

Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Penentuan Bibit Unggul

Bambang Sugito*, Muhammad Iqbal

Program studi Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Medan, Indonesia

Email: ¹bambang.sugito@pertamina.com, ²muhammadiqbal@dosenpancabudi.ac.id

Email Penulis Korespondensi: bambang.sugito@pertamina.com

Submitted: 05/11/2024; Accepted: 20/12/2024; Published: 21/12/2024

Abstrak—Kakao merupakan salah satu tanaman komoditas unggulan dari sub sektor perkebunan yang berkembang, sebagian besar diusahakan oleh petani dalam bentuk perkebunan rakyat. Budidaya yang dilakukan agar kiranya tetap dapat memproduksi bibit dan pohon secara mandiri untuk keberlangsungan komoditi. Proses yang dilakukan pada saat ini pada pemilihan bibit unggul dilakukan oleh staff yang bertugas khusus pada bagian tersebut. Seharusnya proses pemilihan bibit unggul tanaman kakao sudah harus terdata dan tersimpan. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengestraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Berdasarkan masalah tersebut, maka di terapkan algoritma C5.0 dan K-Nearest Neighbor untuk dilakukan proses perbandingan dalam memprediksi kebutuhan bibit unggul yang akan di gunakan. Akan bisa memberikan informasi yang akurat dan bisa dijadikan bahan pertimbangan dalam stok kebutuhan bibit unggul kakako. Proses dengan menggunakan algoritma C5.0 dan K-Nearest Neighbor telah berhasil dilakukan, dari proses pengujian yang dilakukan bahwasannya algoritma C5.0 memiliki hasil kinerja lebih baik sebesar 87,50% dibandingkan dengan algoritma K-Nearest Neighbor sebesar 62,50%.

Kata Kunci: Algoritma; Bibit; Unggul; C5.0; Naive Bayes

Abstract—Cocoa is one of the leading commodity crops from the growing plantation sub-sector, mostly cultivated by farmers in the form of community plantations. Cultivation is carried out so that it can continue to produce seeds and trees independently for the sustainability of the commodity. The process currently carried out in the selection of superior seeds is carried out by staff who are specifically assigned to that section. The process of selecting superior cocoa plant seeds should have been recorded and stored. Data mining is a process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from various large databases. Based on this problem, the C5.0 and K-Nearest Neighbor algorithms are applied to carry out the comparison process in predicting the need for superior seeds to be used. It will be able to provide accurate information and can be used as a consideration in the stock of superior cocoa seed needs. The process using the C5.0 and K-Nearest Neighbor algorithms has been successfully carried out, from the testing process carried out that the C5.0 algorithm has a better performance result of 87.50% compared to the K-Nearest Neighbor algorithm of 62.50%.

Keywords: Algorithm; Seed; Superior; C5.0; Naive Bayes

1. PENDAHULUAN

Kakao merupakan salah satu tanaman komoditas unggulan dari sub sektor perkebunan yang berkembang, sebagian besar diusahakan oleh petani dalam bentuk perkebunan rakyat. Komoditi ini mampu mengangkut perekonomian penduduk dan menjadi salah satu tanaman yang diunggulkan dan juga Kakao merupakan tumbuhan tahunan berbentuk pohon di alam dapat mencapai ketinggian 10 m meskipun demikian, dalam pembudidayaan tingginya dibuat tidak lebih dari 5 m tetapi dengan tajuk menyamping yang meluas. Hal ini dilakukan untuk memperbanyak cabang produktif sehingga memudahkan petani untuk memanennya[1] [2].

Budidaya yang dilakukan agar kiranya tetap dapat memproduksi bibit dan pohon secara mandiri untuk keberlangsungan komoditi. Tanaman yang baik berlandaskan cikal bakal bibit yang baik pula. Oleh sebab itu tidak seluruh bibit digunakan untuk ditanam pada kakao tersebut, hanya bibit unggul saja yang digunakan untuk proses perkembangan pohon hingga mendapatkan hasil yang baik. Pemilihan bibit unggul pada tanaman kakao merupakan sebuah proses yang harus dilakukan secara detail dan teliti. Kesalahan pada pemilihan bibit unggul maka akan berdampak pada hasil yang didapatkan tidak baik sepenuhnya[3] [4].

Proses yang dilakukan pada saat ini pada pemilihan bibit unggul dilakukan oleh staff yang bertugas khusus pada bagian tersebut. Namun pada hasil pemilihan bibit unggul tanaman kakao kadang kala didapatkan hasil yang tidak sesuai atau kekeliruan dalam pemilihan bibit unggul tersebut. Hal tersebut dikarenakan human error dan juga banyak faktor lainnya.

