

PERBANDINGAN METODE NMC DENGAN K-NN UNTUK PEMILIHAN BIBIT KELAPA DALAM

¹Muh.Rasyid Ridha, ²Amelia

^{1,2}Sistem Informasi, Fakultas Teknik Dan Ilmu Komputer, Universitas Islam Indragiri

Jl.Provinsi Parit 1 Tembilahan Hulu, Tembilahan, Riau-Indonesia

Email: rasyid4sky@gmail.com, ameliafbriyani@gmail.com

ABSTRAK

Pemilihan bibit kelapa yang berkualitas tinggi sangat penting untuk mendukung produktivitas dan keberlanjutan pertanian kelapa. Dalam penelitian ini, kami membandingkan dua metode klasifikasi, yaitu Nearest Mean Classifier (NMC) dan K-Nearest Neighbors (K-NN), untuk mengidentifikasi bibit kelapa lokal yang unggul. NMC bekerja dengan menghitung rata-rata fitur dari setiap kelas dan mengklasifikasikan data uji berdasarkan kedekatan ke rata-rata ini. Sebaliknya, K-NN mengklasifikasikan data uji berdasarkan mayoritas label dari k tetangga terdekatnya dalam ruang fitur. Dataset yang terdiri dari berbagai parameter morfologi dan genetik bibit kelapa lokal, kemudian menerapkan kedua metode klasifikasi tersebut. Hasil penelitian menunjukkan bahwa metode NMC memberikan akurasi sebesar 0,56, sementara metode K-NN hanya mencapai akurasi sebesar 0,22. Hasil ini mengindikasikan bahwa NMC lebih unggul dalam mengklasifikasikan bibit kelapa lokal dibandingkan K-NN.

Keywords: Pemilihan Bibit Kelapa Lokal, NMC, K-NN

1 PENDAHULUAN

Kelapa memiliki variasi genetik yang besar, secara umum pembiakannya dilakukan secara *generative*. Pemilihan bibit kelapa merupakan langkah dalam memastikan produksi kelapa yang optimal. Pemilihan bibit yang tepat menjadi faktor penentu dalam meningkatkan hasil panen dan kesejahteraan petani.

Berdasarkan penelati sebelumnya telah melakukan penelitian tentang pemilihan bibit kelapa menggunakan metode *Nearest Mean Classifier (NMC)* sedangkan pada penelitian ini membandingkan metode *Nearest Mean Classifier (NMC)* dan *K-Nearest Neighbor(K-NN)*.

NMC metode klasifikasi yang berfokus pada perbedaan mean atau rata-rata antara kelas-kelas yang berbeda. Metode ini mencoba untuk menemukan mean atau rata-rata dari setiap kelas dan mengklasifikasikan instance baru kekelas dengan mean terdekat. Sementara itu, K-NN metode klasifikasi yang berbasis pada “kedekatan” atau “keberdekatan” antara instance yang akan diklasifikasikan dengan instance-instance yang sudah ada dalam dataset. Instance baru akan diklarifikasikan berdasarkan mayoritas kelas dari K-nearest Neighbor (tetanga terdekat) dari instance tersebut.

2 TINJAUAN PUSTAKA

Nearest Mean Classifier (NMC) di kembangkan dalam penelitian untuk prediksi objek seperti buah-buahan yang bertindak sebagai pemilah menggunakan kemiripan antar pola untuk melakukan klasifikasi. Untuk setiap kelas, NMC menghitung rerata kelas (*centroid*) dari data sampel. Perolehan kemiripan melalui perhitungan jarak *Euclidean* antara pola yang tidak diketahui dengan *centroid* dari data sampel. NMC mengklasifikasikan sembarang pola yang tidak diketahui ke kelas dengan *centroid* terdekat dengan pola yang akan dikenali. NMC berhasil diterapkan pada banyak masalah klasifikasi dan menunjukkan hasil yang baik. Selain itu NMC memberikan kemampuan yang bagus untuk persoalan sampel kecil dimana jumlah sampel jauh lebih kecil dari jumlah fitur.[1]

NMC mengklasifikasikan sebuah titik data berdasarkan jarak terdekatnya ke rata-rata (mean) dari setiap kelas.

Ridha, Perbandingan Metode Nmc Dengan K-Nn Untuk Pemilihan Bibit Kelapa Dalam

Berikut adalah langkah-langkah umum untuk melakukan klasifikasi menggunakan NMC:

- a) Hitung Mean Setiap Kelas: Hitung rata-rata dari setiap fitur untuk setiap kelas dalam data pelatihan.
- b) Hitung Jarak ke Mean Setiap Kelas: Untuk setiap titik data yang akan diklasifikasikan, hitung jarak antara titik tersebut dengan mean dari setiap kelas.
- c) Pilih Kelas dengan Jarak Terkecil: Klasifikasikan titik data ke kelas yang mean-nya memiliki jarak terkecil ke titik data tersebut.

