

K-NEAREST NEIGHBOR UNTUK KLASIFIKASI MUTU PRODUKSI FRESH FRUIT BUNCHES (FFB) DI PT. PADASA ENAM UTAMA KEBUN TELUK DALAM

**Widia Fahwana Br Panjaitan¹, Muhammad Ardiansyah Sembiring^{2*},
Elly Rahayu²**

¹Mahasiswa Prodi Sistem Informasi, STMIK Royal

²Dosen Prodi Sistem Informasi, STMIK Royal

email: adinmantap88@gmail.com

Abstract: PT. Padasa Enam Utama, a palm oil plantation company, currently assesses the quality of Fresh Fruit Bunches (FFB) based solely on physical aspects. They use Microsoft Excel without a specialized application system, which can lead to subjective assessments and the risk of fraud. This research aims to develop a predictive system for the quality of Fresh Fruit Bunches. Data collection was conducted using quantitative methods through direct observation and interviews with relevant parties. The study shows that the use of the K-Nearest Neighbor method provides the best accuracy with a more efficient calculation process. With this system, the company's performance in making decisions regarding FFB quality is expected to improve. The system helps reduce human errors in quality assessments and offers visualizations that make it easier for users to understand the classification of production quality. Although the results are reliable, there is still room for further development, such as improving accuracy through more advanced data preprocessing techniques or using more complex machine learning models.

Keywords: Data Mining; K-Nearest Neighbor algorithm; Fresh Fruit Bunches (FFB); Python; Streamlit.

Abstrak: PT. Padasa Enam Utama, sebuah perusahaan perkebunan kelapa sawit, saat ini menilai kualitas produksi Tandan Buah Segar (TBS) berdasarkan aspek fisik saja. Mereka menggunakan *Microsoft Excel* tanpa sistem aplikasi khusus, yang dapat menyebabkan penilaian tidak objektif dan risiko kecurangan. Penelitian ini bertujuan untuk mengembangkan sistem prediksi mutu kualitas Tandan Buah Segar. Dalam pengumpulan data, digunakan metode kuantitatif melalui observasi langsung dan wawancara dengan pihak terkait. Penelitian menunjukkan bahwa penggunaan metode *K-Nearest Neighbor* memberikan akurasi terbaik dengan proses perhitungan yang lebih efisien. Dengan adanya sistem ini, diharapkan kinerja perusahaan dalam membuat keputusan terkait mutu produksi TBS dapat meningkat. Sistem ini membantu mengurangi kesalahan manusia dalam penilaian mutu TBS dan memberikan visualisasi yang memudahkan pengguna memahami klasifikasi mutu produksi. Meskipun telah memberikan hasil yang dapat diandalkan, masih ada ruang untuk pengembangan lebih lanjut, seperti peningkatan akurasi melalui teknik preprocessing data yang lebih canggih atau penggunaan model-machine learning yang lebih kompleks.

Kata kunci: Data Mining; Algoritma K-Nearest Neighbor; Tandan Buah Segar (TBS); Python; Streamlit.

PENDAHULUAN

Kelapa sawit, sebagai salah satu komoditas hasil perkebunan, memegang peranan krusial dalam aktivitas perekonomian Indonesia. Fakta menunjukkan bahwa Indonesia menduduki peringkat teratas sebagai produsen *Crude Palm Oil* (CPO) terbesar di seluruh dunia[1]. Sejalan dengan meningkatnya permintaan global terhadap minyak sawit, perlu diperhatikan upaya untuk peningkatan kualitas dan jumlah produksi kelapa sawit dengan tepat, guna mencapai sasaran yang diinginkan sebagai tahap awal dalam rangkaian produksi minyak kelapa sawit[2]. Perusahaan perlu memastikan prediksi dari kualitas produksi Tandan Buah Segar (TBS). Hal tersebut dilakukan guna jika kualitas dari produksi Tandan Buah Segar (TBS) menurun, perusahaan dapat dengan cepat mengambil tindakan agar tidak terjadi kerugian seperti mengimplementasikan sistem prediksi, menjalankan pemantauan rutin terhadap parameter-parameter kunci kualitas Tandan Buah Segar (TBS), melibatkan petani dalam program penyuluhan dan pelatihan terkait penerapan sistem prediksi, serta membentuk strategi perencanaan pemeliharaan tanaman yang responsif berdasarkan hasil prediksi, sehingga menghindari penurunan kualitas produksi Tandan Buah Segar (TBS) yang berimbas pada harga pasar.