Seharusnya proses pemilihan bibit unggul tanaman kakao sudah harus terdata dan tersimpan. Kemudian dari data yang digunakan pada proses pemilihan bibit kakao tersebut dapat diolah kembali dengan berbasis komputer sehingga mendapatkan hasil yang lebih baik dan lebih akurat. Dimana proses pengolahan data berbasis komputer tersebut adalah data mining.

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengestraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar[5] [6]. Dengan adanya data mining maka akan didapatkan suatu permata berupa

pengetahuan di dalam kumpulan data-data yang banyak jumlahnya. Pada data mining terdapat banyak teknik atau cara yang dapat dilakukan untuk prosesnya. Seperti teknik klasifikasi dengan algoritma C5.0 dan algoritma K-Nearest Neighbor[7] [10].

Algoritma C5.0 ini, merupakan salah satu teknik untuk penyelesaian masalah dalam memilih bibit terbaik untuk petani, masalah yang terjadi terkadang mengalami kendala di pemilihan bibit unggul, karna tidak mengetahui spesifikasi bibit terbaik untuk di tanam. Berdasarkan masalah tersebut, maka di terapkan algoritma C5.0 dan K-Nearest Neighbor untuk dilakukan proses perbandingan dalam memprediksi kebutuhan bibit unggul yang akan di gunakan. Akan bisa memberikan informasi yang akurat dan bisa dijadikan bahan pertimbangan dalam stok kebutuhan bibit unggul kakako. Agar tidak terjadi kesalahan pemilihan bibit unggul dan bisa memberikan informasi kepada para petani. Penelitian menggunakan penerapan dengan algoritma nearest neighbor untuk analisis pemberian bibit unggul ini diharapkan dapat memberikan rekomendasi dalam melakukan analisis bibit yang akan digunakan . Sehingga pemilihan bibit unggul dapat dilakukan dengan cepat, tepat, dan akurat berdasarkan kesamaan dari data kasus kesalahan dalam penanaman bibit kakao.

Pada penelitian sebelumnya, Vinna Rahmayanti Setyaning Nastiti, dkk “Penerapan Algoritma C5.0 Pada Analisis Faktor-Faktor Pengaruh Kelulusan Tepat Waktu Mahasiswa Teknik Informatika Universitas Muhammadiyah Malang” menyimpulkan Algoritma C5.0 mampu melakukan seleksi fitur untuk menentukan pengaruh kelulusan tepat waktu mahasiswa Teknik Informatika UMM dengan menghasilkan 8 fitur dari total keseluruhan 15 fitur dengan nilai akurasi sebesar 91,9% [11].

Tedi Permana, dkk “Perbandingan Hasil Prediksi Kredit Macet Pada Koperasi Menggunakan Algoritma KNN dan C5.0” menyimpulkan berdasarkan perhitungan menggunakan algoritma KNN dan C5.0 dengan datasets sebanyak 30 data transaksi nilai akurasi yang diperoleh dari algoritam C5.0 lebih baik dari KNN yaitu 86,67%. Sedangkan untuk algoritma KNN memiliki nilai akurasi sebesar 83,33%, sehingga algoritma, C5.0 lebih cocok digunakan sebagai patokan dalam memprediksi kredit macet di koperasi [12].

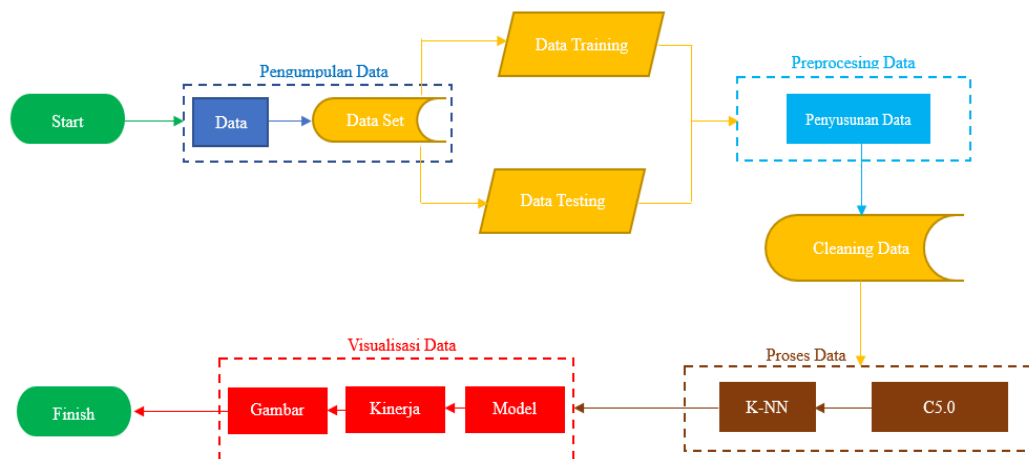
Fatih Nur Umma, dkk “Klasifikasi Status Kemiskinan Rumah Tangga Dengan Algoritma C5.0 di Kabupaten Pematang” menyimpulkan Algoritma C5.0 menghasilkan nilai rata akurasi sebesar 91,16% dalam mengklasifikasi status kemiskinan rumah tangga di Kabupaten Pematang [13].