Rumus untuk menghitung mean dari setiap kelas k adalah:

$$\mu_k = \frac{1}{N_k} \sum_{i \in C_k} x_i$$

Dimana Keterangan Rumus :

- a) μ_k adalah mean dari kelas k
- b) N_k adalah jumlah sampel dalam kelas k
- c) x_i adalah vektor fitur dari sampel i dalam kelas k
- d) C_k adalah kumpulan sampel dalam kelas k [1]

Algoritma K-NN(K-Nearest Neighbor) menentukan nilai jarak pada pengujian data testing dengan data training berdasarkan nilai terkecil dari nilai ketetanggaan terdekat. Tujuannya untuk mengklasifikasikan objek baru berdasarkan atribut dan training samples. K-NN melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data lain. Nilai K pada K-NN berarti K-data terdekat dari data uji. K-NN memiliki kelebihan dapat menghasilkan data yang kuat atau jelas dan efektif jika digunakan pada data yang besar. K-NN juga memiliki kekurangan yaitu membutuhkan nilai k, jarak dari data percobaan tidak jelas dengan tipe jarak yang digunakan, untuk memperoleh hasil yang terbaik, maka harus menggunakan semua atribut atau hanya satu atribut yang telah pasti. Metode K-NN cukup sederhana, tidak ada asumsi mengenai distribusi data dan mudah diaplikasikan. Pemilihan nilai K (jumlah data/tetangga terdekat) ditentukan oleh peneliti. Pemilihan nilai K ini bisa mempengaruhi tingkat akurasi prediksi yang dikerjakan.[2]

Adapun tahapan dari algoritma K-Nearest Neighbor (K-NN) dijelaskan sebagai berikut :

- a) Persiapan data training dan data testing
- b) Menentukan nilai k
- c) Menghitung jarak data testing ke setiap data pelatihan.
- d) Data pelatihan di hitung menggunakan rumus penghitung jarak Euclidean sebagai
- e) berikut :

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

Dimana Keterangan Rumus :

- a) $d(x,y)$ adalah jarak Euclidean antara dua titik data x dan y .
- b) x_i dan y_i adalah komponen ke- i dari vektor x dan y .
- c) n adalah jumlah fitur atau dimensi. [3]

Sistem dapat didefinisikan sebagai kumpulan dari prosedur-prosedur yang memiliki tujuan tertentu. Prosedur adalah urutan yang tepat dari tahapan-tahapan instruksi yang menerangkan apa (*what*) yang harus dikerjakan, siapa (*who*) yang mengerjakannya, kapan (*when*) dikerjakan dan bagaimana (*how*) mengerjakannya.[4] Klasifikasi pengelompokan benda yang sama serta memisahkan benda yang tidak sama. Secara harfiah arti klasifikasi adalah penggolongan, pengelompokan. Dalam kaitannya di dunia perpustakaan klasifikasi diartikan sebagai kegiatan pengelompokan bahan pustaka berdasarkan ciri-ciri yang sama, misalnya pengarang, fisik, isi dan sebagainya. Menentukan sebuah record data baru ke salah satu dari beberapa katagori/kelas yang telah didefinisikan sebelumnya. Disebut juga dengan *supervised learning*. Berikut beberapa aplikasi dari klasifikasi; a) Penjualan langsung (*direct marketing*) tujuannya mengurangi *cost* surat menyurat dengan menentukan (*targeting*) satu set konsumen yang mempunyai kesamaan dalam membeli produk telepon selular baru; b) *Fraud detection* tujuannya memprediksi kasus transaksi curang

dengan menggunakan kartu kredit; c) *Customer Attrition/churn* tujuannya memprediksi pelanggan mana yang akan berpindah ke kompetitor kita.

Citra dengan $f(x,y)$ yang nilainya didigitalisasikan baik dalam koordinat spasialnya maupun dalam *gray level*-nya. Sebuah citra diubah ke dalam bentuk digital agar mudah diolah atau disimpan dalam memori komputer atau media lain, Citra ialah istilah lain untuk gambar sebagai salah satu komponen multimedia yang memiliki peranan penting dalam bentuk informasi visual.[5] *Pattern recognition* teknik yang bertujuan untuk mengklasifikasikan citra yang telah diolah sebelumnya berdasarkan kesamaan atau kemiripan ciri yang dimilikinya. Secara umum teknik pengenalan pola bertujuan mengklasifikasikan dan mendiskripsi pola atau objek yang kompleks melalui pengukuran sifat atau ciri-ciri objek bersangkutan. Prinsip kerjanya menirukan kemampuan manusia mengenali objek-objek berdasarkan ciri-ciri dan pengetahuan yang pernah diamatinya dari objek tersebut. Cara kerjanya, mengklasifikasikan objek ke dalam kelas tertentu berdasarkan parameter yang telah disimpan atau ditentukan sebelumnya. Contoh aplikasi *speech recognition*, *data mining*, *biometrics*, dll.[6]

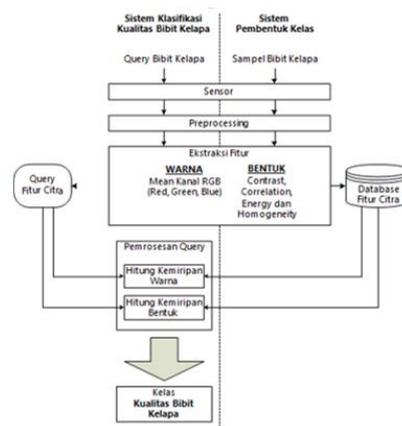
Manusia memiliki kemampuan yang luar biasa didalam pengenalan objek berdasarkan ciri-ciri atau pengetahuan yang pernah diamatinya. Misalnya manusia dengan mudah membedakan antara burung dengan kupu-kupu, kursi dengan meja. Mempelajari cara bagaimana manusia mempelajari pola merupakan masalah yang sangat menarik.

