

KLASIFIKASI MASA AWAL PANEN SAWIT PADA PT. MUSTIKA SEMBULUH MENGUNAKAN ALGORITMA CLASSIFICATION AND REGRESSION TREE

Poppy Sulistiawaty¹⁾, Nurahman²⁾

¹ Fakultas Ilmu Komputer, Universitas Darwan Ali
email: poppys4evr@gmail.com

² Fakultas Ilmu Komputer, Universitas Darwan Ali
email: nurahman@unda.ac.id

Abstract



Permasalahan dalam penentuan masa awal panen kelapa sawit memberikan dampak terhadap hasil produksi berikutnya. Selain nutrisi yang cukup terhadap pohon kelapa sawit dan banyaknya tangkai pelepah serta kondisi buah yang seimbang pada setiap pohon sawit, maka masa awal panen juga merupakan salah satu hal yang penting. Penentuan masa awal panen kelapa sawit dipengaruhi dengan berbagai variabel diantaranya tahun tanam, luas area, jarak tanam, jenis bibit dan berbagai variabel lainnya. Banyaknya variabel dan data tentunya cukup sulit untuk menentuan masa awal panen kelapa sawit. Untuk itu, Algoritma klasifikasi dalam data mining dapat membantu untuk menganalisis penentuan masa awal panen kelapa sawit. Salah satu model dari metode klasifikasi yang mudah digunakan oleh siapa saja adalah pohon keputusan atau yang disebut dengan decision tree. Penelitian ini bertujuan untuk membangun suatu model pohon keputusan yang dapat digunakan untuk memprediksi masa awal panen kelapa sawit. Algoritma yang digunakan pada penelitian ini adalah Classification and Regression Tree. Proses pengolahan data dilakukan pengujian klasifikasi terhadap dataset sebanyak 5 kali. Hasil penelitian menunjukkan bahwa Algoritma Classification and Regression Tree mampu menghasilkan model pohon keputusan dengan nilai performa akurasi tertinggi 96,21%. Node akar pada pohon keputusan dimuai dari variabel tahun tanam dan dilanjutkan dengan variabel yang lainnya.

Keywords: Data Mining, Cart, Panen, Sawit

1. PENDAHULUAN

PT Wilmar menangani produk minyak sawit di Indonesia. PT Wilmar Group adalah salah satu produsen minyak sawit terbesar di Indonesia dengan perkebunan dan fasilitas pengolahan kelapa sawit di beberapa wilayah seperti Sumatera, Kalimantan, dan Sulawesi. Lokasi pabrik dan perkebunan PT Wilmar Group di Kalimantan Tengah memiliki beberapa unit salah satunya adalah PT Mustika Sembuluh yang beroperasi di Kotawaring Timur. Waktu tanam serta mamastikan bahwa tanaman sawit mendapatkan nutrisi yang cukup merupakan variabel yang mempengaruhi hasil produksi. PT

Mustika Sembuluh Estate MS 2 juga melakukan seleksi bibit kelapa sawit yang berkualitas untuk menghasilkan tanaman yang produktif dan berkualitas. Dengan luas lahan yang besar, PT Mustika Sembuluh dapat memangkas produksi kelapa sawit dan memenuhi permintaan pasar yang besar.

Pelaksanaan prosedur yang komprehensif diperlukan untuk memastikan pengelolaan perkebunan kelapa sawit yang tepat. Dalam proses ini, pemahaman yang mendetail tentang pembibitan dan penanaman kelapa sawit sangat penting. Berdasarkan pengamatan dan data yang terkumpul diketahui bahwa umur ideal untuk mulai panen kelapa sawit adalah 31 bulan.



Namun tidak semua buah kelapa sawit dapat dipanen sekaligus, karena tingkat kematangan yang berbeda.

Untuk memastikan hasil yang optimal, buah kelapa sawit harus dipanen pada tahap fraksi kedua. Tahapan ini ditandai dengan 5-10 buah lepas pada lempeng buah, warna buah berubah dari kuning menjadi jingga, dan sekitar 25-75 persen buah bagian luar lepas. Berkaitan dengan pemanenan, penting untuk memperhatikan faktor lain seperti cuaca dan kondisi tanah untuk menjaga kualitas buah yang dipanen. Dengan pengukuran dan pemantauan yang akurat, PT Mustika Sembuluh Estate MS 2 berusaha memastikan proses pemanenan kelapa sawit berjalan dengan baik dan menghasilkan produk berkualitas tinggi.