Inna Alvi Nikmatun dan Indra Waspada “ Implementasi Data Mining Untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor” menyimpulkan Algoritma K-Nearest Neighbor dengan melakukan skenario percobaan yang dilakukan diperoleh nilai akurasi tertinggi pada skenario yang menggunakan atribut mata kuliah pilihan yaitu 75,95%, sehingga dapat disimpulkan bahwa mata kuliah pilihan sangat berpengaruh pada masa studi mahasiswa [14].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian merupakan alur proses yang akan dilalui pada penelitian. Pada tahapan penelitian merupakan model penyelesaian penelitian. Adapun tahapan penelitian dapat dilihat berikut:



Gambar 1. Tahapan Penelitian

2.2 Data Mining

Data Mining ditujukan untuk mengekstrak pengetahuan dari kumpulan data sehingga didapatkan struktur yang dapat dimengerti manusia. Data Mining atau penambangan data adalah teknik yang relatif cepat dan mudah untuk menemukan pengetahuan, pola dan / atau relasi antar data, secara otomatis. Data Mining adalah langkah analisis terhadap proses penemuan pengetahuan di dalam basis data atau knowledge discovery in databases yang disingkat

KDD[15], [16]. Pengetahuan bisa berupa pola data atau relasi antar data yang valid. Data Mining merupakan gabungan sejumlah disiplin ilmu komputer yang didefinisikan sebagai proses penemuan pola-pola baru dari kumpulan-kumpulan data yang sangat besar, meliputi metode-metode yang merupakan irisan dari artificial intelligence, machine learning, statistic, dan database system. Data Mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini diketahui secara manual dari suatu basis data[17]–[19].

2.3 Algoritma C5.0

Algoritma C5.0 merupakan penyempurnaan dari algoritma ID3 dan C4.5. Dalam proses pembentukan pohon keputusan nilai informasi gain tertinggi akan terpilih sebagai root bagi node selanjutnya. Algoritma ini dimulai dengan semua data yang dijadikan akar dari pohon keputusan sedangkan atribut yang dipilih akan menjadi pembagi bagi sampel tersebut. Nilai entropy didefinisikan sebagai suatu parameter untuk mengukur heterogenitas (keberagaman) dalam suatu himpunan data. Semakin heterogen suatu himpunan data, semakin besar pula nilai entropynya. Secara matematis, entropy dirumuskan sebagai berikut dan dapat dilihat pada persamaan dibawah[20]–[22]:

$$\text{Entropy}(S) = \sum_i^c - p_i \log_2 p_i \tag{1}$$

Dimana:

S = Himpunan data

c = jumlah nilai / kelas

p_i = Rasio antara sampel di kelas i dengan jumlah sampel di himpunan data

Gain didefinisikan sebagai ukuran efektivitas suatu atribut dalam mengklasifikasi data. Secara matematis gain dari suatu atribut A, dapat dilihat pada persamaan dibawah[20]–[22].

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{v \in \text{Values } A} \frac{|S_v|}{|S|} * \text{Entropy}(S_v) \tag{2}$$

Dimana:

A = atribut

V = menyatakan suatu nilai yang mungkin untuk atribut A

Values(A) = himpunan nilai-nilai yang mungkin untuk atribut A

|S_v| = jumlah sampel untuk nilai v

|S| = jumlah seluruh sampel data

Entropy(S_v) = entropy untuk sampel-sampel yang memiliki nilai v

2.4 Algoritma K-Nearest Neighbor

K-Nearest Neighbor salah satu teknik klarifikasi data yang kuat, dengan cara mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama berdasarkan pencocokan bobot[4]. K-Nearest Neighbor termasuk dalam kelompok instance-based-learning. K-Nearest Neighbor merupakan contoh algoritma berbasis pembelajaran, dimana data set pelatihan (training) disimpan, sehingga klasifikasi untuk record baru yang tidak diklasifikasi didapatkan dengan membandingkan record yang paling mirip dengan training set. Untuk mendefinisikan jarak antara dua titik yaitu titik pada training (x) dan titik pada data testing (y) maka digunakan rumus Euclidean, seperti pada rumus yang dijabarkan[23]–[25].

$$d(x,y) = \sqrt{\sum_{k=1}^n (X_{\text{training}} - Y_{\text{testing}})^2} \tag{3}$$

Dimana :

X_{training} = data training ke-i

Y_{testing} = data testing

i = record ke-i dari tabel

n = jumlah data training

Langkah-langkah dalam menyelesaikan algoritma K-Nearest Neighbor adalah sebagai berikut[5]:

1. Menentukan parameter K.
2. Menghitung kuadrat jarak euclid (queri instance) masing-masing objek terhadap data sampel yang ada.
3. Mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak eucliden terkecil.
4. Mengumpulkan kategori Y (Klasifikasi Nearest Neighbor).
5. Dengan menggunakan kategori nearest neighbor yang paling mayoritas maka dapat diprediksi nilai queri instance yang telah dihitung.