Adapun permasalahan yang terjadi di PT. Padasa Enam Utama (PEU) yakni dalam menentukan kualitas produksi Tandan Buah Segar (TBS) hanya dari segi fisiknya saja. Sedangkan menentukan kualitas produksi Tandan Buah Segar (TBS) bisa dinilai berdasarkan ukuran, warna serta tingkat kematangan buah. Hal tersebut bisa menyebabkan *human error* karena penilaian yang kurang objektif serta kecurangan-kecurangan yang mungkin bisa dilakukan oleh beberapa oknum.

Oleh karena itu, penerapan klasifikasi produksi TBS menjadi suatu kebutuhan penting untuk mendapatkan pemahaman yang lebih mendalam terhadap hasil produksi selanjutnya. Klasifikasi kualitas produksi Tandan Buah Segar (TBS) tersebut akan menggunakan metode *K-Nearest Neighbor*. Algoritma *K-Nearest Neighbor* juga dikenal sebagai metode klasifikasi dengan rumus yang cukup sederhana, sehingga memudahkan penerapannya dalam implementasi praktis[3].

Penelitian sebelumnya dengan judul Rancang Bangun Alat Pendeteksi Kematangan Buah Sawit Dengan Menggunakan Metode Image Processing Berdasarkan Komposisi Warna. Klasifikasi kematangan buah sawit dari pengujian 30 sampel citra buah sawit, dapat dilihat dari rentang nilai Hue. Ekstrasi RGB ke HSV nilai pada kulit buah Sawit menghasilkan dua klasifikasi nilai rentang Hue, yaitu warna hitam kekuningan dengan nilai Hue (0.25604 - 0.59155) untuk sawit mentah, warna orange merah tua dengan nilai Hue (0.06511 - 0.12985) untuk sawit matang [4].

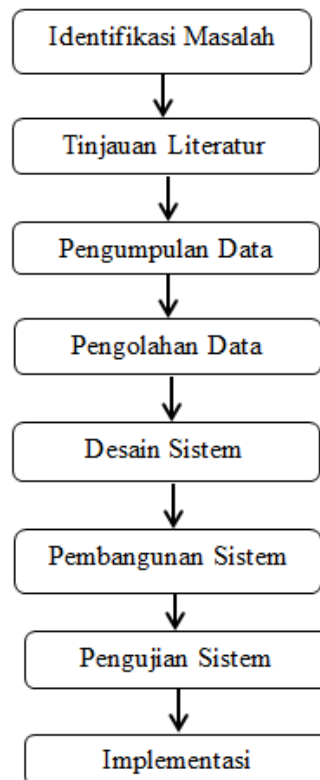
Penelitian sebelumnya Akurasi Citra Image Penyakit Daun Kentang berdasarkan Citra Sehat, Citra Early Blight, dan Citra Late Blight Menggunakan Convolutional Neural Network (CNN). Penelitian ini memanfaatkan algoritma deep learning, yaitu Convolutional Neural Network (CNN), untuk klasifikasi citra daun kentang. Metode CNN menggunakan proses konvolusi di mana citra dipecah menjadi gambar-gambar yang lebih kecil dengan konvolusi yang sama. Hasil dari gambar-gambar kecil tersebut kemudian dimasukkan ke dalam array baru yang digunakan untuk prediksi. Data yang digunakan dalam penelitian ini berjumlah 5400 citra, terbagi menjadi tiga kelas: citra sehat, citra early blight, dan citra late blight. Hasil pengujian menunjukkan akurasi

tertinggi pada data validasi sebesar 99% dengan waktu komputasi per epoch sekitar 2,5 detik [5].