Permasalahan dalam penentuan masa awal panen kelapa sawit akan memberikan dampak terhadap hasil produksi berikutnya. Selain memberi nutrisi yang cukup terhadap pohon kelapa sawit dan banyaknya tangkai pelepah serta kondisi buah yang seimbang, maka masa awal panen juga merupakan salah satu hal yang penting. Klasifikasi merupakan salah satu solusi ampuh untuk menyelesaikan masalah terutama pada data yang berukuran besar [1]. Teknik klasifikasi banyak digunakan di berbagai bidang [2]. Misalnya, di bidang kesehatan, teknik klasifikasi dapat digunakan untuk menghitung kemungkinan terjadinya suatu penyakit berdasarkan data gejala dan faktor risiko terkait [3]. Dalam meteorologi, klasifikasi digunakan untuk memprediksi kejadian cuaca ekstrem seperti badai atau banjir berdasarkan data cuaca historis [4], [5]. Dalam industri, kategorisasi dapat membantu mengelompokkan produk atau pelanggan sesuai dengan preferensi atau perilaku mereka [6], [7]. Di bidang pertanian, teknik klasifikasi dapat digunakan untuk memprediksi hasil berdasarkan variabel seperti cuaca, jenis tanaman, dan faktor pertanian lainnya [8], [9]. Selain itu, teknik klasifikasi dapat diterapkan

pada banyak bidang lain, seperti keuangan, pendidikan, dan penelitian.

Algoritma yang digunakan dalam data mining memiliki beberapa pilihan yang dapat disesuaikan dengan kebutuhan dan karakteristik permasalahan yang dihadapi. Saat memilih algoritma yang tepat, output yang dihasilkan harus dipertimbangkan sedemikian rupa sehingga mudah dipahami oleh semua orang yang terlibat. *Decision tree* adalah salah satu algoritma yang paling populer dan memberikan hasil yang mudah dipahami [10].

Pohon keputusan adalah representasi grafis yang menggambarkan sekumpulan keputusan dan aturan yang digunakan untuk mengklasifikasikan atau memprediksi data. Dalam pohon keputusan, setiap cabang mewakili atribut atau fitur yang digunakan untuk membuat keputusan, sedangkan setiap daun atau simpul terminal mewakili hasil akhir atau klasifikasi yang dihasilkan. Keunggulan pohon keputusan terletak pada kemampuannya untuk memberikan hasil yang mudah diinterpretasikan dan dijelaskan secara visual. Selain itu, pohon keputusan dapat digunakan untuk menganalisis hubungan antar variabel dan mengambil keputusan berdasarkan aturan yang telah ditentukan.

Pohon keputusan merupakan salah satu metode klasifikasi yang paling populer [11]. Beberapa Algoritma yang menghasilkan luaran berupa sebuah pohon keputusan seperti *ID3*, *C4.5*, *CART*, dan beberapa algoritma lainnya. Algoritma *Classification and Regression Tree* atau biasa dikenal dengan Algoritma *CART* adalah salah satu Algoritma yang cukup populer dalam analisis data. Algoritma ini terdiri dari dua jenis pohon yaitu pohon regresi dan pohon klasifikasi [12].

Jika variabel dependen adalah variabel kategori, *CART* membuat pohon klasifikasi untuk mengklasifikasikan data berdasarkan kategori yang ada [13]. Namun, jika variabel

dependen kontinu atau numerik, CART membuat pohon regresi untuk memprediksi nilai kontinu [14]. Pohon klasifikasi CART membantu mengatur dan mengklasifikasikan data berdasarkan karakteristik yang ada. Setiap simpul dari pohon klasifikasi menggambarkan keputusan yang dibuat berdasarkan aturan yang telah ditentukan sebelumnya.

Pohon regresi CART juga digunakan untuk memperkirakan variabel dependen kontinu. Algoritma CART memungkinkan kita membuat model sesuai dengan jenis data yang akan dianalisis. Algoritma ini mengubah himpunan data menjadi pohon keputusan dengan aturan-aturan yang mempermudah proses pengambilan keputusan dengan atribut kompleks [15]. Metode ini berguna di banyak bidang, mulai dari penelitian ilmiah hingga aplikasi komersial.

Penelitian Louis dkk [16] Melakukan perbandingan beberapa Algoritma untuk klasifikasi promo karyawan. Hasil penelitian menunjukkan KNN memiliki tingkat performa lebih baik. Tetapi Algoritma Decision Tree dan random Forest juga tingkat akurasi masing-masing di atas nilai 80%. Pada penelitian mereka pohon keputusan akan dibangun dengan menggunakan Algoritma C45 sedangkan pada artikel ini pohon keputusan akan dibangun menggunakan algoritma CART.

Hasil penelitian [17] menunjukkan bahwa algoritma dengan model pohon keputusan memiliki nilai akurasi lebih baik dari pada algoritma Naïve Bayes. Kemudian Membangun model pohon keputusan dengan algoritma C.45 untuk membantu perusahaan agar mengetahui pengelompokan dan pengolahan limbah cair tersebut agar tidak berdampak buruk bagi sekitar perusahaan [18].