3. HASIL DAN PEMBAHASAN

Analisa merupakan proses menguraikan suatu permasalahan menjadi beberapa elemen terkecil, guna mempermudah dalam menyelesaikan permasalahan yang ada. Tahap analisa memerlukan suatu pendekatan analisa guna



menghindari kesalahan yang mungkin saja muncul pada tahap berikutnya. Tahap analisa ini merupakan tahapan yang sangat penting, pendekatan yang dilakukan adalah mendefinisikan masalah pada sistem yang sedang berjalan dan sekaligus melakukan evaluasi setiap cara kerja sistem yang sedang berjalan berdasarkan.

Permasalahan yang terjadi dalam proses penentuan atau pemilihan terhadap bibit unggul tersebut masih dilakukan oleh staff khusus, dimana dalam proses yang dilakukan masih terdapat human eror ataupun kesalahan dalam penentuan bibit unggul tersebut. Sudah seharusnya proses penyelesaian dilakukan dengan memudahkan proses penentuan dengan melakukan proses penentuan pola yang dilakukan dalam bibit unggul tersebut.

Dalam memilih bibit kakao yang bermutu dan berkualitas pada tentunya membutuhkan sebuah kebijakan untuk kelangsungan pemilihan bibit kakao yang unggul. Masalah yang terjadi pada terkadang mengalami kendala di pemilihan bibit unggul, karna tidak mengetahui spesifikasi bibit terbaik untuk ditanam. Oleh karena itu dilakukan sebuah perbandingan algoritma untuk mengetahui tingkat presentasi terhadap kasus yang sama yakni pemilihan bibit unggul kakao, dengan menggunakan algoritma C5.0 dan Nearest Neighbor berdasarkan data yang ada. Adapun data yang digunakan pada penelitian dapat dilihat berikut:

Tabel 1. Sampel Data Penelitian

Objek	Perkembang Biakan Generatif	Tinggi Bibit (Cm)	Jumlah Daun	Diameter Batang (Cm)	Hasil
Objek 1	Ya	23	8	1	Unggul
Objek 2	Ya	35	12	0,4	Tidak
Objek 3	Ya	43	11	0,8	Unggul
Objek 4	Tidak	50	15	0,7	Unggul
Objek 5	Ya	55	16	0,7	Unggul
Objek 6	Tidak	20	9	0,5	Tidak
Objek 7	Tidak	32	10	0,5	Tidak
Objek 8	Ya	40	11	0,6	Unggul

3.1 Penerapan Algoritma C5.0

Tahapan pertama yang dilakukan pada penelitian adalah proses klasifikasi dengan menggunakan algoritma C5.0. Sebelum dilakukan proses, terlebih dahulu dilakukan preprocessing terhadap data. Tujuan dilakukan preprocessing guna untuk penyesuaian dan penyesuaian terhadap data yang akan digunakan pada penggunaan algoritma. Adapun proses preprocessing dapat dilihat berikut:

Tabel 2. Preprocessing Tinggi Bibit

No	Data Awal	Hasil Preprocessing
1	10 - 30	Pendek
2	31 - 45	Sedang
3	>45	Tinggi

Pada Tabel 2 dapat terlihat hasil dari proses preprocessing data tinggi bibit, kemudian dilakukan proses preprocessing data terhadap data lainnya seperti tabel berikut:

Tabel 3. Preprocessing Jumlah Daun

No	Data Awal	Hasil Preprocessing
1	5 - 10	Sedikit
2	11 - 15	Sedang
3	>15	Banyak

Pada Tabel 3 dapat terlihat hasil dari proses preprocessing data jumlah daun, kemudian dilakukan proses preprocessing data terhadap data lainnya seperti tabel berikut:

Tabel 4. Preprocessing Diameter Batang

No	Data Awal	Hasil Preprocessing
1	0,0 - 0,3	Kecil
2	0,4 - 0,6	Sedang
3	0,7 - 1	Besar

Setelah diketahui terhadap proses preprocessing data, maka selanjutnya dilakukan proses perubahan ataupun penyesuaian data. Adapun sampe data hasil preprocessing data dapat dilihat berikut:

