

## Klasifikasi Kualitas Kopi Robusta dengan Algoritma *K-Nearest Neighbor* di PT. Indokom Citra Persada Situbondo

Nur Aizah <sup>1</sup>, Zaehol Fatah <sup>2</sup>

<sup>1,2</sup> Universitas Ibrahimy Situbondo, Indonesia

<b>Info Artikel</b>	<b>ABSTRAK</b>
<p><u>Riwayat Artikel</u>  Diterima: 02-11-2024  Disetujui: 15-12-2024</p> <p><u>Kata Kunci</u>  Klasifikasi Kopi;  Data Mining;  KNN;  Kualitas Kopi;  PT. Indokom Citra Persada.</p> <p><a href="mailto:*aizaheshal39@gmail.com">*aizaheshal39@gmail.com</a></p>	<p>Algoritma <i>K-Nearest Neighbors</i> KNN merupakan metode Data Mining klasifikasi yang diterapkan untuk memperoleh kualitas kopi robusta di PT. Indokom Citra Persada Situbondo. Kopi robusta dengan karakteristik tumbuh di berbagai kondisi lingkungan, memiliki variabilitas kualitas yang signifikan berdasarkan faktor fisik dan kimia seperti pH tanah, curah hujan, suhu, kadar air, dan ketinggian lahan. Pendekatan kuantitatif diterapkan pada data yang dikumpulkan melalui wawancara dan studi data historis. Hasil pengujian nyaris sempurna dengan model KNN yang dikembangkan sehingga mampu mencapai tingkat akurasi 98,81% dalam mengklasifikasikan kualitas kopi. Evaluasi menggunakan <i>confusion matrix</i> mengindikasikan bahwa model berhasil mengidentifikasi kualitas kopi dengan baik, dengan <i>recall</i> mencapai <i>confidence</i> 97,67% untuk kelas 'Baik' dan 100% untuk kelas 'Cukup'. Temuan ini menunjukkan efektivitas KNN sebagai alat untuk klasifikasi kualitas produk pertanian, memberikan solusi bagi PT. Indokom Citra Persada Situbondo untuk meningkatkan konsistensi dan kualitas produk kopi robusta.</p>

### 1. PENDAHULUAN

Kopi robusta, secara ilmiah dikenal sebagai *Coffea canephora var. robusta*, merupakan salah satu varietas asli dari spesies *Coffea canephora*. Varietas ini terkenal karena kemampuannya tumbuh di berbagai kondisi lingkungan dan ketahanannya terhadap penyakit[1], sehingga menjadi pilihan utama bagi banyak petani kopi di seluruh dunia, termasuk di Kabupaten Situbondo. PT. Indokom Citra Persada Situbondo merupakan salah satu perusahaan pengolahan kopi terkemuka di Kabupaten Situbondo yang berfokus pada produksi berbagai jenis kopi dari banyak daerah di Situbondo. Setiap lokasi penanaman kopi di Kabupaten Situbondo memiliki karakteristik yang unik, seperti pH tanah, kadar air, dan suhu yang berbeda-beda. Seperti Sumbermalang memiliki curah hujan tinggi, suhu stabil, dan tanah subur di ketinggian sekitar 1200-1300 mdpl, ideal untuk robusta berkarakter kuat. Kayumas dengan lokasi lebih tinggi (1300-1400 mdpl) dengan suhu dingin malam hari dan pH tanah lebih basa, menghasilkan robusta bercita rasa lebih kompleks.

PT. Indokom Citra Persada Situbondo sebagai salah satu perusahaan penghasil kopi robusta berfokus pada kualitas dan konsistensi produk untuk memenuhi permintaan pasar yang terus meningkat. Kondisi lingkungan yang beragam di setiap lokasi kebun kopi di Situbondo dapat menyebabkan variasi kualitas yang signifikan, mencakup faktor-faktor fisik dan kimia di sekitar tanaman yang memengaruhi pertumbuhan, hasil panen, bahkan mutu produk akhir[2]. Variabilitas ini menjadi tantangan bagi PT. Indokom Persada Situbondo dalam memenuhi standar mutu yang diharapkan. Tanpa sistem klasifikasi yang efektif, pemantauan kualitas ini

akan menjadi proses yang kurang efisien, bahkan dapat mengakibatkan penurunan kualitas produk dalam jangka panjang.

Klasifikasi merupakan teknik dalam data *mining* sebagai metode analisis data yang digunakan untuk mengelompokkan data ke dalam kategori atau kelas tertentu berdasarkan fitur-fitur atau karakteristik yang relevan[3]. Dalam klasifikasi, sebuah model dibangun dengan menggunakan data pelatihan yang berisi data yang sudah diklasifikasikan. Model ini kemudian digunakan untuk menentukan kelas dari data baru. Dalam konteks prediksi kualitas kopi robusta, klasifikasi menggunakan metode *K-Nearest Neighbors* (KNN) sangat berguna karena KNN dapat mengklasifikasikan data berdasarkan kesamaan pola antara kondisi lingkungan dan hasil kualitas kopi sebelumnya[4]. Metode *K-Nearest Neighbors* (KNN) dapat menjadi solusi dengan memanfaatkan data lingkungan sebagai dasar klasifikasi kualitas kopi secara otomatis.