Penelitian Andi menghasilkan dua model pohon keputusan untuk membantu dalam penyelesaian masalah [19]. Permasalahan perusahaan dalam memprediksi produksi

komoditi tanaman diselesaikan dengan menggunakan pohon keputusan [20].

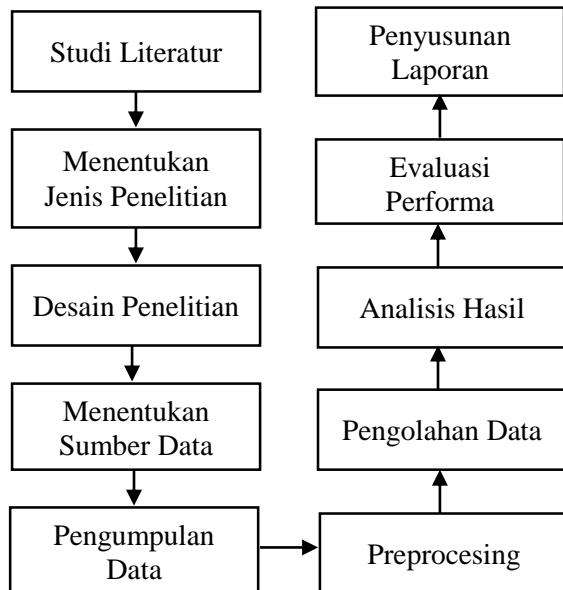
Selain tingkat performanya pada model pohon keputusan cukup baik model pohon keputusan sangat mudah difahami semua kalangan oleh karena itu penelitian ini berupaya membangun pohon keputusan. Penelitian ini memiliki perbedaan dalam teknik atau Algoritma yang digunakan dalam membangun pohon keputusan. Pada penelitian ini model pohon keputusan dibangun dengan menggunakan Algoritma CART. Selain itu, pada penelitian ini juga dilakukan pengujian performa untuk mengetahui model terbaik dilakukan 5 kali.

Penelitian ini dilakukan dengan tujuan utama membangun pohon keputusan menggunakan algoritma CART. Penelitian ini juga mengevaluasi tingkat performa model yang dihasilkan oleh algoritma CART untuk mengetahui seberapa akurat model tersebut pada klasifikasi. Selain itu, penelitian ini bertujuan untuk menentukan masa panen awal tanaman kelapa sawit yang lebih akurat, yang dapat membantu petani dan industri kelapa sawit untuk merencanakan dan mengelola panen mereka. Hasil penelitian ini diharapkan dapat membuka peluang untuk pengembangan lebih lanjut penggunaan algoritma CART dibidang pertanian khususnya untuk penentuan waktu panen tanaman kelapa sawit.

2. METODE PENELITIAN

Metode penelitian merupakan suatu strategi atau pendekatan yang digunakan untuk merencanakan dan menjalankan suatu penelitian. Metode ini perlu ditetapkan karena akan mempengaruhi prosedur yang digunakan untuk menemukan jawaban atas pertanyaan penelitian serta untuk mencapai tujuan awal penelitian. Untuk itu pada penelitian ini dirancang tahapan penelitian yang dapat digunakan sebagai langkah sistematis dalam pelaksanaan penelitian. Adapun tahapan penelitian dapat dilihat berdasarkan

gambar 1. Tahapan penelitian dilakukan mulai dari studi literatur hingga kesimpulan hasil penelitian.



Gambar 1 Tahapan Penelitian

2.1 Lokasi Penelitian

Lokasi penelitian ini adalah PT Mustika Sembuluh. PT Mustika Sembuluh merupakan bagian dari Grup Wilmar. Perkebunan PT. Mustika Sembuluh berlokasi di Kabupaten Kotawaringin Timur. Perusahaan ini memiliki lahan yang sangat luas untuk kegiatan perkebunan. Selain itu, PT. Mustika Sembuluh juga dikenal memiliki jumlah karyawan yang sangat banyak. Keberadaan perusahaan ini memberikan kontribusi signifikan terhadap perekonomian di wilayah tersebut. Dengan memiliki lahan yang luas dan jumlah karyawan yang banyak, PT. Mustika Sembuluh memiliki potensi untuk menghasilkan produksi kelapa sawit yang besar dan berperan penting dalam industri agribisnis di daerah tersebut.

2.2 Sumber Data

Sumber data yang akan digunakan yaitu bersumber dari laporan crop statistik perusahaan.

Karena Sumber data yang digunakan adalah berupa laporan crop statistik perusahaan maka jenis data yang digunakan dalam penelitian ini adalah tergolong pada data sekunder.