Tabel 5. Data Hasil Preprocessing

Objek	Perkembang Biakan Generatif	Tinggi Bibit	Jumlah Daun	Diameter Batang	Hasil
-------	-----------------------------	--------------	-------------	-----------------	-------



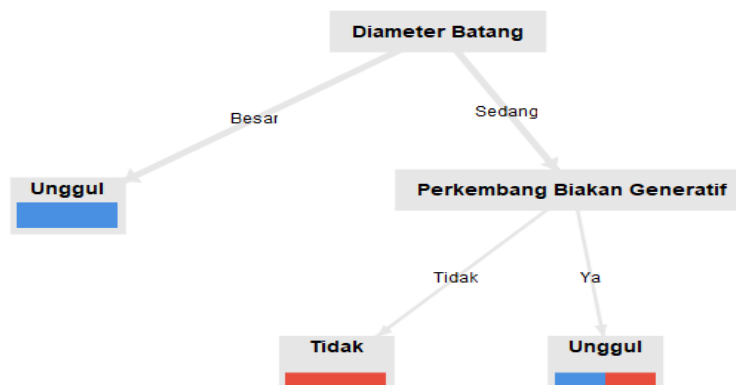
Objek	Perkembang Biakan Generatif	Tinggi Bibit	Jumlah Daun	Diameter Batang	Hasil
Objek 1	Ya	Pendek	Sedikit	Besar	Unggul
Objek 2	Ya	Sedang	Sedang	Sedang	Tidak
Objek 3	Ya	Sedang	Sedang	Besar	Unggul
Objek 4	Tidak	Tinggi	Sedang	Besar	Unggul
Objek 5	Ya	Tinggi	Banyak	Besar	Unggul
Objek 6	Tidak	Pendek	Sedikit	Sedang	Tidak
Objek 7	Tidak	Sedang	Sedikit	Sedang	Tidak
Objek 8	Ya	Sedang	Sedang	Sedang	Unggul

Setelah didapatkan hasil data preprocessing, maka langkah selanjutnya dapat dilakukan terhadap proses penerapan algoritma C5.0 terhadap penyelesaian proses klasifikasi. Dimana proses pertama sekali dilakukan dengan mencari nilai akar atau node 1. Adapun hasil perhitungan nilai entropy dan gain pada node 1 dapat dilihat berikut:

Tabel 6. Hasil Perhitungan Node I

Nod e	Atribut	Nilai	Jumlah Kasus / Data Set	Unggu l	Tida k	Entrop y	Gain
1			8	5	3	0,95443	
	Perkembang Biakan Generatif						0,00323
		Ya	5	3	2	0,97095	
		Tidak	3	1	2	0,9183	
	Tinggi Bibit						0,20443
		Pendek	2	1	1	1	
		Sedang	4	2	2	1	
		Tinggi	2	2	0	0	
	Jumlah Daun						0,20443
		Sedikit	3	1	2	0,9183	
		Sedang	4	3	1	0,81128	
		Banya k	1	1	0	0	
	Diameter Batang						0,45443
		Sedang	4	2	2	1	
		Besar	4	4	0	0	

Dari proses yang dilakukan pada tabel 6, maka dapat dilihat bahwasannya atribut Diameter Batang terpilih sebagai akar pada porses. Hal tersebut dikarenakan atribut Diameter Batang merupakan atribut dengan nilai Gain tertinggi. Dari proses yang dilakukan masih terdapat percabangan. Maka proses dilanjutkan dengan menggunakan cara yang sama, atapun hasil akhir pohon keputusan dapat dilihat berikut:



Gambar 2. Hasil Pohon Keputusan

3.2 Penerapan Algoritma K-NN

Data yang dianalisa adalah data pemilihan bibit unggul kakao, data dikonversikan terlebih dahulu ke dalam bentuk angka. Atribut data yang dikonversikan ke dalam angka yaitu atribut perkembangbiakan generatif.

Tabel 7. Konversi Data

Perkembang Biakan Generatif	Keterangan
Ya	1
Tidak	0

Berikut adalah contoh perhitungan pada pembahasan algoritma Nearsset Neighbor dengan menggunakan data sampel.