Sejumlah penelitian dalam lima tahun terakhir, diantaranya penelitian oleh Aldi Tangkelayuk pada tahun 2022 dengan judul "The Klasifikasi Kualitas Air Menggunakan Metode KNN, *Naïve Bayes*, dan *Decision Tree*" menunjukkan bahwa metode *K-Nearest Neighbors* (KNN) lebih unggul dalam klasifikasi kualitas air dibandingkan dengan *Naïve Bayes* dan *Decision Tree*. Dengan tingkat akurasi tertinggi mencapai 71,19%, KNN memberikan presisi yang baik dalam mengidentifikasi air aman dan tidak aman untuk diminum[5]. Penelitian selanjutnya dengan judul "Perbandingan Metode *Naïve Bayes* dan KNN dalam Penerapan Data *Mining* untuk Klasifikasi Menu Potensial (Studi Kasus : Ateku Kopi Medan )" oleh Siboro dan Niken Alfredo Donatus pada tahun 2024, menunjukkan bahwa penggunaan KNN dalam klasifikasi ini membantu dalam menentukan kategori potensi menu (tinggi, sedang, atau rendah), sehingga dapat meningkatkan efisiensi pemasaran dan pengelolaan stok makanan. Hasil evaluasi menunjukkan bahwa akurasi KNN mencapai 88,23%, yang memperkuat relevansi metode ini dalam mendukung pengambilan keputusan di sektor kuliner yang kompetitif[6]. Juga penelitian oleh Muhammad Jundanuddin, Zaehol Fatah, dan Akhlis Munazilin pada tahun 2024 yang berjudul "Implementasi Data Mining Pada Penilaian Kinerja Guru Menggunakan Metode *K-Nearest Neighbor* (K-NN) Studi Kasus MI Alfaqiri Silo-Jember" yang menyatakan bahwa implementasi *K-Nearest Neighbor* (KNN) untuk evaluasi penilaian kinerja guru dipilih karena kemampuannya mengklasifikasi data tanpa asumsi distribusi, sehingga mampu menghasilkan hasil evaluasi lebih akurat dan rinci tentang kapasitas pengajaran buku[7].

Menjadi hal menarik dan penting untuk mendapatkan hasil Klasifikasi Kualitas Kopi Robusta dengan Algoritma *K-Nearest Neighbor* yang dilakukan pada data dari PT. Indokom Citra Persada Situbondo.

## 2. METODE

### 1) Metode Penelitian

Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi kualitas kopi robusta menggunakan metode *K-Nearest Neighbors* (KNN) dengan mempertimbangkan faktor-faktor lingkungan. Penelitian ini merupakan penelitian kuantitatif, yakni pendekatan yang mengutamakan pengumpulan dan analisis data berupa angka atau statistik untuk mengidentifikasi pola, mengukur variabel, serta membuat generalisasi tentang populasi melalui perhitungan matematis dan penggunaan alat statistik[8] yang dilakukan di PT. Indokom Citra Persada Situbondo.

### 2) Instrumen dan Teknik Pengumpulan Data

Pertama, dilakukan wawancara dengan petani dan manajer kebun kopi untuk mendapatkan informasi yang komprehensif mengenai kondisi pertanian, praktik budidaya, serta berbagai faktor lingkungan yang berpengaruh signifikan terhadap kualitas kopi. Selain itu, studi data

akan dilakukan dengan mengumpulkan informasi dari catatan historis mengenai pH tanah, curah hujan, suhu rata-rata, kadar air, dan ketinggian lahan di setiap lokasi kebun kopi yang menjadi subjek penelitian. Data yang dikumpulkan melalui wawancara dan studi data ini akan digunakan sebagai dasar dalam melakukan analisis lebih lanjut menggunakan metode *K-Nearest Neighbors* (KNN) untuk mengklasifikasikan kualitas kopi berdasarkan kondisi lingkungan yang telah diidentifikasi.

### 3) Metode Data Mining

Data *mining* adalah proses menganalisis data besar untuk menemukan pola, hubungan, dan pengetahuan baru yang bermanfaat untuk pengambilan keputusan[9]. Melalui teknik-teknik analisis tertentu, data *mining* dapat membantu organisasi memahami data mereka dengan lebih baik dan menemukan informasi tersembunyi yang bermanfaat. Metode data *mining* mencakup beberapa teknik utama, seperti klasifikasi, klusterisasi, regresi, asosiasi, dan deteksi anomali[10]. Metode ini sangat berguna untuk berbagai kebutuhan, seperti prediksi tren pasar, analisis perilaku konsumen, dan, dalam konteks ini, klasifikasi kualitas produk pertanian.

Akurasi adalah ukuran dalam klasifikasi yang menunjukkan seberapa banyak prediksi yang benar dalam keseluruhan prediksi yang dilakukan oleh model. Secara matematis, akurasi dihitung dengan membagi jumlah prediksi yang benar dengan total prediksi yang dilakukan[11]. Akurasi memberikan gambaran umum tentang performa model, tetapi pada kasus data yang tidak seimbang, akurasi dapat menjadi metrik yang kurang informatif.