2.3. Pengumpulan Data

Dalam penelitian ini, pengumpulan data dilakukan dengan melakukan komunikasi aktif dengan pihak terkait penyaji laporan *crop statistic* perusahaan. Melalui komunikasi ini, peneliti memperoleh data berupa jenis rekap laporan yang telah tersusun dalam bentuk beberapa tabel. Proses pengumpulan data secara langsung dari sumbernya memastikan keakuratan dan keandalan data yang digunakan dalam penelitian. Dengan menggunakan data dari laporan *crop statistic* perusahaan, penelitian ini dapat memberikan wawasan yang komprehensif dan mendalam terkait topik penelitian yang sedang diteliti.

2.4 Pengolahan Data, Analisis dan Performa

Pengolahan data pada penelitian ini dilakukan dengan menggunakan algoritma *CART* (*Classification and Regression Trees*). Penggunaan algoritma *CART* ini memungkinkan peneliti untuk mendapatkan wawasan yang mendalam dari data dan memahami karakteristik serta faktor-faktor yang berpengaruh dalam penelitian ini. Dengan demikian, algoritma *CART* menjadi salah satu pendekatan yang kuat untuk mengolah data dalam penelitian ini. Perhitungan pada Algoritma *CART* dilakukan dua tahap yaitu mencari nilai *index gini* dan mencari nilai *gini gain*. Hal ini dilakukan berulang hingga terbentuk pohon keputusan. Adapun persamaan pada algoritma *CART* disajikan pada persamaan (1) dan (2).

Mencari nilai *Index Gini* :

$$Index\ Gini = 1 - \sum_{i=1}^k p_i^2 \quad (1)$$

Mencari nilai *Index gini* terlebih dahulu dilakukan dengan mengetahui kasus dalam setiap

fitur. Kemudian pada kasus akan memiliki beberapa pilihan kondisi. Setelah semua kasus dan kondisi diketahui maka langkah berikutnya melakukan perhitungan dengan persamaan pada index gini $1 - ((kondisi/kasus)^2 + (kondisi/kasus)^2)$. Mencari nilai *Gini Gain*:

$$Gini\ Gain = Gini(A, S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times Gini(S_i)$$

Nilai pada *Gini Gain* dapat diperoleh dengan melakukan perhitungan sesuai pada persamaan 2. Perhitungan persamaan 2 yaitu nilai *index gini* label dikurang dengan nilai jumlah kasus dibagi nilai seluruh *record* pada *dataset* kemudian dikali dengan nilai *gini index* setiap kasus. Jika diperhatikan persamaan bahwa terdapat *symbol sigma* (Σ) artinya hal ini akan menjumlah bagian-bagian pada kasus yang lain juga yang terdapat pada fitur tersebut. Sehingga nilai *Gini Gain* pada suatu fitur telah diketahui.

Evaluasi performa pada algoritma *CART* dilakukan dengan memperhatikan nilai tingkat performa yang dihasilkan. Salah satu metrik evaluasi yang penting adalah akurasi. Semakin tinggi persentase akurasi yang dihasilkan, semakin baik performa dari model pohon keputusan tersebut dalam mengklasifikasikan data. Nilai akurasi (*acc*) adalah proporsi jumlah prediksi yang benar. Dapat dihitung dengan menggunakan persamaan (3):

$$acc = \frac{tp + tn}{tp + tn + fp + fn} \quad (3)$$

Untuk mengetahui nilai akurasi maka dalam penelitian ini akan menyajikan tabel *Confusion Matrix*. Pada tabel *Confusion Matrix* akan berisi tentang nilai true positive (tp), true negative (tn), false positive (fp), dan false negative (fn). Dengan diperolehnya tabel *Confusion Matrix* maka nantinya akan dapat mencari nilai akurasi.

3. HASIL DAN PEMBAHASAN

Penelitian ini dilakukan berdasarkan tahapan yang telah di desain dalam metode penelitian. Tahapan penelitian dilakukan mulai dari studi literature hingga penyajian hasil

laporan. Sebagaimana yang telah ditentukan bahwa sumber data yang digunakan dalam penelitian ini yaitu dari laporan crop statistik perusahaan. Untuk itu, hasil dari komunikasi yang dilakukan dalam pengumpulan data bahwa data penelitian diperoleh sebanyak 147 *record* sebagai sample dalam penelitian ini yang akan diolah dan dianalisis untuk dapat diketahui *posisi* dari *dataset* tersebut.

3.1. Dataset

Dataset pada penelitian ini berjumlah 147 *record* sebagai sample penelitian. Untuk kebutuhan pengolahan data maka *dataset* perlu disiapkan berdasarkan kebutuhan dari Algoritma yang akan digunakan. Penelitian ini menggunakan Algoritma *CART*. Algoritma *CART* dapat mengolah data dengan *type data* binominal, polinomial, integer, real dan sebagainya. Pada penelitian ini *type data* yang digunakan adalah sebagai mana yang disajikan pada tabel 1.