Penggunaan pendekatan *K-Nearest Neighbor* telah diterapkan sebelumnya dalam penelitian, seperti yang dilakukan oleh [6]. Dalam penelitian tersebut, tujuh metode dibandingkan, dan metode *K-Nearest Neighbor* mencapai akurasi sebesar 95,00%, sehingga mempercepat proses pengambilan keputusan. Penelitian lainnya, yakni [7], mengawali tahap klasifikasi dengan menghitung klasifikasi pada *immunotherapy dataset* menggunakan metode *K-Nearest Neighbor*. Pengujian dilakukan dengan variasi nilai *k* dari 1 hingga 10, dan hasilnya dievaluasi untuk mendapatkan data yang benar dan yang salah dari proses klasifikasi. Selanjutnya, dilakukan perhitungan nilai *Accuracy*, *Precision*, dan *Recall* berdasarkan *Confusion Matrix*. Hasil akhir penelitian ini menunjukkan bahwa metode *K-Nearest Neighbor* mencapai akurasi sebesar 90,00%.

METODE

Dalam penelitian ini menggunakan beberapa tahap. Tahapan ini akan dijelaskan dalam gambar 1.



Gambar 1. Alur Penelitian

Gambar 1 menjelaskan aliran penelitian, mulai dari identifikasi masalah, tinjauan literatur, pengumpulan data, pengolahan data, desain sistem, pembangunan sistem, pengujian sistem dan implementasi.

Pengumpulan Data

Dataset terdiri dari 161 entri data sortasi panen periode januari 2024.

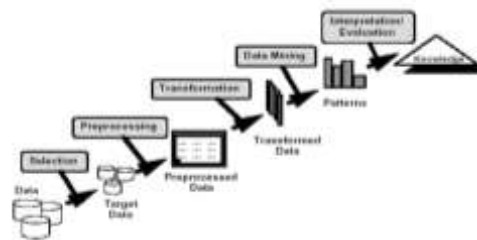
$$\text{dis} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (1)$$

Untuk penjelasan rumus diatas adalah sebagai berikut:

- dis = jarak euclidean
- x = Nilai data *training* ke-
- y = Nilai data *testing* ke-

Data Mining

Data Mining disebut juga *Knowledge Discovery from Database* (KDD) adalah serangkaian proses yang digunakan untuk mengeksplorasi nilai tambah dari suatu set data, menghasilkan pengetahuan yang sebelumnya tidak terdeteksi secara manual [8].



Gambar 2. Tahapan Proses KDD

Proses data mining melibatkan langkah-langkah seperti membersihkan data dari *noise* dan informasi tidak relevan, menggabungkan data ke dalam basis data baru, memilih data yang relevan untuk analisis, mengubah format data agar cocok untuk proses, menerapkan metode pada proses mining untuk menemukan informasi berharga, dan mengevaluasi pola atau prediksi untuk menentukan tingkat keberhasilannya.

Klasifikasi

Klasifikasi adalah suatu teknik analisis data yang dimaksudkan untuk konstruksi model atau algoritma yang mampu mengategorikan data ke dalam kelas tertentu berdasarkan pola atau aturan yang ditemukan dari data yang telah diberi label sebelumnya [9].

K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* berfungsi untuk mengenali kesamaan antara data baru dengan data yang telah ada, lalu menempatkan data baru tersebut dalam kategori yang paling serupa dengan kategori yang sudah ada sebelumnya [10].

HASIL DAN PEMBAHASAN

Pengumpulan Data

Data dikumpulkan secara langsung di PT. Padasa Enam Utama Kebun Teluk Dalam, Kecamatan Teluk Dalam, Kabupaten Asahan. Berikut adalah tabel beberapa dari Data Sortasi Panen periode Januari 2024, khususnya data fraksi atau tingkat kematangan Tandan Buah Segar.