*K-Nearest Neighbors* adalah algoritma yang sederhana namun efektif untuk klasifikasi dan regresi dalam data *mining*. KNN bekerja dengan prinsip "kesamaan"[12], untuk mengklasifikasikan titik data baru, algoritma ini mencari beberapa titik data terdekat (*neighbors*) di dalam dataset yang sudah ada. "K" adalah jumlah tetangga terdekat yang akan digunakan untuk memprediksi kategori dari titik data baru tersebut. Misalnya, jika  $K=3$ , maka KNN akan mencari tiga tetangga terdekat dan menentukan kelas yang paling umum di antara ketiga tetangga tersebut. KNN berguna dalam berbagai aplikasi, termasuk klasifikasi citra, deteksi pola, dan prediksi kualitas produk seperti kopi. Berikut tahapan pemecahan masalah menggunakan algoritma KNN :

1. Menentukan Parameter K  
Tentukan nilai K (jumlah tetangga terdekat) yang akan digunakan dalam algoritma. Nilai K harus disesuaikan dengan dataset karena jika nilai K terlalu kecil atau terlalu besar dapat mengurangi akurasi model. Untuk mencari K yang optimal, bisa dilakukan uji coba (*cross-validation*) untuk melihat nilai K yang menghasilkan akurasi terbaik.
2. Mengumpulkan Data Latih  
Dataset berisi data latih yang sudah memiliki label atau kelas masing-masing. Setiap titik data berisi atribut-atribut yang relevan, misalnya data lingkungan seperti suhu, pH tanah, curah hujan, kelembapan, dan ketinggian untuk klasifikasi kualitas kopi.
3. Menghitung Jarak *Euclidean Distance*  
Untuk setiap titik data baru yang ingin diklasifikasikan, hitung jaraknya dengan setiap titik dalam data latih. Umumnya, jarak yang digunakan adalah jarak Euclidean yang dihitung dengan rumus :

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

11

**Gambar 1.** Rumus Perhitungan *Euclidean Distance*

Dalam rumus ini,  $x$  adalah data baru,  $y$  adalah data latih, serta  $x_i$  dan  $y_i$  yang merupakan nilai atribut pada dimensi  $i$  yang relevan.

4. Mengidentifikasi K Tetangga Terdekat  
 Setelah menghitung jarak, pilih K titik data dengan jarak terpendek sebagai tetangga terdekat dari data baru. Data ini adalah sampel yang paling mirip dengan data baru berdasarkan atribut yang ada.
5. Mengklasifikasikan Data Baru Berdasarkan Mayoritas Kelas Tetangga Terdekat  
 Dari K tetangga terdekat yang telah diidentifikasi, perhatikan kelas (label) yang dimiliki oleh masing-masing tetangga tersebut. Data baru kemudian akan diberi label berdasarkan kelas yang paling sering muncul di antara tetangga-tetangga terdekat tersebut. Sebagai contoh, jika sebagian besar tetangga terdekat berada dalam kelas "Kualitas Tinggi," maka data baru tersebut juga akan diklasifikasikan ke dalam kelas "Kualitas Tinggi."
6. Menentukan Hasil Akhir Klasifikasi  
 Kelas yang terpilih dari langkah sebelumnya menjadi prediksi kelas untuk data baru tersebut. Dengan demikian, data baru telah memiliki kelas berdasarkan kedekatannya dengan data latih yang telah memiliki label.

*Confusion matrix* adalah tabel yang menggambarkan kinerja model klasifikasi dengan cara menunjukkan jumlah prediksi benar dan salah yang dilakukan model berdasarkan kelas yang sebenarnya[13]. Tabel ini terdiri dari empat elemen utama: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). *Confusion matrix* memberikan gambaran rinci tentang jenis kesalahan yang terjadi pada model, sehingga memudahkan analisis kelemahan model. *Confusion matrix* adalah tabel yang menunjukkan hasil prediksi model terhadap data aktual, dengan menguraikan jumlah prediksi untuk setiap kategori dalam klasifikasi biner atau multi-klas. Dalam kasus biner, *confusion matrix* memiliki empat elemen.

**Tabel 1.** *Confusion Matrix*

	Prediksi Positif	Prediksi Negatif
Aktual Positif	True Positive (TP)	False Negative (FN)
Aktual Negatif	False Positive (FP)	True Negative (TN)

Penjelasan:

- True Positive (TP) : Prediksi positif dan benar.
- False Positive (FP) : Prediksi positif tetapi salah (juga disebut Type I Error).
- False Negative (FN) : Prediksi negatif tetapi salah (juga disebut Type II Error).
- True Negative (TN) : Prediksi negatif dan benar.