Tabel 1. *Type data* Pada *Dataset*

Fitur	Type data	Keterangan
TT	Tahun Tanam	Tahun Tanam
Blok	Polynomial	Nama Tempat Tanam Sawit
DM	Nominal	Devisi Manager
BT	Nominal	Jenis Bibit dan Tahun
Ha	Real	Luas Lahan
JPSH	Integer	Jumlah Pohon Dalam Satu Hektar
PB	Integer	Umur Pohon berdasarkan Bulan
BP	Nominal	Mulai Panen/ Buah Pasir/Label

Tabel 1 menunjukkan bahwa v_i (3) yang terdapat dalam penelitian ini berjumlah 8 variable. 8 variabel ini akan dipertimbangkan sebelum melakukan proses pengolahan data. Untuk mempertimbangkan variabel dalam

pengolahan data masuk pada tahapan preprocessing data.

3.2. Preprocessing Data

Tahapan *preprocessing* data ini mempertimbangkan data yang akan diproses dalam pengolahan. Untuk ini pada penelitian ini diperlukan pemahaman dan deskripsi setiap fitur yang diperoleh dari sumber data. Peneliti melakukan wawancara kepada pihak terkait tentang deskripsi dari fitur yang akan digunakan dalam penelitian. Dalam deskripsi dan penjelasan dari sumber data bahwa terdapat satu fitur yang dianggap tidak diperlukan dalam penelitian. Hasil pemilihan fitur yang relevan dalam penelitian disajikan pada tabel 2 berikut.

Tabel 2. Fitur Hasil *Preprocessing* Data

Fitur	Type data	Keterangan
TT	Tahun Tanam	Tahun Tanam
DM	Nominal	Devisi Manager
BT	Nominal	Jenis Bibit dan Tahun
Ha	Real	Luas Lahan
JPSH	Integer	Jumlah Pohon Dalam Satu Hektar
PB	Integer	Umur Pohon berdasarkan Bulan
BP	Nominal	Mulai Panen/ Buah Pasir/Label

Pada tabel 2 diketahui bahwa fitur blok tidak digunakan dalam proses pengolahan dan analisis data. Hal ini dipertimbangkan bahwa fitur blok hanya merupakan nama tempat yang menjadi lokasi penanaman pohon sawit. Selain itu juga bahwa mengenai lokasi penanaman sawit sudah diwakili dengan fitur luas area dan jumlah penanaman pohon, untuk itu maka fitur blok tidak digunakan dalam penelitian ini.

Pada tahapan preprocessing juga ditemukan beberapa *record* dengan value typo. Kemudian peneliti berusaha untuk menganalisa dengan kesesuaian data yang dibutuhkan terhadap value yang lainnya. Jika peneliti tidak

menemukan mengenai value yang tepat dalam analisa maka dilakukan konfirmasi terhadap sumber data.

3.3 Pengolahan Data

Pada penelitian ini pengujian klasifikasi dilakukan sebanyak 5 kali. Pada setiap tahapan klasifikasi data akan di bagi menjadi dua bagian yaitu sebagai data training dan sebagai data testing. Pada pengujian klasifikasi pertama data training yang digunakan sebanyak 60% sedangkan data testing 40% dari data sebanyak 147 record data hingga 5 kali sampai data training yang digunakan 100%. Dari pengujian klasifikasi data dapat diketahui hasil pengujian klasifikasi ditunjukkan pada tabel 3 berikut.

Tabel 3. Klasifikasi *Dataset*

Data Training	Persentase Data Training	Kebenaran Prediksi	Kesalahan Prediksi
88	60%	84	4
102	70%	98	4
117	80%	111	6
132	90%	127	5
147	100%	141	6

Tabel 3 tersebut merupakan hasil dari suatu algoritma klasifikasi pada berbagai tingkat persentase data training yang berbeda. Data training merupakan bagian dari *dataset* yang digunakan untuk melatih algoritma dan mengembangkan model prediksi. Persentase data training menunjukkan seberapa besar proporsi data yang digunakan untuk pelatihan dari keseluruhan *dataset*. Dalam tabel ini, terdapat lima tingkat persentase data training yang diuji, yaitu 60%, 70%, 80%, 90%, dan 100%.

Hasil evaluasi klasifikasi dicantumkan dalam dua metrik, yaitu "Kebenaran Prediksi" dan "Kesalahan Prediksi". Kebenaran prediksi menunjukkan jumlah data yang berhasil diprediksi dengan benar oleh algoritma,

sedangkan kesalahan prediksi menunjukkan jumlah data yang salah diprediksi oleh algoritma. Contohnya, pada persentase data training 70%, algoritma berhasil memprediksi dengan benar sebanyak 98 data, namun juga melakukan kesalahan dalam memprediksi 4 data.

