritis. Pertama, dilakukan perhitungan jarak antar data menggunakan Euclidean Distance untuk mengukur tingkat kemiripan antara sampel uji dan data latih. Kedua, dipilih nilai  $K$  yang optimal beserta tetangga terdekat berdasarkan urutan jarak terpendek. Ketiga, diterapkan mekanisme voting untuk menentukan kelas mayoritas dari tetangga terdekat yang terpilih. Keempat, hasil voting tersebut menjadi dasar penentuan klasifikasi akhir kategori kualitas CPO. Terakhir, validasi ketat dilakukan terhadap proses pembagian data training–testing dan parameterisasi nilai  $K$  untuk memastikan keandalan dan akurasi model secara keseluruhan. Pengujian dilakukan melalui unit testing untuk memastikan setiap fungsi berjalan sesuai logika yang telah dirancang.

## 3. Hasil Dan Pembahasan

### 3.1 Implementasi sistem (UI)

Antarmuka sistem dirancang sederhana dan interaktif agar mudah digunakan oleh petugas laboratorium. Modul login digunakan untuk validasi kredensial sebelum pengguna dapat mengakses sistem. Dashboard menampilkan ringkasan data harian dan distribusi label kualitas (Grade 1 dan Grade 2), sehingga admin dapat memantau proporsi mutu CPO secara real time. Selain itu, tersedia modul input

data untuk mengunggah file hasil uji laboratorium dalam format Excel, serta menu pengelolaan data (edit, hapus, dan pencarian berdasarkan tanggal).

### 3.1.1 Input Data

Halaman input data digunakan untuk mengunggah file Excel berisi parameter mutu (ALB, kadar air, kadar kotoran). Sistem menampilkan data dalam tabel dengan fitur pencarian, edit, dan hapus, sehingga admin dapat mengelola data dengan mudah. Untuk tampilan Pada halaman input data dapat dilihat pada Gambar 4.

Tanggal	ALB	Air	Kotoran	Aksi
31-12-2024	3.392	0.192	0.025	Edit Hapus
30-12-2024	4.317	0.269	0.026	Edit Hapus
29-12-2024	3.498	0.205	0.028	Edit Hapus
28-12-2024	3.436	0.202	0.030	Edit Hapus

**Gambar 4.** Halaman Input Data

### 3.1.2 Modul Labelling

Antarmuka labelling menampilkan data rata-rata harian beserta label kualitas (Grade 1/Grade 2). Label diberikan otomatis berdasarkan standar mutu, dan hasilnya digunakan untuk pelatihan model KNN. Gambar 10 merupakan halaman modul labelling.

ID	Tanggal	Avg ALB	Avg Air	Avg Kotoran	Label	Actions
8939	2024-12-31	3.392	0.192	0.025	Grade 1	Edit Delete
8938	2024-12-30	4.317	0.269	0.026	Grade 2	Edit Delete
8937	2024-12-29	3.498	0.205	0.028	Grade 1	Edit Delete
8936	2024-12-28	3.436	0.202	0.030	Grade 1	Edit Delete
8935	2024-12-27	3.576	0.263	0.021	Grade 2	Edit Delete
8934	2024-12-26	3.947	0.254	0.025	Grade 2	Edit Delete
8933	2024-12-25	4.011	0.234	0.029	Grade 2	Edit Delete
8932	2024-12-24	3.666	0.219	0.029	Grade 2	Edit Delete

**Gambar 5.** Halaman Modul Labelling

### 3.2. Evaluasi Confusion Matrix

Evaluasi performa model K-Nearest Neighbor (KNN) dilakukan menggunakan confusion matrix dengan pembagian data 70% sebagai data latih dan 30% sebagai data uji. Confusion matrix memberikan gambaran jumlah data yang berhasil diprediksi dengan benar maupun yang salah, serta menjadi dasar untuk menghitung metrik evaluasi lainnya, yaitu accuracy, precision, recall, dan F1-score. Berikut hasil dari Confusion Matrix dapat dilihat pada Gambar 6.