Proses pengenalan pola mendeskripsikan tahap-tahapan dalam pengenalan pola. Hal ini di maksudkan agar adanya langkah-langkah terstruktur dalam melakukan pengenalan pola yaitu: a) Data (gambar, bunyi, teks) untuk dikelaskan; b) Menghilangkan gangguan/menormalkan gambar (*Image processing*); c) Pengiraan citra; dan d) Pengenalan kelas (mengambil keputusan). Sebagai penelitian yakni Kelapa (*Cocos nucifera*) pohon yang mempunyai nilai ekonomi yang tinggi. Buah kelapa terbungkus serabut dan batok yang cukup kuat. Setiap bagian dari pohon kelapa dapat dimanfaatkan bagi kepentingan manusia. Oleh karena itu, tanaman kelapa sering disebut sebagai pohon kehidupan.[7]

3 METODE PENELITIAN

Agar pembuatan sistem klasifikasi kualitas bibit kelapa dapat berjalan dengan baik dan terstruktur, maka tahapan dalam penelitian ini digambarkan pada sketsa pemrosesan data model *waterfall* [8] sebagaimana berikut:

- a) Tahapan Pengumpulan data. Pada tahap Pengumpulan data dalam penelitian dimaksudkan guna mendapatkan bahan, keterangan, kenyataan, dan informasi yang bisa dipercaya. Analisa dan ekstraksi fitur bibit kelapa dalam tahapan perencanaan meliputi fitur pengumpulan data dan observasi yang dilakukan di Kabupaten Indragiri Hilir, memilah data bibit kelapa yang akan di klasifikasi, klasifikasi didasarkan pada bentuk, ukuran dan warna RGB. Memilah citra input dan uji, melakukan klasifikasi citra latih dan uji, mengklasifikasikan citra fitur bibit kelapa.
- b) Tahapan Desain Program. Setelah data-data yang diperlukan terkumpul langkah selanjutnya adalah perancangan program sistem klasifikasi kualitas bibit kelapa. Penelitian ini menggunakan konsep *data mining* dengan menggunakan metode *Neares Mean Classifier (NMC)* dan *K-Nearest Neighbor(K-NN)*. Merancang bentuk arsitektur sistem klasifikasi kualitas bibit kelapa yang menggambarkan cara kerja aplikasi sistem klasifikasi kualitas bibit kelapa.



Gambar 1. Arsitektur Sistem Klasifikasi Kualitas Bibit Kelapa

- a) Tahapan Pembuatan Program. Tahapan pembuatan program dilakukan dengan mengacu pada hasil perancangan yang di uraikan pada tahapan perancangan yang sesuai dengan kebutuhan.
- b) Tahapan Pengujian. Tahapan pengujian sistem ini dilakukan untuk menguji apakah sistem sudah siap digunakan, pengujian sistem dilakukan dengan prosedur White-Box dan Black-Box
- c) Tahapan Analisa. Tahap ini merupakan tahapan menganalisa, yaitu apakah sistem yang digunakan membantu pengguna sistem atau sesuai dengan kebutuhan.

4 HASIL DAN PEMBAHASAN

Penelitian ini yang menjadi inputan sistem klasifikasi kualitas citra bibit kelapa adalah berupa citra bibit kelapa. Citra sampel berjumlah 30 buah (3 kelas kualitas bibit kelapa, masing-masing kelas 10 citra).

Tabel I. Data Citra Latih

No.	Kelas	Jumlah
1	Kelas rata-rata Grade A	10
2	Kelas rata-rata Grade B	10
3	Kelas rata-rata Grade C	10
Total		30

Citra digital adalah citra dengan $f(x,y)$ yang nilainya didigitalisasikan baik dalam koordinat spasialnya maupun dalam *gray level*-nya., Citra ialah istilah lain untuk gambar sebagai salah satu komponen multimedia yang memiliki peranan penting dalam bentuk informasi visual.

Citra uji berjumlah 30 buah (3 kelas kualitas kopra, masing-masing kelas 10 citra). Semua citra, baik citra sample ataupun citra uji terlebih dahulu akan dikonversi ke format BMP dan diubah resolusinya menjadi 640 x 480 piksel.