3.4 Evaluasi Performa

Evaluasi performa pada algoritma CART dilakukan dengan memperhatikan nilai akurasi yang dihasilkan oleh model. Akurasi merupakan metrik yang penting dalam mengukur sejauh mana model algoritma CART mampu melakukan klasifikasi data dengan benar. Nilai akurasi mengindikasikan persentase keberhasilan model dalam memprediksi kategori atau label yang tepat berdasarkan data training yang telah digunakan. Semakin tinggi nilai akurasi yang diperoleh, semakin baik performa dari model algoritma CART dalam mengklasifikasikan data dengan akurat. Evaluasi performa dengan mempertimbangkan nilai akurasi ini penting untuk memastikan kualitas dan efektivitas model algoritma CART dalam melakukan klasifikasi data dengan tepat dan akurat. Pada penelitian ini nilai performa akurasi di sajikan pada tabel 4 berikut ini.

Tabel 4. Hasil Evaluasi Performa Model

Data Training	Persentase Data Training	Performa Akurasi
88	60%	95,45%
102	70%	96,08%
117	80%	94,87%
132	90%	96,21%
147	100%	95,92%

Tabel 4 tersebut merupakan hasil evaluasi performa dari suatu model algoritma klasifikasi yang dihasilkan oleh algoritma CART. Pada tabel menunjukkan adanya perbedaan tingkat persentase data training sehingga menghasilkan tingkat performa yang berbeda pula. Data training adalah bagian dari *dataset* yang digunakan untuk melatih model, dan persentase data training menunjukkan seberapa besar persentase data tersebut dari keseluruhan *dataset*.

Hasil evaluasi performa ditunjukkan dalam bentuk metrik akurasi, yang mengukur sejauh mana model mampu melakukan klasifikasi data dengan benar. Akurasi dihitung sebagai persentase dari jumlah data yang berhasil diprediksi dengan benar dibandingkan dengan keseluruhan data yang ada. Semakin tinggi nilai akurasi yang diperoleh, semakin baik performa dari model algoritma klasifikasi dalam mengklasifikasikan data dengan akurat.

Model pohon keputusan yang dihasilkan dalam mengklasifikasi *dataset* menunjukkan tingkat performa yang berbeda-beda. Tingkat performa pada model pohon keputusan berkisar antara 94,87% hingga 96,21%. Tingkat performa tertinggi berda pada pengujian data training sebanyak 90%. Untuk itu model pohon keputusan yang direkomendasikan pada penelitian ini terdapat pada data uji klasifikasi dengan data training 90% yang memiliki nilai performa akurasi 96,21%.

Nilai akurasi 92,21% diperoleh dari tabel *confusion matrix*. Pada tabel *confusion matrix* memperlihatkan adanya nilai kebenaran dalam memprediksi dan nilai kesalahan dalam memprediksi suatu algoritma. Tabel *confusion matrix* dapat dilihat pada tabel 5 berikut.

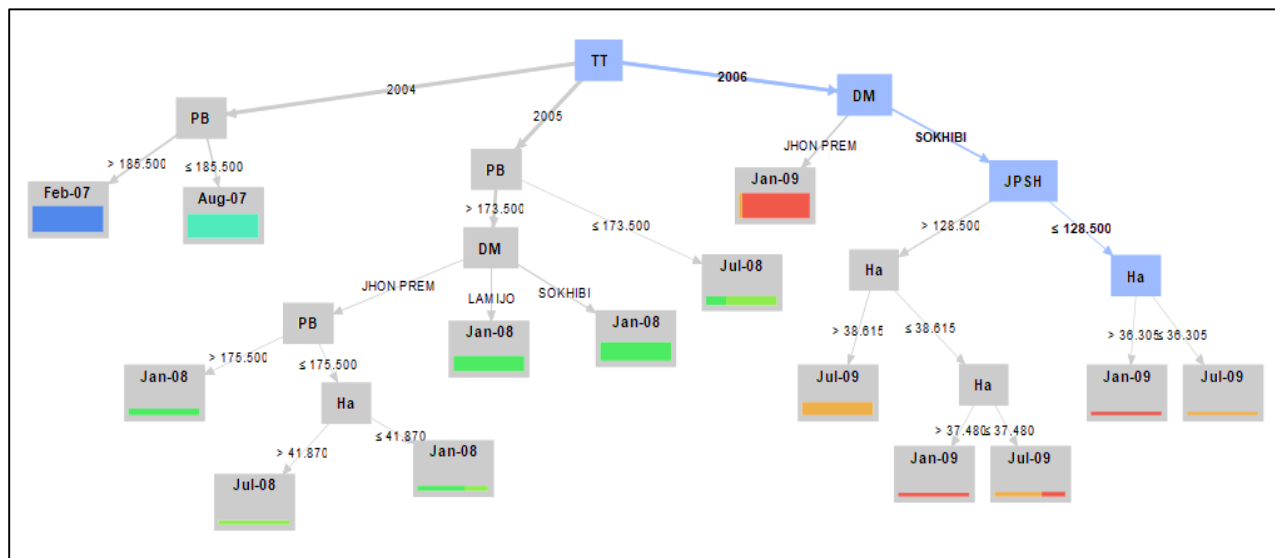
Tabel 5. *Confusion Matrix*

True Feb-07	True Aug-07	True Jan-08	True Jul-08	True Jul-09	True Jan-09	Class Precision
-------------	-------------	-------------	-------------	-------------	-------------	-----------------