Akurasi Model: 96.35%				
Confusion Matrix				
	Pred Grade 1	Pred Grade 2		
Actual Grade 1	138	7		
Actual Grade 2	1	73		

  

Classification Report				
Kelas	Precision	Recall	F1 Score	Support
Grade 1	99.28%	95.17%	97.18%	145
Grade 2	91.25%	98.65%	94.81%	74
<b>Weighted Avg</b>	<b>96.57%</b>	<b>96.35%</b>	<b>96.38%</b>	<b>219</b>

**Gambar 6.** Hasil Confusion Matrik

Berdasarkan hasil pengujian, model KNN menghasilkan akurasi uji sebesar 96,35%, Nilai precision dan recall yang tinggi pada kedua kelas menunjukkan bahwa model KNN mampu mengenali mutu minyak sawit dengan sangat baik, baik pada Grade 1 maupun Grade 2. Nilai rata-rata tertimbang (weighted average) yang konsisten di atas 96% memperlihatkan bahwa model tidak bias terhadap salah satu kelas, serta mampu mempertahankan keseimbangan performa. Hasil ini menegaskan bahwa metode KNN efektif dalam melakukan klasifikasi kualitas minyak kelapa sawit mentah (CPO) dengan tingkat keandalan yang tinggi. Model dapat dijadikan dasar dalam pengambilan keputusan mutu CPO secara cepat, objektif, dan berbasis data.

## 4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penggunaan metode K-Nearest Neighbor (KNN) dalam klasifikasi kualitas minyak kelapa sawit mentah dapat berjalan dengan baik. Sistem yang dikembangkan mencakup fitur input data (baik secara manual maupun melalui impor Excel), pelatihan model, proses klasifikasi, dan tampilan hasil klasifikasi yang menyeluruh. Data yang dianalisis mencakup parameter kualitas seperti kadar Asam Lemak Bebas (ALB), kadar air, dan kadar kotoran. Model KNN dilatih menggunakan data yang telah diberikan label sesuai dengan standar perusahaan, selanjutnya digunakan untuk mengklasifikasikan data baru secara otomatis ke dalam kategori Grade 1 dan Grade 2. Metode K-Nearest Neighbor (KNN) terbukti berhasil dalam klasifikasi kualitas minyak kelapa sawit mentah. Model ini menghasilkan akurasi mencapai 96,35%, dengan rata-rata akurasi cross-validation sebesar 96,50%. Penilaian menggunakan confusion matrix menunjukkan tingkat ketepatan

klasifikasi yang tinggi, dengan nilai F1-score, precision, dan recall untuk kedua kelas berada di atas 95%. Rata-rata tertimbang dari ketiga metrik ini berada di kisaran 96,38%. yang menandakan performa model yang stabil, konsisten, dan layak digunakan dalam mendukung proses klasifikasi mutu CPO yang lebih akurat dan terstandar.