4.1 Analisa Proses

Tiap kelas kualitas dibedakan berdasarkan nilai fitur warna (*red, green, blue*) dan tekstur (*contrast, correlation, energy dan homogeneity*). Semakin berwarna coklat keabuan-abuan, dan bentuk lonjong dan pipih bagian bawah maka bibit kelapa tersebut semakin baik kualitasnya. Setelah diketahui karakteristik kelas kualitas bibit kelapa, maka analisa proses pembentukan kelas bisa dilakukan. Proses pembentukan kelas diawali dengan melakukan ekstraksi fitur citra data latih. Tiap citra akan diekstrak untuk mendapatkan nilai *red, green, blue* dari variabel warna kemudian *contrast, correlation, energy dan homogeneity* dari variabel tekstur. Hasilnya akan disimpan di dalam database fitur citra, yakni di dalam tabel fitur citra latih.

Proses Perhitungan Jarak dan Identifikasi Kualitas Bibit Kelapa, proses ini diawali dengan pemanggilan citra uji (Citra Query) yang ingin dicari kelas kualitasnya, hasil citra query akan diekstrak untuk mendapatkan vector fitur citra query, Setelah didapatkan vektor fitur *Red, Green dan Blue* dari variabel warna kemudian *energy, contrast, correlation dan homogeneity* dari variabel tekstur. Perbandingan tersebut dapat dilakukan dengan cara menghitung jarak *euclidean distance*

yang merupakan selisih nilai piksel antara 2 vektor tersebut. Adapun rumus *euclidean distance* ditunjukkan pada persamaan berikut ini:

$$\text{Dist}(i,k) = \sqrt{\sum_{i=j}^D (i_j - k_j)^2},$$

Dengan $\text{dist}(i,k)$ adalah jarak *euclidean* antara vektor i dan vektor k_j = komponen ke j dari vektor i

k_j = komponen ke j dari vektor k

D = adalah jumlah komponen pada vektor i dan vektor k .

Dari hasil perhitungana jarak *euclidean* tersebut dapat ditentukan suatu citra query adalah mirip bila memiliki jarak yang paling dekat atau nilainya paling kecil.[9]

4.2 Analisa Citra Latih

Pada Tabel 4 merupakan hasil dari proses fitur latih, kualitas bibit kelapa dibagi kedalam kelas bibit Grade A (80-85)%, Grade A (70-7)%, dan Grade C (60-65)% berdasarkan variabel warna RGB (*Red Green Blue*).dimana dilakukan citra latih untuk 30 Jenis Bibit Kelapa yang berbeda yang dikelompokkan berdasarkan *Grade A, Grade B dan Grade C*.

Table 2. Database Citra Latih

No	Warna RGB			Texsture				Kelas (Grade)
	Red	green	Blue	contrast	correlation	energy	homogeneity	
1	207	159	127	0.19549	0.92197	0.52511	0.94904	A
2	191	165	141	0.38007	0.85114	0.53196	0.91931	A
3	198	182	156	0.28098	0.85695	0.59057	0.94675	A
4	194	177	150	0.38497	0.82575	0.59906	0.92838	A
5	190	170	149	0.344	0.88702	0.51422	0.92907	A
6	216	189	156	0.16532	0.89522	0.49762	0.95973	A
7	197	162	132	0.15753	0.94251	0.60499	0.9676	A
8	179	152	126	0.19365	0.93941	0.55527	0.96952	A
9	167	143	121	0.50924	0.87592	0.51844	0.93899	A
10	170	148	123	0.39577	0.90232	0.47072	0.94541	A
11	169	151	132	0.53944	0.85987	0.49505	0.92654	B
12	176	136	108	0.37075	0.92512	0.40051	0.9498	B
13	182	148	119	0.19421	0.95022	0.43443	0.95427	B
14	166	133	117	0.3873	0.9244	0.43754	0.93494	B
15	188	145	119	0.28961	0.92834	0.42736	0.94186	B
16	185	166	146	0.24684	0.92728	0.48688	0.94574	B
17	190	159	136	0.24722	0.93015	0.4345	0.94791	B
18	193	160	134	0.21519	0.935	0.46796	0.96324	B
19	213	185	151	0.34537	0.842	0.49011	0.94002	B
20	182	155	132	0.22212	0.93376	0.51685	0.96795	B
21	189	160	136	0.18885	0.92558	0.56785	0.96571	C
22	191	165	141	0.14484	0.94504	0.52099	0.96843	C
23	170	141	122	0.93561	0.80219	0.45287	0.92195	C
24	182	155	136	0.45029	0.89404	0.40869	0.95395	C
25	189	161	138	0.29611	0.92038	0.46095	0.95042	C
26	195	169	148	0.52735	0.85075	0.47835	0.92733	C
27	212	187	169	0.28504	0.86161	0.42752	0.91608	C
28	195	173	149	0.21311	0.92172	0.50081	0.94931	C
29	183	145	120	0.39593	0.90532	0.37351	0.91268	C
30	180	148	131	0.53324	0.86049	0.46606	0.9237	C

4.3 Analisa Citra Uji

Misalkan kita ambil satu contoh bibit kelapa ada sebuah citra *query* yang belum di ketahui masuk label kelasnya dan masuk kualitas *Grade A, Grade B atau Grade C*. Vektor Tersebut yaitu: **red**