Pred. Feb-07	23	0	0	0	0	0	100.00%
Pred. Aug-07	0	21	0	0	0	0	100.00%
Pred. Jan-08	0	0	36	1	0	0	97.30%
Pred. Jul-08	0	0	2	7	0	0	77.78%
Pred. Jul-09	0	0	0	0	15	1	93.75%
Pred. Jan-09	0	0	0	0	1	25	96.15%
Class Recall	100.00%	100.00%	94.74%	87.50%	93.75%	96.15%	

Pohon Keputusan yang disarankan pada hasil penelitian ini memiliki akar pada fitur TT atau disebut tahun tanam. Akar node menjadi start dalam memprediksi tahun awal panen sawit. Kemudian jika fitur tahun tanam berada pada

tahun 2004 maka akan masuk fitur PB, jika tahun tanam 2005 masuk pada fitur PB, dan jika tahun tanam 2006 maka masuk pada fitur DM demikian selanjutnya. Pohon keputusan hasil dari data training 90% ditunjukkan pada gambar 2 berikut.



Gambar 2. Pohon Keputusan

Model pohon keputusan yang direkomendasikan dalam penelitian ini memiliki kemampuan untuk memprediksi masa awal panen kelapa sawit. Model ini dapat digunakan untuk memprediksi masa tanam sawit yang berada dalam rentang tahun 2004 sampai 2006.

Dengan menggunakan data masa tanam dan variabel-variabel lain yang relevan, model pohon keputusan dapat memberikan prediksi mengenai waktu awal panen kelapa sawit

berdasarkan karakteristik dan faktor-faktor yang mempengaruhinya.

Model pohon keputusan digunakan untuk mengambil keputusan berdasarkan serangkaian kondisi atau atribut yang terdapat pada data training. Model ini akan membentuk pohon keputusan yang mewakili klasifikasi dan aturan-aturan yang digunakan untuk memprediksi masa awal panen kelapa sawit. Hasil prediksi dari model dapat memberikan estimasi waktu panen

yang lebih akurat berdasarkan informasi dari data masa tanam sebelumnya.

Dengan menggunakan model pohon keputusan ini, peneliti dapat memanfaatkan hasil prediksi untuk perencanaan dan pengambilan keputusan terkait kelapa sawit, seperti pengaturan jadwal panen, estimasi produksi, dan pengelolaan lahan. Model pohon keputusan ini dapat menjadi alat yang berguna dalam mendukung pengelolaan kebun kelapa sawit secara efisien dan optimal.

4. KESIMPULAN

Penelitian ini melakukan klasifikasi terhadap *dataset* penelitian sebanyak 147 *record*. Dari sebanyak 147 *record* data dilakukan pengujian klasifikasi menggunakan algoritma CART sebanyak 5 kali yang dilakukan dengan pembagian data training menjadi 60%, 70%, 80%, 90%, dan 100%.

Model pohon keputusan yang direkomendasikan pada penelitian ini adalah hasil dari pengujian data training 90%. Pada pengujian data training 90% memberikan root akar awal pada pohon keputusan adalah tahun tanam pohon sawit yang selanjutnya mengarah ke fitur-fitur lainnya. Performa dari model pohon keputusan yang direkomendasikan pada penelitian ini berada pada nilai akurasi sebesar 96,21%.