*Recall*, atau dikenal juga sebagai sensitivitas, adalah rasio antara prediksi positif yang benar (TP) dengan total data yang seharusnya diprediksi positif (TP + FN). *Recall* menunjukkan seberapa baik model dapat menangkap semua data positif yang ada.

$$\text{Recall} = \frac{TP}{TP + FN}$$

*Precision* adalah rasio antara prediksi positif yang benar (TP) dengan total prediksi positif yang dilakukan (TP + FP). *Precision* menunjukkan sejauh mana prediksi positif yang dibuat oleh model benar-benar relevan.

$$\text{Precision} = \frac{TP}{TP + FP}$$

### 3. HASIL DAN PEMBAHASAN

Berikut ini adalah pembahasan mengenai perhitungan algoritma *K-Nearest Neighbors* (KNN), tahap implementasi menggunakan *Microsoft Excel*, serta pengujian model dan evaluasi akurasi menggunakan perangkat lunak *RapidMiner*.

#### 1) Perhitungan dengan *Microsoft Excel*

*Microsoft Excel* adalah aplikasi *spreadsheet* yang memungkinkan pengguna untuk mengelola, menganalisis, dan memvisualisasikan data [14]. *Excel* menyediakan berbagai fungsi, rumus, dan alat grafik untuk mendukung analisis data yang efektif dan efisien. Pada kasus ini, kita akan menghitung jarak antara titik uji dan seluruh data dalam dataset untuk menemukan K tetangga terdekat yang memiliki kemiripan karakteristik lingkungan. Setiap fitur lingkungan, seperti pH Tanah, Curah Hujan, Suhu Rata-rata, Kadar Air (%), dan Ketinggian, digunakan dalam perhitungan jarak Euclidean. Berikut adalah langkah-langkah penerapan algoritma KNN di *Excel*:

##### 1. Input Data

Penulis memasukkan data observasi dari setiap titik lokasi penanaman kopi sebagai data latih/data *training* serta data titik uji yang akan diklasifikasikan ke dalam lembar kerja *Excel* sebagai berikut.

**Tabel 2.** Data Observasi Kualitas Kopi

Bulan	Lokasi	pH Tanah	Curah Hujan	Suhu Rata-rata	Kadar Air (%)	Ketinggian	Kualitas Kopi
Januari	Sumbermalang	6,45	175	23,7	12,5	1355	Baik
Januari	Mlandingan	6,28	148	25,5	10,9	890	Cukup
Januari	Kayumas	6,37	165	23	11,6	1460	Baik
Januari	Jatibanteng	6,22	155	24,9	11,4	960	Cukup
Februari	Sumbermalang	6,48	180	23,5	12,7	1360	Baik
Februari	Mlandingan	6,31	150	25,6	10,8	900	Cukup
Februari	Kayumas	6,42	170	23,3	11,7	1470	Baik
Februari	Jatibanteng	6,25	158	24,6	11,5	970	Cukup
Maret	Sumbermalang	6,5	185	23,4	12,8	1370	Baik
Maret	Mlandingan	6,34	152	25,7	10,6	910	Cukup
Maret	Kayumas	6,44	175	23,1	11,8	1480	Baik
Maret	Jatibanteng	6,28	160	24,5	11,6	980	Cukup
April	Sumbermalang	6,52	190	23,3	12,9	1375	Baik
April	Mlandingan	6,36	155	25,9	10,7	920	Cukup
April	Kayumas	6,46	180	23	11,9	1490	Baik

2. Tentukan Titik yang Akan Diklasifikasikan

Tambahkan satu baris kosong di bawah data sebagai titik uji/data *testing* yang akan kita klasifikasikan menggunakan KNN. Misalnya, kita ingin mengetahui kualitas kopi dari titik dengan karakteristik berikut:

- pH Tanah: 6.25
- Curah Hujan: 130
- Suhu Rata-rata: 25.5
- Kadar Air (%): 11.0
- Ketinggian : 1300 mdpl

3. Perhitungan Jarak *Euclidean*

Gunakan rumus jarak *Euclidean* untuk menghitung jarak dari setiap titik kopi ke titik uji. Rumus jarak *Euclidean* untuk dua titik dengan beberapa fitur (atau dimensi) adalah sebagai berikut:

$$\text{Jarak} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

Masukkan rumus berikut untuk baris pertama data, sesuaikan kolomnya sesuai lokasi data di *Excel*. Misalnya, data pertama berada di baris 2 dan titik uji di baris 18, dan salin untuk baris setelahnya:

$$=\text{SQRT}((\text{C2} - \text{C18})^2 + (\text{D2} - \text{D18})^2 + (\text{E2} - \text{E18})^2 + (\text{F2} - \text{F18})^2 + (\text{G2} - \text{G18})^2)$$