Ridha, Perbandingan Metode Nmc Dengan K-Nn Untuk Pemilihan Bibit Kelapa Dalam

= 181, green = 149 dan blue = 132. Adapun tahapan perhitungan menggunakan Nearest mean classifier (NMC) adalah sebagai berikut: Menghitung jarak antara vektor fitur citra query dengan database vektor fitur data citra latih pada Tabel 3 menggunakan rumus euclidean distance.

$$A = (x_1, y_1) \quad B = (x_2, y_2)$$

$$D(A, B) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Dalam R^n

$$A = (x_1, x_2, x_3, \dots, x_n) \quad B = (y_1, y_2, y_3, \dots, y_n)$$

Disederhanakan menjadi:

$$D(A, B) = \sqrt{(y_n - x_n)^2 + (y_{n-1} - x_{n-1})^2 + \dots + (y_1 - x_1)^2}$$

Masing Grade A, Grade B dan Grade C, Nilai RGB diambil dari data citra latih kemudian dicari nilai Mean untuk masing RGB, untuk menghitung jarak ke citra query menggunakan rumus euclidean distance, setelah dilakukan perhitungan jarak Euclidean distance[10] kemudian mengumpulkan kelas klasifikasinya, lebih jelas perhitungan dapat dilihat pada Tabel 3.

Tabel 3. Perhitungan Jarak Euclidean

Kelas (Grade)	i1 = Red	i2 = Green	i3 = Blue	Mean			Jarak ke citra query [181 149 132]
Grade A	207	159	127	190	164	138	$\sqrt{(190 - 181)^2 + (164 - 149)^2 + (138 - 132)^2}$ $\sqrt{9^2 + (15)^2 + (6)^2}$ $\sqrt{81 + 225 + 36}$ $=\sqrt{342} = 18,49$
	191	165	141				
	198	182	156				
	194	177	150				
	190	170	149				
	216	189	156				
	197	162	132				
	179	152	126				
	167	143	121				
	170	148	123				
Grade B	169	151	132	184	153	129	$\sqrt{(184 - 181)^2 + (153 - 149)^2 + (129 - 132)^2}$ $\sqrt{9 + 16 + 9}$ $=\sqrt{34} = 5,83$
	176	136	108				
	182	148	119				
	166	133	117				
	188	145	119				
	185	166	146				
	190	159	136				
	193	160	134				
	213	185	151				
	182	155	132				
Grade C	189	160	136	188	160	139	$\sqrt{(188 - 181)^2 + (160 - 149)^2 + (139 - 132)^2}$ $\sqrt{(7)^2 + (11)^2 + (7)^2}$
	191	165	141				
	170	141	122				
	182	155	136				
	189	161	138				
	195	169	148				
	212	187	169				
	195	173	149				
183	145	120					
180	148	131					

Tabel 4. Database Vektor Fitur Citra Uji

No	Warna RGB			Textsture							Kelas
	Red	Green	Blue	Contrast	correlation	Energy	Homo-geneity	Jarak	(%)		
1	197	165	141	0.34102	0.8844	0.44258	0.93037	0.330428	66.96	Grade A	
2	189	165	141	0.3477	0.88449	0.46412	0.93576	0.207606	79.24	Grade A	
3	201	172	150	0.19858	0.92978	0.41243	0.94963	0.536621	46.34	Grade A	
4	195	162	135	0.15248	0.95553	0.42583	0.96984	0.399954	60	Grade A	
5	197	171	146	0.19083	0.91921	0.50444	0.95074	0.436817	56.32	Grade A	
6	196	164	140	0.26337	0.92136	0.46682	0.96309	0.292944	70.71	Grade A	
7	205	171	139	0.21954	0.91745	0.43591	0.95715	0.542776	45.72	Grade A	
8	200	171	151	0.43385	0.84292	0.43686	0.92405	0.535985	46.4	Grade A	
9	201	172	150	0.19858	0.92978	0.41243	0.94963	0.536621	46.34	Grade A	
10	195	173	150	0.66621	0.81456	0.37999	0.91988	0.501212	49.88	Grade A	
11	169	141	125	0.3566	0.91679	0.46204	0.93431	0.588703	41.13	Grade B	
12	175	152	130	0.63044	0.81322	0.59353	0.93723	0.4381	56.19	Grade B	
13	170	143	120	0.51662	0.89499	0.3503	0.90987	0.577785	42.22	Grade B	
14	179	155	132	0.18739	0.94763	0.44662	0.96662	0.370053	62.99	Grade B	
15	183	143	125	0.5005	0.88628	0.41405	0.93033	0.442443	55.76	Grade B	
16	187	147	124	0.41284	0.90136	0.41521	0.93079	0.39475	60.53	Grade B	
17	179	146	118	0.36522	0.92268	0.35266	0.92219	0.495668	50.43	Grade B	
18	183	147	116	0.26978	0.93224	0.43549	0.94625	0.513396	48.66	Grade B	
19	177	144	119	0.45779	0.90731	0.39377	0.9438	0.512364	48.76	Grade B	
20	178	147	124	0.57499	0.873	0.41699	0.93217	0.414226	58.58	Grade B	
21	184	156	137	0.2825	0.91693	0.44927	0.94638	0.598038	40.2	Grade C	
22	187	165	151	0.59699	0.84998	0.24348	0.85432	0.529347	47.07	Grade C	
23	188	158	140	0.40422	0.89383	0.38088	0.9165	0.223277	77.67	Grade C	
24	183	161	138	0.38606	0.88281	0.53322	0.93886	0.506079	49.39	Grade C	
25	185	166	146	0.24684	0.92728	0.48688	0.94574	0.468781	53.12	Grade C	
26	190	159	136	0.24722	0.93015	0.4345	0.94791	0.356312	64.37	Grade C	
27	189	160	136	0.18885	0.92558	0.56785	0.96571	0.301436	69.86	Grade C	
28	189	161	138	0.29611	0.92038	0.46095	0.95042	0.102079	89.79	Grade C	
29	187	160	135	0.18875	0.92441	0.51232	0.94743	0.495049	50.5	Grade C	
30	190	160	137	0.24722	0.93015	0.4345	0.94791	0.219485	78.05	Grade C	