5. REFERENSI

- [1] M. Mardiani, "Desain Model Data Mining pada Model SECI untuk Pemetaan dan Ekstraksi Pengetahuan Kompetensi Lulusan," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 3, 2021, doi: 10.35957/jatisi.v8i3.1349.
- [2] C. R. Sari, "Teknik Data Mining Menggunakan Classification Dalam Sistem Penunjang Keputusan Peminatan SMA Negeri 1 Polewali," *IJNS – Indones. J. Netw. Secur.*, vol. 5, no. 1, 2016.
- [3] W. Apriliah, I. Kurniawan, M. Baydhowi, and T. Haryati, "Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest," *SISTEMASI*, vol. 10, no. 1, 2021, doi: 10.32520/stmsi.v10i1.1129.
- [4] B. P. T.P and R. D. Indah Sari, "Penerapan Data Mining Untuk Prakiraan Cuaca Di Kota Malang Menggunakan Algoritma Iterative Dichotomiser Tree (ID3)," *Jouticla*, vol. 2, no. 2, 2017, doi: 10.30736/jti.v2i2.68.
- [5] A. Luthfiarta, A. Febriyanto, H. Lestiawan, and W. Wicaksono, "Analisa Prakiraan Cuaca dengan Parameter Suhu, Kelembaban, Tekanan Udara, dan Kecepatan Angin Menggunakan Regresi Linear Berganda," *JOINS (Journal Inf. Syst.)*, vol. 5, no. 1, 2020, doi: 10.33633/joins.v5i1.2760.
- [6] S. Rahayu and J. J. Purnama, "Klasifikasi Konsumsi Energi Industri Baja Menggunakan Teknik Data Mining," *J. Teknoinfo*, vol. 16, no. 2, 2022, doi: 10.33365/jti.v16i2.1984.
- [7] S. R. S. Octavianty, "Implementasi Algoritma Cart Untuk Memprediksi Pembayaran Kredit Customer (Studi Kasus: Pt. Napoleon Light Industries)," *Pelita Inform. Inf. dan ...*, vol. 7, 2019.
- [8] A. Rismayana and D. Rosdiana, "Penerapan Algoritma C4.5 Pada Bidang Pertanian," *J. TEDC,13*, vol. 13, no. 3, 2019.
- [9] E. Novalia and A. Voutama, "Prediction of Rice Field Planted Area with CRISP-DM Using Classification and Regression Tree (Cart) Algorithms," 2023.
- [10] R. Nursyahfitri, A. N. Maharadja, R. A. Farissa, and Y. Umaidah, "Klasifikasi Penentuan Jenis Obat Menggunakan Algoritma Decision Tree," *J. Inform.*



- Polinema*, vol. 7, no. 3, 2021, doi: 10.33795/jip.v7i3.629.
- [11] P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno, "Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5," in *PRISMA (Prosiding Seminar Nasional Matematika)*, 2020, vol. 3.
- [12] N. Indah Prabawati, Widodo, and H. Ajie, "Kinerja Algoritma Classification And Regression Tree (Cart) dalam Mengklasifikasikan Lama Masa Studi Mahasiswa yang Mengikuti Organisasi di Universitas Negeri Jakarta," *PINTER J. Pendidik. Tek. Inform. dan Komput.*, vol. 3, no. 2, 2019, doi: 10.21009/pinter.3.2.9.
- [13] Y. R. Fajriati and S.-, "Pengklasifikasian Status Kerja pada Angkatan Kerja di Kabupaten Tanah Datar Menggunakan Metode CART dan Metode CHAID," *J. Math. UNP*, vol. 7, no. 3, 2022, doi: 10.24036/unptomath.v7i3.12720.
- [14] A. Y. Sari, "Penerapan Metode Classification and Regression Trees pada Klasifikasi Kelayakan Peserta BPJS-PBI di Kelurahan Meranti Pandak Kota Pekanbaru," 2020.
- [15] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *J. Appl. Informatics Comput.*, vol. 5, no. 2, 2021, doi: 10.30871/jaic.v5i2.3200.
- [16] LOUIS MADAERDO SOTARJUA and DIAN BUDHI SANTOSO, "Perbandingan Algoritma KNN, Decision Tree, Dan Random Forest Pada Data Imbalanced Class Untuk Klasifikasi Promosi Karyawan," *J. INSTEK (Informatika Sains dan Teknol.)*, vol. 7, no. 2, 2022, doi: 10.24252/instek.v7i2.31385.
- [17] Risqianti and B. Ismanto, "Analisis Komparasi Algoritma Naive Bayes Dan C4-5 Untuk Waktu Kelulusan Mahasiswa," *IC-Tech*, vol. XII, no. 1, 2017.
- [18] W. S. Rejeki, "Implementasi Algoritma C.45 Pada Data Pengolahan Limbah Kelapa Sawit," *BEES Bull. Electr. Electron. Eng.*, vol. 2, no. 2, 2021, doi: 10.47065/bees.v2i2.696.
- [19] A. Nurkholis, Muhaqiqin, and Tr. Susanto, "Analisis Kesesuaian Lahan Padi Gogo Berbasis Sifat Tanah dan Cuaca Menggunakan ID3 Spasial (Land Suitability Analysis for Upland Rice based on Soil and Weather Characteristics using Spatial ID3)," *J. Inform.*, vol. 8, no. 2, 2020.
- [20] D. Damayanti, "Implementasi Algoritma C4.5 Prediksi Produksi Komoditas Tanaman Perkebunan Berdasarkan Luas Lahan," *TIN Terap. Inform. Nusant.*, vol. 2, no. 10, 2022, doi: 10.47065/tin.v2i10.1026.