	A	B	C	D	E	F	G	H	I
1	Bulan	Lokasi	pH Tanah	Curah Huj	Suhu Rata-ra	Kadar Air (%)	Ketinggian	Kualitas	ED
2	Januari	Sumbermalang	6,45	175	23,7	12,5	1355	Baik	71,10225
3	Januari	Mlandingan	6,28	148	25,5	10,9	890	Cukup	410,39495
4	Januari	Kayumas	6,37	165	23	11,6	1460	Baik	163,80362
5	Januari	Jatibanteng	6,22	155	24,9	11,4	960	Cukup	340,91864
6	Februari	Sumbermalang	6,48	180	23,5	12,7	1360	Baik	78,14693
7	Februari	Mlandingan	6,31	150	25,6	10,8	900	Cukup	400,49975
8	Februari	Kayumas	6,42	170	23,3	11,7	1470	Baik	174,65783
9	Februari	Jatibanteng	6,25	158	24,6	11,5	970	Cukup	331,18735
10	Maret	Sumbermalang	6,5	185	23,4	12,8	1370	Baik	89,06578
11	Maret	Mlandingan	6,34	152	25,7	10,6	910	Cukup	390,62029
12	Maret	Kayumas	6,44	175	23,1	11,8	1480	Baik	185,55710
13	Maret	Jatibanteng	6,28	160	24,5	11,6	980	Cukup	321,40529
14	April	Sumbermalang	6,52	190	23,3	12,9	1375	Baik	96,09122
15	April	Mlandingan	6,36	155	25,9	10,7	920	Cukup	380,82182
16	April	Kayumas	6,46	180	23	11,9	1490	Baik	196,48691
17	<b>DATA BARU</b>		<b>6,25</b>	<b>130</b>	<b>25,5</b>	<b>11,0</b>	<b>1300</b>		

**Gambar 2.** Hasil Perhitungan *Euclidean Distance*

4. Menentukan Nilai K dan Urutkan Jarak

Tentukan nilai K (misalnya, K=3 atau K=5), lalu urutkan kolom jarak dari nilai terkecil hingga terbesar. Pilih K nilai teratas dalam daftar jarak ini untuk menemukan tetangga terdekat.

5. Menentukan Nilai K dan Urutkan Jarak

Dari K tetangga terdekat, periksa frekuensi kualitas kopi (misalnya, "Baik" atau "Cukup"). Klasifikasi untuk titik uji diambil berdasarkan mayoritas kualitas kopi pada K tetangga terdekat.

	A	B	C	D	E	F	G	H	I	J
1	Bulan	Lokasi	pH Tanah	Curah Huj	Suhu Rata-ra	Kadar Air (%)	Ketinggian	Kualitas	ED	
2	Januari	Sumbermalang	6,45	175	23,7	12,5	1355	Baik	71,10225	
3	Januari	Mlandingan	6,28	148	25,5	10,9	890	Cukup	410,39495	1
4	Januari	Kayumas	6,37	165	23	11,6	1460	Baik	163,80362	
5	Januari	Jatibanteng	6,22	155	24,9	11,4	960	Cukup	340,91864	5
6	Februari	Sumbermalang	6,48	180	23,5	12,7	1360	Baik	78,14693	
7	Februari	Mlandingan	6,31	150	25,6	10,8	900	Cukup	400,49975	2
8	Februari	Kayumas	6,42	170	23,3	11,7	1470	Baik	174,65783	
9	Februari	Jatibanteng	6,25	158	24,6	11,5	970	Cukup	331,18735	
10	Maret	Sumbermalang	6,5	185	23,4	12,8	1370	Baik	89,06578	
11	Maret	Mlandingan	6,34	152	25,7	10,6	910	Cukup	390,62029	3
12	Maret	Kayumas	6,44	175	23,1	11,8	1480	Baik	185,55710	
13	Maret	Jatibanteng	6,28	160	24,5	11,6	980	Cukup	321,40529	
14	April	Sumbermalang	6,52	190	23,3	12,9	1375	Baik	96,09122	
15	April	Mlandingan	6,36	155	25,9	10,7	920	Cukup	380,82182	4
16	April	Kayumas	6,46	180	23	11,9	1490	Baik	196,48691	
17	DATA BARU		6,25	130	25,5	11,0	1300			
18	NILAI K						5			

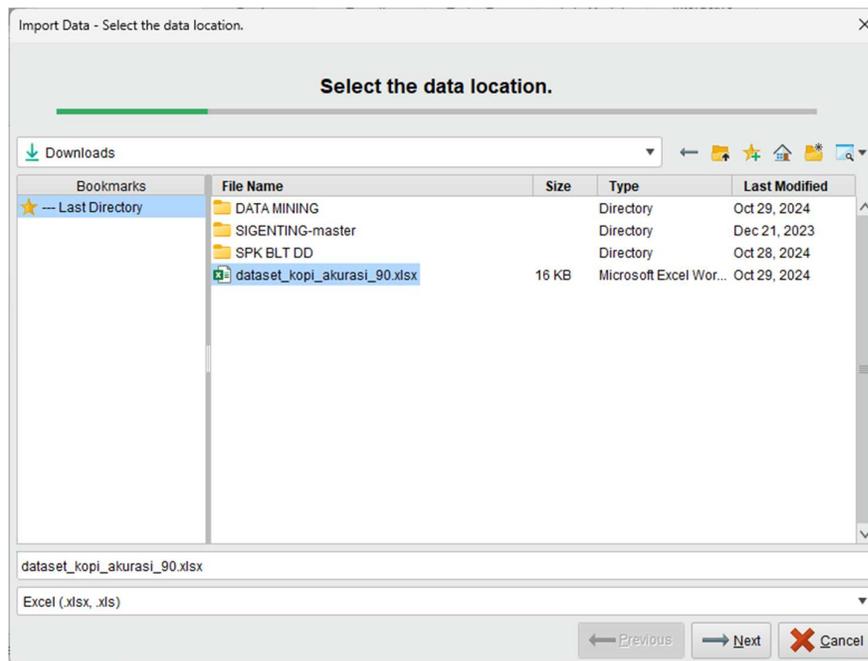
Gambar 3. Hasil Prediksi K-Nearest Neighbor