Tabel 4. merupakan Hasil dari citra uji dari 30 citra uji, dimana masing data 10 grade A, 10 grade B, 10 grade C dimana data citra uji itu masing – masing didapat dalam proses Klasifikasi berdasarkan perhitungan citra query dengan menggunakan rumus *euclidean distance*, kemudian dicari jarak terkecil dalam proses klasifikasi dan perhitungan persentase (%) kemiripan berdasarkan dari data citra latih. dengan metode *Nearest mean classifier*, nilai rata-rata terendah dari citra query yang nantinya akan menjadi hasil kualitas bibit kelapa.

Hasil Klasifikasi Metode NMC

```

...
# Split into train and test sets
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.3, random_state=42)
# Calculate the mean of each class
class_means = defaultdict(list)
for i, label in enumerate(y_train):
    class_means[label].append(X_train[i])
class_means = {label: np.mean(features, axis=0) for label, features in class_means.items()}
# Output the mean vectors for each class
print("Class Mean Vectors:")
for label, mean_vector in class_means.items():
    print(f"Class {label}: {mean_vector}")
# Predict function using nearest mean
def predict_nmc(X, class_means):
    predictions = []

```

```

for x in X:
    distances = {label: np.linalg.norm(x - mean) for label, mean in class_means.items()}
    predictions.append(min(distances, key=distances.get))
return predictions
# Make predictions
y_pred_nmc = predict_nmc(X_test, class_means)
# Calculate accuracy
nmc_accuracy = accuracy_score(y_test, y_pred_nmc)
print(f'NMC Accuracy: {nmc_accuracy:.2f}')
...

```

Tabel 5. Hasil Klasifikasi NMC

Class Mean Vectors:								
Class A:	[0.41749276	0.19907841	0.09747337	-0.01405737	-0.02149558	0.63138105	0.07987223]
Class B:	[-0.6834864	-0.63705091	-0.76860149	1.04013231	0.75982222	-1.17284507	0.35815634]
Class C:	[0.11192983	-0.01592627	0.73121007	-0.22036174	-0.70761848	-0.54257265	-0.19561821]
NMC Accuracy: 0.56								

Tabel 5. menunjukkan hasil dari analisis klasifikasi menggunakan metode Nearest Mean Classifier (NMC). Berikut penjelasan elemen-elemennya:

a) Class Mean Vectors:

Ini adalah vektor rata-rata untuk masing-masing kelas dalam dataset. Vektor ini mewakili titik tengah dari fitur-fitur yang ada untuk setiap kelas.

- 1) Class A: Vektornya adalah [0.41749276, 0.19907841, 0.09747337, -0.01405737, -0.02149558, 0.63138105, 0.07987223].
- 2) Class B: Vektornya adalah [-0.6834864, -0.63705091, -0.76860149, 1.04013231, 0.75982222, -1.17284507, 0.35815634].
- 3) Class C: Vektornya adalah [0.11192983, -0.01592627, 0.73121007, -0.22036174, -0.70761848, -0.54257265, -0.19561821].

b) NMC Accuracy:

Akurasi dari klasifikasi yang diperoleh menggunakan metode Nearest Mean Classifier adalah `0.56`, atau 56%. Ini menunjukkan seberapa sering model klasifikasi ini memberikan prediksi yang benar untuk dataset yang diuji.

Metode Nearest Mean Classifier bekerja dengan mengklasifikasikan titik data baru berdasarkan kelas dengan rata-rata terdekat. Jadi, setiap titik data baru diukur jaraknya ke vektor rata-rata dari setiap kelas, dan data tersebut diberi label sesuai dengan kelas yang memiliki vektor rata-rata terdekat. Akurasi 56% berarti lebih dari setengah data yang diuji diklasifikasikan dengan benar, namun masih ada ruang untuk peningkatan.