Tabel 1. Data *Training*

TANGGAL	AFDELING	F00	F0	F1	F2	F3	F4	F5	KELAS
03-Jan-24	AFD I	18	85	206	309	231	77	0	TB
03-Jan-24	AFD II	10	68	188	282	211	70	0	B
03-Jan-24	AFD III	11	65	187	280	210	70	0	B
03-Jan-24	AFD IV	8	53	146	219	165	55	0	B
03-Jan-24	AFD V	16	85	216	324	243	81	0	TB
03-Jan-24	AFD VI	20	96	222	332	249	83	0	TB
03-Jan-24	KEBUN	83	452	1164	1745	1309	436	0	TB

Tabel 2. Data *Testing*

TANGGAL	AFDELING	X1	X2	X3	X4	X5	X6	X7	KELAS
01-Feb-24	AFD I	20	100	175	200	305	0	3	?
01-Feb-24	AFD II	5	150	200	107	230	74	0	?
01-Feb-24	AFD III	0	65	100	260	220	40	1	?
01-Feb-24	AFD IV	8	55	185	219	145	35	1	?
01-Feb-24	AFD V	10	65	310	312	245	65	0	?
01-Feb-24	AFD VI	15	90	135	250	255	80	0	?
01-Feb-24	KEBUN	58	525	1105	1348	1400	294	5	?

Menentukan Nilai K

Menentukan nilai K tidak memiliki rumus yang pasti. Namun, ada cara yang dapat dipertimbangkan, yaitu jika jumlah kelas adalah genap, disarankan untuk menggunakan nilai K yang ganjil. Sebaliknya, jika jumlah kelas adalah ganjil, disarankan untuk menggunakan nilai K yang genap. Dalam penelitian ini menggunakan nilai K=3.

Menghitung Jarak Euclidean

Menghitung jarak antara data yang akan diuji dengan seluruh data latih.

D1=

$$\sqrt{(18 - 20)^2 + (85 - 100)^2 + (206 - 175)^2 + (309 - 200)^2 + (231 - 305)^2 + (77 - 0)^2 + (0 - 3)^2}$$

= 156,4768353

D2=

$$\sqrt{\begin{array}{l} (10 - 20)^2 + (68 - 100)^2 + (188 - 175)^2 + (282 - 200)^2 + \\ (211 - 305)^2 + (70 - 0)^2 + (0 - 3)^2 \end{array}}$$

$$= 147,5194902$$

D3=

$$\sqrt{\begin{array}{l} (11 - 20)^2 + (65 - 100)^2 + (187 - 175)^2 + (280 - 200)^2 + \\ (210 - 305)^2 + (70 - 0)^2 + (0 - 3)^2 \end{array}}$$

$$= 147,5940378$$

D4=

$$\sqrt{\begin{array}{l} (8 - 20)^2 + (53 - 100)^2 + (146 - 175)^2 + (219 - 200)^2 + \\ (165 - 305)^2 + (55 - 0)^2 + (0 - 3)^2 \end{array}}$$

$$= 161,8301579$$

D5=

$$\sqrt{\begin{array}{l} (16 - 20)^2 + (85 - 100)^2 + (216 - 175)^2 + (324 - 200)^2 + \\ (243 - 305)^2 + (8177 - 0)^2 + (0 - 3)^2 \end{array}}$$

$$= 166,4692164$$

D6=

$$\sqrt{\begin{array}{l} (20 - 20)^2 + (96 - 100)^2 + (222 - 175)^2 + (332 - 200)^2 + \\ (249 - 305)^2 + (83 - 0)^2 + (0 - 3)^2 \end{array}}$$

$$= 172,2875503$$

D7=

$$\sqrt{\begin{array}{l} (83 - 20)^2 + (452 - 100)^2 + (1164 - 175)^2 + (1745 - 200)^2 + \\ (1309 - 305)^2 + (436 - 0)^2 + (0 - 3)^2 \end{array}}$$

$$= 2165,903968$$

Proses dilakukan secara berulang dengan metode yang sama, yaitu membandingkan setiap data uji dengan seluruh data latih yang ada, sehingga dapat menghasilkan total jarak seperti yang tertera dalam tabel berikut:

Tabel 3. Hasil Perhitungan Euclidean Data Satu Terhadap Data *Training*

TANGGAL	AFDELING	F00	F0	F1	F2	F3	F4	F5	EUCLIDEAN
03-Jan-24	AFD I	18	85	206	309	231	77	0	156,4641
03-Jan-24	AFD II	10	68	188	282	211	70	0	147,1802
03-Jan-24	AFD III	11	65	187	280	210	70	0	147,3194
03-Jan-24	AFD IV	8	53	146	219	165	55	0	161,3846
03-Jan-24	AFD V	16	85	216	324	243	81	0	166,4212
03-Jan-24	AFD VI	20	96	222	332	249	83	0	172,2876
03-Jan-24	KEBUN	83	452	1164	1745	1309	436	0	2164,988

Mencari Jarak Terkecil Ke Terbesar

Mengurutkan euclidean yang diperoleh dari jarak terkecil ke terbesar.