## 2) Pengujian Model dan Akurasi Menggunakan *RapidMiner*

Setelah perhitungan KNN dilakukan di *Excel*, langkah berikutnya adalah melakukan pengujian akurasi model dan mengevaluasi performa model dengan menggunakan perangkat lunak *RapidMiner*. *RapidMiner* adalah aplikasi yang menyediakan fitur lengkap untuk pengujian model *machine learning* [15], termasuk algoritma KNN. Langkah-langkah pengujian model di *RapidMiner* adalah sebagai berikut:

### 1. Impor Data ke *RapidMiner*

Impor dataset yang telah dibuat di *Excel* ke dalam *RapidMiner*. Data ini harus mencakup fitur lingkungan yang relevan dan label kualitas kopi sebagai variabel target.



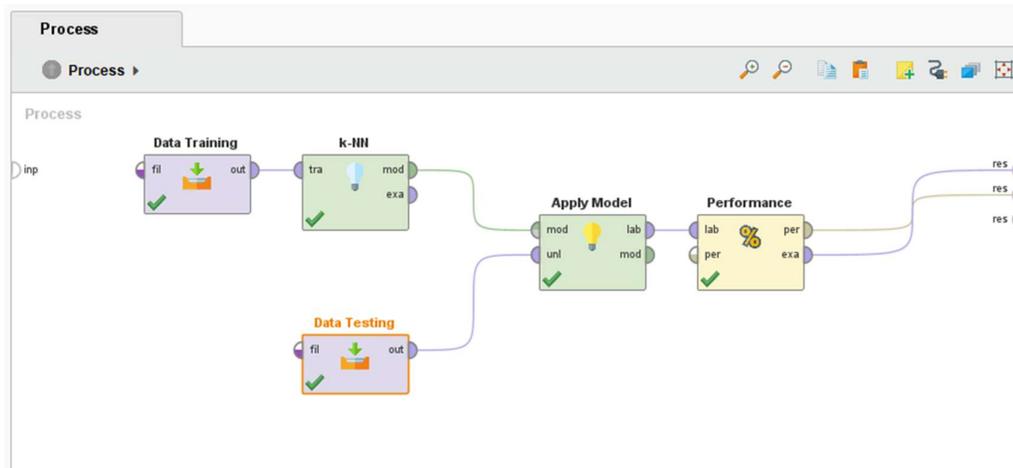
Gambar 4. Halaman Impor Data

### 2. Pembagian Data

Bagi dataset menjadi dua bagian, yaitu data latih dan data uji, biasanya dengan proporsi 80% untuk data latih dan 20% untuk data uji. Pembagian ini bertujuan untuk melatih model dengan sebagian besar data, sementara sebagian kecil data digunakan untuk menguji performa model.

### 3. Pengaturan Algoritma KNN

Pilih operator *K-Nearest Neighbors* di *RapidMiner* dan atur parameter nilai K sesuai dengan jumlah tetangga yang diinginkan, misalnya K=3 atau K=5, sesuai hasil eksperimen di *Excel*.



Gambar 5. Proses Pengujian Model

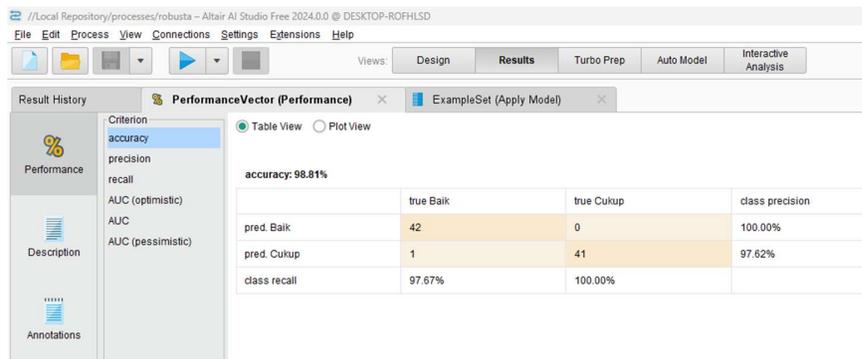
### 4. Evaluasi Model

Jalankan proses untuk melatih dan menguji model KNN. *RapidMiner* akan menghasilkan laporan akurasi yang mencakup berbagai metrik evaluasi, seperti *Precision*, *Recall*, dan *Confusion Matrix*, untuk menunjukkan seberapa baik model dalam mengklasifikasikan data uji berdasarkan kondisi lingkungan. *Confusion Matrix* akan memberikan informasi mengenai jumlah prediksi yang benar dan salah dalam klasifikasi kualitas kopi.