Hasil klasifikasi metode K-nn

```

...
# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y_encoded, test_size=0.3,
random_state=42)
# Create and train the K-NN classifier
knn = KNeighborsClassifier(n_neighbors=3)
knn.fit(X_train, y_train)
# Predict the labels for the test set
y_pred = knn.predict(X_test)
# Evaluate the classifier
accuracy = accuracy_score(y_test, y_pred)

```

```

classification_rep = classification_report(y_test, y_pred,
target_names=label_encoder.classes_)
print(f'Accuracy: {accuracy}')
print('Classification Report:')
print(classification_rep)
...

```

Tabel 6. Hasil Klasifikasi K-Nearest Neighbor

Accuracy: 0.2222222222222222				
Classification Report:				
	Precision	recall	f1-score	support
A	0.14	0.50	0.22	2
B	0.00	0.00	0.00	3
C	1.00	0.25	0.40	4
<hr/>				
Accuracy		0.22		9
macro avg	0.38	0.25	0.21	9
weighted avg	0.48	0.22	0.23	9

Tabel 6. menunjukkan laporan klasifikasi dari model machine learning untuk sebuah masalah klasifikasi. Berikut penjelasan elemen-elemennya :

a) Akurasi

- 1) Accuracy : 0.2222222222222222
- 2) Ini berarti bahwa model ini hanya benar dalam 22.22% dari kasus.

b) Classification Report

Laporan klasifikasi memberikan rincian tentang kinerja model untuk masing-masing kelas dalam hal presisi, recall, dan f1-score.

Kelas A

- 1) Precision: 0.14
- 2) Dari semua prediksi yang mengatakan bahwa kelas tersebut adalah A, hanya 14% yang benar-benar A.
- 3) Recall: 0.50
- 4) Dari semua data yang benar-benar kelas A, model ini berhasil mendeteksi 50% di antaranya.
- 5) F1-Score: 0.22
- 6) Merupakan rata-rata harmonis dari precision dan recall, memberikan gambaran keseluruhan tentang kinerja model untuk kelas A.
- 7) Support: 2
- 8) Jumlah sebenarnya dari kasus yang termasuk dalam kelas A.

Kelas B

- 1) Precision: 0.00
- 2) Model ini tidak berhasil memprediksi kelas B dengan benar.
- 3) Recall: 0.00
- 4) Model ini tidak berhasil mendeteksi satupun dari kelas B yang sebenarnya.
- 5) F1-Score: 0.00
- 6) Kombinasi dari precision dan recall menunjukkan performa yang sangat buruk untuk kelas B.
- 7) Support: 3
- 8) Jumlah sebenarnya dari kasus yang termasuk dalam kelas B.

Kelas C

- 1) Precision: 1.00
- 2) Dari semua prediksi yang mengatakan bahwa kelas tersebut adalah C, 100% benar-benar C.
- 3) Recall: 0.25

- 4) Dari semua data yang benar-benar kelas C, model ini hanya berhasil mendeteksi 25% di antaranya.
 - 5) F1-Score: 0.40
 - 6) Kombinasi dari precision dan recall menunjukkan bahwa meskipun prediksi benar, model sering kali gagal mendeteksi kelas C.
 - 7) Support: 4
 - 8) Jumlah sebenarnya dari kasus yang termasuk dalam kelas C.
- c) Rata-rata
- Macro avg:
- 1) Precision: 0.38
 - 2) Recall: 0.25
 - 3) F1-Score: 0.21
- Ini adalah rata-rata dari semua metrik di atas, memberikan gambaran keseluruhan tanpa mempertimbangkan jumlah sampel di masing-masing kelas.
- Weighted avg:
- 1) Precision: 0.48
 - 2) Recall: 0.22
 - 3) F1-Score: 0.23

Ini adalah rata-rata tertimbang dari semua metrik di atas, memberikan gambaran keseluruhan dengan mempertimbangkan jumlah sampel di masing-masing kelas.

Secara keseluruhan, laporan ini menunjukkan bahwa model memiliki kinerja yang sangat rendah, terutama untuk kelas B yang tidak terdeteksi sama sekali dan kinerja model untuk kelas C cukup baik dalam hal precision tetapi sangat rendah dalam recall. Hal ini menunjukkan bahwa model ini membutuhkan perbaikan signifikan untuk menjadi dapat diandalkan.

Berdasarkan hasil yang diberikan, metode Nearest Mean Classifier (NMC) dan K-Nearest Neighbor (K-NN) memiliki kinerja yang berbeda dalam klasifikasi dataset. Berikut adalah ringkasan dan kesimpulan mengenai kinerja masing-masing metode:

Nearest Mean Classifier (NMC)

- a) Akurasi: 56%
- b) Kinerja:
- c) Kinerja keseluruhan NMC cukup baik dengan akurasi lebih dari setengah data yang diuji diklasifikasikan dengan benar.
- d) Metode ini menghitung vektor rata-rata untuk masing-masing kelas dan menggunakan vektor ini untuk klasifikasi, yang sering kali sederhana namun efektif untuk dataset yang terdistribusi dengan baik.