Tabel 8. Data Training

Objek	Perkembang Biakan Generatif	Tinggi Bibit (Cm)	Jumlah Daun	Diameter Batang (Cm)	Hasil
Objek 1	1	23	8	1	Unggul
Objek 2	1	35	12	0,4	Tidak
Objek 3	1	43	11	0,8	Unggul
Objek 4	0	50	15	0,7	Unggul
Objek 5	1	55	16	0,7	Unggul
Objek 6	0	20	9	0,5	Tidak
Objek 7	0	32	10	0,5	Tidak
Objek 8	1	40	11	0,6	Unggul

Tabel 9. Data Testing

Objek	Perkembang Biakan Generatif	Tinggi Bibit (Cm)	Jumlah Daun	Diameter Batang (Cm)	Hasil
Objek	1	55	10	0,5	?

a. Menentukan parameter k .

$$k = 5$$

b. Menghitung jarak antara data uji dan data training dengan euclidean distance.

$$d = \sqrt{(1 - 1)^2 + (23 - 55)^2 + (8 - 10)^2 + (1 - 0,5)^2}$$

$$= \sqrt{0 + 1.024 + 4 + 0,25}$$

$$= 2,296$$

$$d = \sqrt{(1 - 1)^2 + (35 - 55)^2 + (12 - 10)^2 + (0,4 - 0,5)^2}$$

$$= \sqrt{0 + 400 + 4 + 0,01}$$

$$= 20,1$$

$$d = \sqrt{(1 - 1)^2 + (43 - 55)^2 + (11 - 10)^2 + (0,8 - 0,5)^2}$$

$$= \sqrt{0 + 144 + 1 + 0,09}$$

$$= 12,045$$

$$d = \sqrt{(0 - 1)^2 + (50 - 55)^2 + (15 - 10)^2 + (0,7 - 0,5)^2}$$

$$= \sqrt{1 + 25 + 25 + 0,04}$$

$$= 7,144$$

$$d = \sqrt{(1 - 1)^2 + (55 - 55)^2 + (16 - 10)^2 + (0,7 - 0,5)^2}$$

$$= \sqrt{0 + 0 + 36 + 0,04}$$

$$= 6,003$$

$$d = \sqrt{(0 - 1)^2 + (20 - 55)^2 + (9 - 10)^2 + (0,5 - 0,5)^2}$$

$$= \sqrt{1 + 1.225 + 1 + 0}$$

$$= 1,795$$

$$d = \sqrt{(0 - 1)^2 + (32 - 55)^2 + (10 - 10)^2 + (0,5 - 0,5)^2}$$

$$= \sqrt{1 + 529 + 0 + 0}$$

$$= 23,021$$

$$d = \sqrt{(1 - 1)^2 + (40 - 55)^2 + (11 - 10)^2 + (0,6 - 0,5)^2}$$

$$= \sqrt{0 + 225 + 1 + 0,01}$$

$$= 15,033$$

Tabel 10. Hasil Perhitungan Jarak

Objek	Perkembang Biakan Generatif	Tinggi Bibit (Cm)	Jumlah Daun	Diameter Batang (Cm)	Euclidean Distance	Hasil
Objek 1	1	23	8	1	2,296	Unggul
Objek 2	1	35	12	0,4	20,1	Tidak
Objek 3	1	43	11	0,8	12,045	Unggul
Objek 4	0	50	15	0,7	7,144	Unggul
Objek 5	1	55	16	0,7	6,003	Unggul
Objek 6	0	20	9	0,5	1,795	Tidak
Objek 7	0	32	10	0,5	23,021	Tidak
Objek 8	1	40	11	0,6	15,033	Unggul

c. Mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak euclidean terkecil.

Tabel 11. Hasil Penyusunan Data Jarak Terdekat

Objek	Perkembang Biakan Generatif	Tinggi Bibit (Cm)	Jumlah Daun	Diameter Batang (Cm)	Euclidean Distance	Hasil	Peringkat Jarak
Objek 6	1	20	9	0,5	1,795	Tidak	1
Objek 1	0	23	8	1	2,296	Unggul	2
Objek 5	1	55	16	0,7	6,003	Unggul	3
Objek 4	1	50	15	0,7	7,144	Unggul	4
Objek 3	0	43	11	0,8	12,045	Unggul	5
Objek 8	1	40	11	0,6	15,033	Unggul	6
Objek 2	1	35	12	0,4	20,1	Tidak	7
Objek 7	1	32	10	0,5	23,021	Tidak	8

d. Mengumpulkan kategori Y (Klasifikasi Nearest Neighbor).