Row No.	Kualitas	prediction(K...	confidence(...	confidence(...	Bulan	Lokasi	pH Tanah	Curah Hujan	Suhu R...
36	Cukup	Cukup	0.179	0.821	Juni	Jatibanteng	6.38	168	24.3
37	Baik	Baik	1	0	Juli	Sumbermala...	6.48	180	22.7
38	Cukup	Cukup	0	1	Juli	Mlandingan	6.25	170	26.0
39	Baik	Baik	1.000	0	Juli	Kayumas	6.38	188	23.6
40	Baik	Cukup	0.250	0.750	Juli	Jatibanteng	6.32	175	24.8
41	Baik	Baik	1	0	Juli	Sumbermala...	6.6	205	22.9
42	Cukup	Cukup	0	1	Juli	Mlandingan	6.43	162	25.5
43	Baik	Baik	1	0	Juli	Kayumas	6.52	195	22.6
44	Cukup	Cukup	0	1	Juli	Jatibanteng	6.36	180	24.1
45	Baik	Baik	1	0	Agustus	Sumbermala...	6.62	210	22.8
46	Cukup	Cukup	0	1	Agustus	Mlandingan	6.45	165	25.4
47	Baik	Baik	1	0	Agustus	Kayumas	6.54	200	22.5
48	Cukup	Cukup	0.185	0.815	Agustus	Jatibanteng	6.38	185	24
49	Cukup	Cukup	0	1	Agustus	Mlandingan	6.40	180	25.8

Gambar 6. Hasil Pengujian Model

Hasil pengujian menunjukkan bahwa model memiliki tingkat akurasi sebesar 98,81%, yang berarti 98,81% dari keseluruhan data uji berhasil diprediksi dengan benar sesuai dengan label asli data.

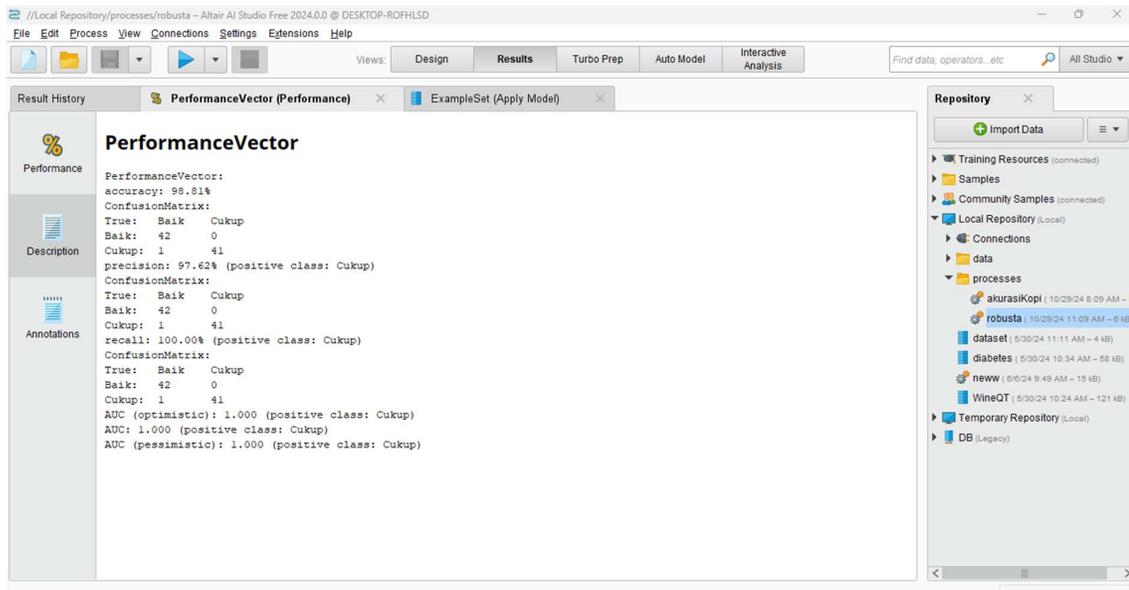


**Gambar 7.** Confusion Matrix

*Confusion matrix* menunjukkan distribusi hasil prediksi model, di mana dari 84 sampel, model berhasil memprediksi 40 data dengan benar sebagai kelas 'Baik' dan 41 data sebagai kelas 'Cukup.' Namun, ada 1 data kelas 'Cukup' yang salah diprediksi sebagai 'Baik'.

*Recall* untuk kelas 'Baik' mencapai 97,67%, menunjukkan bahwa dari seluruh data yang sebenarnya masuk dalam kelas 'Baik', model berhasil mengidentifikasi 97,67% dengan benar. Namun, *recall* untuk kelas 'Cukup' sebesar 100%, menandakan bahwa model sangat optimal dalam mendeteksi semua data yang masuk dalam kelas 'Cukup'.

Model menunjukkan *precision* sebesar 100% untuk kelas 'Baik', yang berarti model sudah optimal akurasi dari semua prediksi yang masuk dalam kelas 'Baik'. Sementara untuk kelas 'Cukup,' *precision* tercatat sebesar 97,62%, menandakan bahwa model menghasilkan hanya satu *false positive* pada prediksi kelas ini.