K-Nearest Neighbor (K-NN)

- a) Akurasi: 22.22%
- b) Kinerja:
- c) Kinerja keseluruhan K-NN sangat rendah dengan hanya 22.22% akurasi.
- d) Kelas B tidak terdeteksi sama sekali (precision dan recall adalah 0).
- e) Precision tinggi untuk kelas C (100%), tetapi recall rendah (25%), menunjukkan ketidakseimbangan dalam deteksi kelas.
- f) Kinerja rata-rata untuk precision, recall, dan f1-score sangat rendah.

5 KESIMPULAN

Berdasarkan hasil di atas, metode Nearest Mean Classifier (NMC) memiliki kinerja yang lebih baik dibandingkan dengan K-Nearest Neighbor (K-NN) dalam konteks dataset ini. NMC memiliki akurasi yang lebih tinggi (56%) dibandingkan dengan K-NN (22.22%) dan menunjukkan kinerja yang lebih konsisten di berbagai metrik evaluasi.

Rekomendasi:

- a) Pilih NMC: Mengingat kinerja keseluruhan yang lebih baik, NMC adalah pilihan yang lebih baik untuk digunakan dalam klasifikasi dataset ini.
- b) Peningkatan Model: Meskipun NMC lebih baik dalam konteks ini, akurasi 56% masih menunjukkan ruang yang signifikan untuk perbaikan. Pertimbangkan untuk:
- c) Mengoptimalkan parameter model.
- d) Menggunakan teknik praproses data untuk menyeimbangkan kelas.
- e) Mengeksplorasi metode klasifikasi lain yang mungkin lebih cocok untuk dataset ini.
- f) Dengan demikian, untuk dataset ini, NMC adalah metode yang lebih disarankan dibandingkan K-NN.

REFERENSI

- [1] M. R. Ridha and F. Yunita, "Pemilihan Bibit Kelapa Menggunakan Metode Nearest Mean Classifier Untuk Masyarakat Petani Kelapa Di Kabupaten Indragiri Hilir," *J. Perangkat Lunak*, vol. 2, no. 3, pp. 101–114, 2020, doi: 10.32520/jupel.v2i3.1306.
- [2] F. Yunita, "Diabetes Mellitus Menggunakan Metode K-Nearest Neighbor (K-Nn)," *Bappeda*, vol. 2, pp. 223–230, 2016.
- [3] K. Kartarina, N. K. Sriwinarti, and N. luh P. Juniarti, "Analisis Metode K-Nearest Neighbors (K-NN) Dan Naive Bayes Dalam Memprediksi Kelulusan Mahasiswa," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 3, no. 2, pp. 107–113, 2021, doi: 10.35746/jtim.v3i2.159.
- [4] F. Fitriyana and A. Sucipto, "Sistem Informasi Penjualan Oleh Sales Marketing Pada Pt Erlangga Mahameru," *J. Teknol. dan Sist. Inf.*, vol. 1, no. 1, pp. 105–110, 2020, doi: 10.33365/jtsi.v1i1.239.
- [5] S. Ratna, "Pengolahan Citra Digital Dan Histogram Dengan Phyton Dan Text Editor Phycharm," *Technol. J. Ilm.*, vol. 11, no. 3, p. 181, 2020, doi: 10.31602/tji.v11i3.3294.
- [6] W. E. Sumantri, Y. A. Nasution, Suyono, and R. Rosnelly, "Pengenalan Pola Aksara Batak menggunakan Backpropagation Recognition of Batak Script Patterns using Backpropagation," *J. SISFOTENIKA*, vol. 14, no. 1, pp. 57–67, 2024, [Online]. Available: <https://stmikpontianak.org/ojs/index.php/sisfotenika>
- [7] J. Andilan, D. S.M.Engka, and J. I.Sumual, "Pengaruh Biaya Produksi, Luas Lahan, Harga Jual Terhadap Pendapatan Petani Kelapa (Kopra) Di Kecamatan talawaan," *J. Berk. Ilm. Efisiensi*, vol. 21, no. 06, pp. 102–111, 2021.
- [8] O. Irnawati and I. Darwati, "Penerapan Model Waterfall Dalam Analisis Perancangan Sistem Informasi Inventarisasi Berbasis Web," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 6, no. 2, pp. 109–116, 2020, doi: 10.33330/jurteks.v6i2.406.
- [9] R. Alya Shafira, Yahfizham, and A. Muliani Harahap, "Menentukan Jarak Terpendek Dalam Pengiriman Barang Dengan Perbandingan Euclidean Distance Dan Manhattan Distance," *J. Sci. Soc. Res.*, vol. VI, no. 3, pp. 678–685, 2023.
- [10] M. S. Pangestu and M. A. Fitriani, "Perbandingan Perhitungan Jarak Euclidean Distance, Manhattan Distance, dan Cosine Similarity dalam Pengelompokan Data Bibit Padi Menggunakan Algoritma K-Means," *Sainteks*, vol. 19, no. 2, p. 141, 2022, doi: 10.30595/sainteks.v19i2.14495.