Tabel 12. Tetangga Terdekat k = 5

Objek	Perkembang Biakan Generatif	Tinggi Bibit (Cm)	Jumlah Daun	Diameter Batang (Cm)	Euclidean Distance	Hasil
Objek 6	1	20	9	0,5	1,795	Tidak
Objek 1	0	23	8	1	2,296	Unggul
Objek 5	1	55	16	0,7	6,003	Unggul
Objek 4	1	50	15	0,7	7,144	Unggul
Objek 3	0	43	11	0,8	12,045	Unggul

e. Berdasarkan pada tabel jarak k = 5, label yang banyak muncul adalah unggul, maka hasil pemilihan bibit kakao adalah **“Unggul”**

3.3 Pembahasan

Setelah dilakukan proses pengujian ataupun penerapan terhadap algoritma c5.0 dan K-Nearest Neighbor, maka selanjut dilakukan proses perhitungan kinerja ataupun akurasi terhadap hasil. Adapun hasil kinerja akurasi dapat dilihat berikut:

accuracy: 87.50%

	true Unggul	true Tidak	class precision
pred. Unggul	5	1	83.33%
pred. Tidak	0	2	100.00%
class recall	100.00%	66.67%	

Gambar 3. Hasil Kinerja Algoritma C5.0

Pada gambar diatas dapat dilihat terhadap hasil kinerja, dari gambar tersebut untuk hasil kinerja algoritma C5.0 sebesar **87,50%**. Selanjutnya dilakukan proses pengujian pada algoritma K-Nearest Neighbor seperti terlihat gambar berikut:

accuracy: 62.50%

	true Unggul	true Tidak	class precision
pred. Unggul	4	2	66.67%
pred. Tidak	1	1	50.00%
class recall	80.00%	33.33%	

Gambar 4. Hasil Kinerja Algoritma K-Nearest Neighbor

Pada gambar 4 diatas dapat dilihat bahwasannya hasil kinerja untuk algoritma K-Nearest Neighbor sebesar **62,50%**. Hal ini menandakan bahwasannya algoritma C5.0 memiliki kinerja lebih baik untuk melakukan proses klasifikasi pada bibit unggul

4. KESIMPULAN

Proses akhir dari penelitian merupakan proses penarikan kesimpulan, adapun hasil kesimpulan yang didapatkan dari penelitian bahwasannya dengan menggunakan algoritma data mining dapat membantu mempermudah dalam proses klasifikasi yang dilakukan pada penentuan bibit unggul. Proses dengan menggunakan algoritma C5.0 dan K-Nearest Neighbor telah berhasil dilakukan, dari proses pengujian yang dilakukan bahwasannya algoritma C5.0 memiliki hasil kinerja lebih baik sebesar **87,50%** dibandingkan dengan algoritma K-Nearest Neighbor sebesar **62,50%**.

REFERENCES

- [1] U. Surapati and M. Jannah, "Penerapan Data Mining Menggunakan Metode K-Means Untuk Mengetahui Minat Customer Dalam Pembelian Merchandise Kpop," *J. Sains dan Teknol.*, vol. 5, no. 3, pp. 875–884, 2024, doi: 10.55338/saintek.v5i3.2739.
- [2] R. E. Marpaung, "Penerapan Algoritma K-Means Dalam Mengclustering Kualitas Bibit Kelapa Sawit Di PPKS Marihat," *Progr. Stud. Sist. Inf.*, vol. 1, no. 1, pp. 7–15, 2022.
- [3] Rizki Ananda Putra Fajar, Rakhmat Kurniawan, and Sriani, "Penerapan Metode Vikor dalam Pemilihan Bibit Unggul Pohon Karet," *Da'watuna J. Commun. Islam. Broadcast.*, vol. 4, no. 4, pp. 1550–1560, 2024, doi: 10.47467/dawatuna.v4i4.1842.
- [4] S. Vrtiwi and N. Padang, "Penerapan Metode Naïve Bayes Pada Sistem Penunjang Keputusan Bibit Unggul Kelapa Sawit," *J. Pustaka*, vol. 4, no. 2, pp. 31–37, 2024.
- [5] 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.
- [6] A. D. Adhi Putra, "Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.
- [7] A. Fikri Sallaby, "Analysis of Missing Value Imputation Application with K-Nearest Neighbor (K-NN) Algorithm in Dataset," *Int. J. Informatics Comput. Sci.*, vol. 5, no. 2, pp. 141–144, 2021, doi: 10.30865/ijics.v5i2.3185.
- [8] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, "Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 81–86, 2020, doi: 10.33096/ilkom.v12i2.507.81-86.
- [9] R. Ella Sari, Solikhun, and F. Rizky, "Penerapan Algoritma C5.0 dalam Memprediksi Persediaan Buah pada UD. Bunda Syafira Buah," *JUKI J. Komput. dan Inform.*, vol. 3, no. 2, pp. 59–63, 2021, doi: 10.53842/juki.v3i2.62.
- [10] F. N. Giustin, B. N. Sari, and T. N. Padilah, "Application of C5.0 Algorithm in Prediction of Learning Outcomes in Calculus Subject," *J. Appl. Eng. Technol. Sci.*, vol. 3, no. 2, pp. 90–97, 2022, doi: 10.37385/jaets.v3i2.673.
- [11] A. T. Yuliandari, Z. Sari, and V. R. S. Nastiti, "Pemetaan Mata Kuliah Yang Berpengaruh Pada Kelulusan Tidak Tepat Waktu Mahasiswa Informatika UMM Menggunakan SOM," *J. Repos.*, vol. 3, no. 1, pp. 111–120, 2024, doi: 10.22219/repository.v3i1.31016.
- [12] T. Permana, A. M. Siregar, A. F. N. Masruriyah, and A. R. Juwita, "Perbandingan Hasil Prediksi Kredit Macet pada Koperasi Menggunakan Algoritma KNN dan C5.0," *Conf. Innov. Appl. Sci. Technol.*, vol. 3, no. 1, pp. 737–746, 2020, [Online]. Available: <http://publishing-widyagama.ac.id/ejournal-v2/index.php/ciastech/article/view/1970>.
- [13] F. N. Umma, B. Warsito, and D. A. I. Maruddani, "Klasifikasi Status Kemiskinan Rumah Tangga Dengan Algoritma C5.0 Di Kabupaten Pematang," *J. Gaussian*, vol. 10, no. 2, pp. 221–229, 2021, doi: 10.14710/j.gauss.v10i2.29934.
- [14] E. Novianto, A. Hermawan, and D. Avianto, "Klasifikasi Algoritma K-Nearest Neighbor, Naive Bayes, Decision Tree Untuk Prediksi Status Kelulusan Mahasiswa S1," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 8, no. 2, pp. 146–154, 2023, doi: 10.36341/rabit.v8i2.3434.
- [15] D. P. Indini, Mesran, and Dito Putro Utomo, "Penerapan Data Mining Dalam Pengelompokan Data Reseller di Telkomsel Authorized Partner (TAP) Deli Tua Dengan Algoritma K-Means," *J. Ilm. Media Sisfo*, vol. 17, no. 2, pp. 189–202, 2023, doi: 10.33998/mediasisfo.2023.17.2.1391.
- [16] M. Mesran, M. Syahrizal, S. Sarwandi, S. Aripin, D. P. Utomo, and A. Karim, "A comparison of the performance of data mining classification algorithms on medical datasets with the application of data normalization," *AIP Conf. Proc.*, vol. 3048, no. 1, 2024, doi: 10.1063/5.0207994.
- [17] U. R. Amanda and D. P. Utomo, "Penerapan Data Mining Algoritma Hash Based Pada Data Pemesanan Buah Impor Cv.