Tabel 4. Hasil Perhitungan Euclidean Terurut

TANGGAL	AFDELING	F00	F0	F1	F2	F3	F4	F5	EUCLIDEAN
03-Jan-24	AFD I	18	85	206	309	231	77	0	3
03-Jan-24	AFD II	10	68	188	282	211	70	0	1
03-Jan-24	AFD III	11	65	187	280	210	70	0	2
03-Jan-24	AFD IV	8	53	146	219	165	55	0	4
03-Jan-24	AFD V	16	85	216	324	243	81	0	5
03-Jan-24	AFD VI	20	96	222	332	249	83	0	6
03-Jan-24	KEBUN	83	452	1164	1745	1309	436	0	7

Identifikasi Jarak K Terdekat

Menentukan jarak terdekat dari K=3, kemudian masukkan label sesuai dengan data latih yang ada.

Tabel 5. Penentuan Jarak Terdekat dan Kelas

TANGGAL	AFDELING	F00	F0	F1	F2	F3	F4	F5	EUCLIDEAN	URUTAN JARAK	KE-LAS
03-Jan-24	AFD I	18	85	206	309	231	77	0	156,4641	3	TB
03-Jan-24	AFD II	10	68	188	282	211	70	0	147,1802	1	B
03-Jan-24	AFD III	11	65	187	280	210	70	0	147,3194	2	B
03-Jan-24	AFD IV	8	53	146	219	165	55	0	161,3846	4	B
03-Jan-24	AFD V	16	85	216	324	243	81	0	166,4212	5	TB
03-Jan-24	AFD VI	20	96	222	332	249	83	0	172,2876	6	TB
03-Jan-24	KEBUN	83	452	1164	1745	1309	436	0	2164,988	7	TB

Menentukan Jumlah Kelas Dari Jarak Eclidean

Berdasarkan tabel yang diberikan, Data Sortasi Panen AFD 1 dengan nilai F00=20, F0=100, F1=175, F2=200, F3=305, F4=0, dan F5=3 dapat diklasifikasikan se-

bagai mutu produksi Tandan Buah Segar yang “Bagus”. Hal ini disimpulkan berdasarkan nilai K yang diambil yaitu $K=3$, kriteria yang dimiliki mendekati 2 jarak Euclidean Bagus dan 1 jarak Euclidean yang Tidak Bagus.

Tabel 6. Hasil Kedekatan Mutu Produksi Tandan Buah Segar

TANGGAL	AFDELING	X1	X2	X3	X4	X5	X6	X7	KELAS
03-Jan-24	AFD I	18	85	206	309	231	77	0	TB
03-Jan-24	AFD II	10	68	188	282	211	70	0	B
03-Jan-24	AFD III	11	65	187	280	210	70	0	B
03-Jan-24	AFD IV	8	53	146	219	165	55	0	B
03-Jan-24	AFD V	16	85	216	324	243	81	0	TB
03-Jan-24	AFD VI	20	96	222	332	249	83	0	TB
03-Jan-24	KEBUN	83	452	1164	1745	1309	436	0	TB
01-Feb-24	AFD 1	20	100	175	200	305	0	3	B

Proses ini akan terus dilakukan untuk menghitung Euclidean data uji pada setiap data *testing* hingga semua data yang akan diolah selesai.