**Gambar 8.** Performance Vector

#### 4. KESIMPULAN DAN SARAN

Algoritma *K-Nearest Neighbor* (KNN) efektif dalam mengklasifikasikan kualitas kopi robusta di PT. Indokom Citra Persada Situbondo, dengan tingkat akurasi mencapai 98,81%. Model ini menunjukkan kemampuan yang baik dalam mengidentifikasi kualitas kopi berdasarkan faktor lingkungan yang beragam, sehingga dapat membantu perusahaan dalam mempertahankan konsistensi dan meningkatkan mutu produk. Untuk penelitian selanjutnya, disarankan agar dilakukan pengujian lebih lanjut dengan mempertimbangkan lebih banyak

metode klasifikasi untuk data *mining* sehingga dapat dihasilkan nilai perbandingan dan dapat memilih metode yang paling tepat dan akurat sebagai solusi.

## 5. DAFTAR PUSTAKA

- [1] M. P. Ir Muhammad Rizwan, *Budidaya Kopi*. CV. AZKA PUSTAKA, 2022.
- [2] S. Prijono, A. A. Hanuf, J. Y. Saputri, A. Khoirunnisak, Y. M. Nurin, and D. M. Yunita, *Pengelolaan Tanah di Kebun Kopi*. Universitas Brawijaya Press, 2021.
- [3] R. M. Sari, *Klasifikasi Data Mining*. Serasi Media Teknologi, 2024.
- [4] V. R. Prasetyo, I. A. Ryanda, and D. A. Prima, “Analisis Sentimen Dan Kategorisasi Review Pelanggan Pada Cafe Kopi Paste Dengan Metode Naive Bayes Dan K-Nearest Neighbor,” *NERO (Networking Eng. Res. Oper.*, vol. 8, no. 1, pp. 1–8, 2023.
- [5] A. Tangkelayuk, “The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naive Bayes, dan Decision Tree,” *JATISI (Jurnal Tek. Inform. Dan Sist. Informasi)*, vol. 9, no. 2, pp. 1109–1119, 2022.
- [6] N. A. D. Siboro, “Perbandingan Metode Naive Bayes dan KNN dalam Penerapan Data Mining untuk Klasifikasi Menu Potensial (Studi Kasus: Ateku Kopi Medan).” Universitas Sumatera Utara, 2024.
- [7] M. Jundanuddin, Z. Fatah, and A. Munazilin, “Implementasi Data Mining Pada Penilaian Kinerja Guru Menggunakan Metode K-Nearest Neighbor (K-Nn) Studi Kasus pada Mi Alfagiri Silo-Jember,” *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 13, no. 4, pp. 869–880, 2024.
- [8] E. Ardyan *et al.*, *Metode Penelitian Kualitatif dan Kuantitatif: Pendekatan Metode Kualitatif dan Kuantitatif di Berbagai Bidang*. PT. Sonpedia Publishing Indonesia, 2023.
- [9] Shudiq, W. J. F., As, A. H., & Rahman, M. F. (2020). Penentuan Metode Terbaik Dalam Menentukan Jenis Pohon Pisang Menurut Tekstur Daun (Metode K-NN dan SVM). *Jurnal Teknologi dan Manajemen Informatika*, 6(2), 128-136.3.
- [10] F. Sulianta, *Basic Data Mining from A to Z*. Feri Sulianta, 2023.
- [11] F. Anggraeni, N. Kristiawan, R. Lutfiati, Y. Dirgantara, and P. Rosyani, “Prediksi Cuaca Yang Akan Datang Menggunakan Metode Data Mining,” *Newt. J. Mat. Fis. Algoritm. dan Sains*, vol. 1, no. 1, pp. 73–83, 2023.
- [12] V. Artanti, M. Faisal, and F. Kurniawan, “Klasifikasi Cardiovascular Diseases Menggunakan Algoritma K-Nearest Neighbors (KNN),” *Techno. Com*, vol. 23, no. 2, pp. 469–481, 2024.
- [13] M. Putri, “Prediksi Penyakit Stroke Menggunakan Machine Learning Dengan Algoritma Random Forest,” *J. Infomedia Tek. Inform. Multimedia, dan Jar.*, vol. 9, no. 1, pp. 16–21, 2024.
- [14] A. Subarno and P. Ninghardjanti, *Teknologi Administrasi Kantor*. PENERBIT KBM INDONESIA, 2023.
- [15] Permatasari, U. O. R., Shudiq, W. J. F., & Jasri, M. (2024). Prediksi Kelayakan Mahasiswa sebagai Penerima Beasiswa Bank Indonesia pada Tahap Seleksi Administrasi di Universitas Nurul Jadid Menggunakan Algoritma K Nearest Neighbor. *Journal of Electrical Engineering and Computer (JEECOM)*, 6(1), 252-260.
- [16] D. Jollyta, A. Hajjah, E. Haerani, and M. Siddik, *Algoritma Klasifikasi untuk Pemula Solusi Python dan RapidMiner*. Deepublish, 2023.