## Referensi

- [1] Suryani, D., Yulianti, A., Maghfiroh, E. L., & Alber, J. (2022). Klasifikasi kualitas produk kelapa sawit menggunakan metode Naïve Bayes. *Sistemasi: Jurnal Sistem Informasi*, 11(1), 251–259. <https://sistemasi.ftik.unisi.ac.id>
- [2] Rosdiana, Utomo, W., Daud, M., & Muhammad. (2023). Analisa deteksi tingkat kualitas minyak pada buah sawit berdasarkan tingkat kematangan warna buah menggunakan drone berbasis pengolahan citra di PTPN Membang Muda Sumut. *Sang Pencerah: Jurnal Ilmiah Universitas Muhammadiyah Buton*, 9(1), 184. <http>
- [3] Sidik, A. (2024). Data mining menggunakan metode Naive Bayes untuk menetapkan standar untuk produk minyak sawit mentah. *EJECTS: E-Journal Computer, Technology and Informations System*, 3(2), 52.
- [4] Bayu, D. P., Priyambada, & Supriyanto, G. (2023). Analisis rendemen minyak kelapa sawit (CPO) berdasarkan tingkat kematangan buah di PT. Bumitama Gunajaya Agro (Karya Bakti Agro Sejahtera). *Jurnal Teknik Pertanian*, 1(3).
- [5] Efendi, R., Faurina, R., & Hamimmah, T. S. (2023). Implementasi metode Naïve Bayes pada penentuan mutu CPO (Crude Palm Oil). *JSAI: Journal Scientific and Applied Informatics*, 6(3), 381–390. <https://doi.org/10.36085>
- [6] Sukmawati, S., Rahmi, S., & Nurhidayatullah. (2023). Analisis penentuan kadar air dan kadar kotoran terhadap kualitas minyak Crude Palm Oil (CPO) di Daily Tank di PT. Socfindo Kebun Seunagan. *Jurnal Teknologi Pengolahan Pertanian*, 5(1), 27–32.
- [7] Indriati, & Ridok, A. (2016). Sentiment analysis for review mobile applications using neighbor method weighted k-nearest neighbor (NWKNN). *Journal of Environmental Engineering & Sustainable Technology (JEEST)*, 3(1), 23–32. <http://jeest.ub.ac.id>
- [8] Ghifari, H. S., & Utamingrum, F. (2022). Klasifikasi kualitas minyak goreng berdasarkan fitur warna dan kejernihan dengan metode k-nearest neighbour berbasis Arduino Uno. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(7), 3269–3274. <http://j-ptiik.ub.ac.id>
- [9] Nurhasanah, D., Lestari, D. A., & Simatupang, S. (2023). Pemilihan kualitas produk kelapa sawit menggunakan metode Naïve Bayes di Labuhanbatu Selatan. *Jurnal Teknisi: Jurnal Teknologi Komputer dan Sistem Informasi*, 3(1), 24–31. <http://jurnal.goretanpena.com/index.php/teknisi>
- [10] Rosalina, E., & Agustin, S. (2019). Klasifikasi umur lahan perkebunan kelapa sawit pada citra foto udara berdasarkan tekstur menggunakan metode Naïve Bayes. *INDEXIA*, 1(1), 6. <http://journal.umg.ac.id/index.php/indexia/>
- [11] Mahmud, S. F. (2019). Proses pengolahan CPO (Crude Palm Oil) menjadi RBDPO (Refined Bleached and Deodorized Palm Oil) di PT XYZ Dumai. *UNITEK*, 12(1), Januari–Juni. p-ISSN 2089-3957, e-ISSN 2580-2585.
- [12] Levia, D., & Mhubaligh. (2023). Analisis proses produksi CPO untuk mengidentifikasi faktor-faktor yang mempengaruhi kualitas mutu CPO. *Jurnal Teknologi dan Manajemen Industri Terapan (JTMIT)*, 2(2), 82–89. [13] Medikano, A., & Pardila, L. I. (2022). Analisis pengendalian kualitas produk akhir minyak kelapa sawit di PT. Bengkulu Sawit Lestari. *Jurnal Teknologi Industri*, 2(4). e-ISSN: 2774-8308.
- [14] Loka, S. K. P., & Marsal, A. (2023). Comparison algorithm of K-nearest neighbor and Naïve Bayes classifier for classifying nutritional status in toddlers. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(1), 8-14.
- [15] Uminingsih, Ichsanudin, M. N., Yusuf, M., & Suraya. (2022). Pengujian fungsional perangkat lunak sistem informasi perpustakaan dengan metode black box testing bagi pemula. *STORAGE – Jurnal Ilmiah Teknik dan Ilmu Komputer*, 1(2), 1–8. <https://doi.org/10.55123>.