SIMPULAN

Simpulan dari penelitian ini adalah bahwa implementasi model K-Nearest Neighbor (K-NN) dalam menilai mutu produksi Tandan Buah Segar (TBS) memberikan kontribusi signifikan dalam pengambilan keputusan yang efisien dan efektif di PT. Padasa Enam Utama Kebun Teluk Dalam. Sistem ini membantu mengurangi kesalahan manusia dalam penilaian mutu TBS dan memberikan visualisasi yang memudahkan pengguna memahami klasifikasi mutu produksi. Meskipun telah memberikan hasil yang dapat diandalkan, masih ada ruang untuk pengembangan lebih lanjut, seperti peningkatan akurasi melalui teknik preprocessing data yang lebih canggih atau penggunaan model-machine learning yang lebih kompleks. Keseluruhan, penelitian ini memberikan wawasan berharga dan dapat menjadi pijakan untuk penelitian lebih lanjut dalam bidang evaluasi mutu produksi TBS.

DAFTAR PUSTAKA

- [1] Y. A. Iskandar, I. Sukarno, and I. N. Sabani, “Pengembangan Model Optimasi Suplai Tandan Buah Segar dengan Goal Programming di PKS Tanjung Seumantoh PTPN I,” *Matrik J. Manaj. dan Tek. Ind. Produksi*, vol. 23, no. 2, p. 165, 2023, doi: 10.30587/matrik.v23i2.3959.
- [2] M. H. Nasution, “Pengaruh Pemberian Sludge Limbah Kelapa Sawit dan POC Kulit Pisang Pada Pertumbuhan Bibit Kelapa Sawit (*Elaeis guineensis* Jacq) di Pre-Nursery,” *J. Ilm. Mhs. Pertan.*, vol. 3, pp. 1–12, 2023.
- [3] A. Christopher and T. M. S. Mulyana, “Klasifikasi Tumbuhan Angiospermae

- Menggunakan Algoritma K-Nearest Neighbor Berdasarkan Pada Bentuk Daun,” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 7, no. 4, pp. 1233–1243, 2022, doi: 10.29100/jipi.v7i4.3211.
- [4] R. Bangun *et al.*, “Design of a Palm Fruit Ripeness Detection Tool Using Image Processing Methods Based on Color Composition,” *J. Garuda Pengabd. Kpd. Masy.*, vol. 1, no. 2, 2023, [Online]. Available: <http://creativecommons.org/licenses/by-sa/4.0/>
- [5] H. Ulfa, H. Maulana, L. Fimawahib, and F. Erwis, “Akurasi Citra Image Penyakit Daun Kentang berdasarkan Citra Sehat , Citra Early Blight , dan Citra Late Blight Menggunakan Convolutional Neural Network (CNN),” vol. 10, no. 2, pp. 167–174, 2024.
- [6] N. R. Muntiari and K. H. Hanif, “Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma Machine Learning,” *J. Ilmu Komput. dan Teknol.*, vol. 3, no. 1, pp. 1–6, 2022, doi: 10.35960/ikomti.v3i1.766.
- [7] U. Erdiansyah, A. Irmansyah Lubis, and K. Erwansyah, “Komparasi Metode K-Nearest Neighbor dan Random Forest Dalam Prediksi Akurasi Klasifikasi Pengobatan Penyakit Kutil,” *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 208, 2022, doi: 10.30865/mib.v6i1.3373.
- [8] K. Annisa, B. S. Ginting, and M. A. Syar, “Penerapan Data Mining Pengelompokan Data Pengguna Air Bersih Berdasarkan Keluhannya Menggunakan Metode Clustering Pada Pdam Langkat,” *J. Sist. Inf. Kaputama*, vol. 6, no. 2, pp. 165–179, 2022, doi: 10.59697/jsik.v6i2.167.
- [9] I. P. Ninditama, W. Cholil, M. Akbar, and D. Antoni, “Klasifikasi keluarga sejahtera study kasus : Kecamatan Kota Palembang,” *J. TEKNO KOMPAK*, vol. 15, no. 2, pp. 37–49, 2020.
- [10] G. F. Ramadhan and E. Winarno, “Sistem Diagnosa Penyakit Ikan Menggunakan Metode Case Based Reasoning Dengan Algoritma Similaritas Sorgenfrei Dan K-Nearest Neighbor,” *J. Ilm. Inform.*, vol. 10, no. 01, pp. 44–50, 2022, doi: 10.33884/jif.v10i01.4634.