- Green Uni Fruit,” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 5, no. 1, 2021.
- [18] B. S. Pranata and D. P. Utomo, “Penerapan Data Mining Algoritma FP-Growth Untuk Persediaan Sparepart Pada Bengkel Motor (Study Kasus Bengkel Sinar Service),” *Bull. Inf. Technol.*, vol. 1, no. 2, pp. 83–91, 2020.
- [19] I. Arfyanti, M. Fahmi, and P. Adytia, “Penerapan Algoritma Decision Tree Untuk Penentuan Pola Penerima Beasiswa KIP Kuliah,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1196–1201, 2022, doi: 10.47065/bits.v4i3.2275.
- [20] N. H. Harani and F. S. Damayanti, “Implementasi Algoritma C5.0 Untuk Menentukan Pelanggan Potensial Di Kantor Pos Cimahi,” *J. SITECH Sist. Inf. dan Teknol.*, vol. 4, no. 1, pp. 69–76, 2021, doi: 10.24176/sitech.v4i1.6281.
- [21] D. Fitriana, W. Gunawan, and A. Puspita Sari, “Studi Komparasi Algoritma Klasifikasi C5.0, SVM dan Naive Bayes dengan Studi Kasus Prediksi Banjir Comparative Study of Classification Algorithm between C5.0, SVM and Naive Bayes with Case Study of Flood Prediction,” *Februari*, vol. 21, no. 1, pp. 1–11, 2022.
- [22] R. N. Amalda, N. Millah, and I. Fitria, “Implementasi Algoritma C5.0 Dalam Menganalisa Kelayakan Penerima Keringanan Ukt Mahasiswa Itk,” *Teorema Teor. dan Ris. Mat.*, vol. 7, no. 1, p. 101, 2022, doi: 10.25157/teorema.v7i1.6692.
- [23] A. Putri *et al.*, “Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 20–26, 2023, doi: 10.57152/malcom.v3i1.610.
- [24] J. Supriyanto, D. Alita, and A. R. Isnain, “Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring,” *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 4, no. 1, pp. 74–80, 2023, doi: 10.33365/jatika.v4i1.2468.
- [25] R. Sari, “Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-Nn),” *EVOLUSI J. Sains dan Manaj.*, vol. 8, no. 1, pp. 10–17, 2020, doi: 10.31294/evolusi.v8i1.7371.