 ISSN: 2830-0408

Identifikasi Parameter Kritis Mesin Pengolah CPO untuk Reduksi *Oil Losses* Melalui Pemodelan *Decision Tree*

Tosty Maylangi Sitorus[a*], Zhau Vic Andrico IF[b]
[a,b] Teknik Industri Agro, Politeknik ATI Padang, Jalan Bungo Pasang - Tabing, Kota padang, Sumatera Barat, 25171
* Corresponding author: tosty.sitorus@poltekatipdg.ac.id

ABSTRAK

Tingginya *oil losses* pada stasiun *digester* dan *screw press* yang mencapai 3–4% (standar operasional 2–3%) berpotensi menurunkan rendemen CPO dan efisiensi produksi. Penelitian ini bertujuan mengklasifikasikan level *oil losses* menggunakan algoritma *Decision Tree* berbasis data variabel proses (tekanan, suhu, dan waktu proses). Hasil pemodelan menunjukkan bahwa tekanan merupakan parameter paling dominan dalam menentukan klasifikasi *oil losses*, bertindak sebagai *root node* pada pemisahan pertama (ambang batas 37 mbar), diikuti oleh parameter suhu (ambang batas 95°C) dan waktu proses (ambang batas 39 menit). Evaluasi menggunakan *Confusion Matrix* menghasilkan akurasi model sebesar 57,45% (27 dari 47 data uji terklasifikasi dengan benar), dengan sensitivitas 56,52%, spesifisitas 58,33%, serta nilai Cohen's Kappa sebesar 0,1486. Meskipun model Decision Tree terbukti mampu memetakan pola dan hubungan antar-parameter operasional secara transparan, performa klasifikasi masih perlu dioptimalkan agar prediksi yang dihasilkan lebih akurat dan andal sebagai dasar pengambilan keputusan teknis.

Kata Kunci: Decision Tree, Oil Losses, CPO, RStudio

ABSTRACT

High oil losses in the digester and screw press stations, reaching 3–4% and exceeding the operational standard of 2–3%, significantly reduce CPO yield and production efficiency. This study aims to classify CPO oil loss levels using the Decision Tree algorithm based on operational process data points (pressure, temperature, and process time). The modeling results reveal that pressure is the most dominant parameter in determining oil loss classification, acting as the root node in the initial split (threshold of 37 mbar), followed by temperature (threshold of 95°C) and process time (threshold of 39 minutes). Model evaluation using a Confusion Matrix achieved an accuracy of 57.45% (correctly classifying 27 out of 47 test data), with a sensitivity of 56.52%, specificity of 58.33%, and a Cohen's Kappa value of 0.1486. Although the Decision Tree model successfully mapped the transparent relationships among operational parameters, further optimization is required to improve classification reliability for technical decision-making.

Keywords: CPO, Decision Tree, Oil Losses, RStudio

 ISSN: 2830-0408

1. Pendahuluan

Indonesia tetap menjadi produsen minyak kelapa sawit mentah (Crude Palm Oil / CPO) terbesar di dunia, yang berkontribusi secara signifikan terhadap pertumbuhan ekonomi domestik maupun rantai pasok minyak nabati global. Untuk mempertahankan daya saing di pasar internasional yang kian menuntut, para produsen minyak kelapa sawit harus terus meningkatkan efisiensi produksi, memperbaiki kualitas produk, serta meminimalkan kehilangan hasil [1]. Unit usaha pengolahan kelapa sawit berperan penting dalam mendukung rantai pasok industri minyak sawit nasional. Salah satu unit usaha mengolah tandan buah segar (TBS) menjadi *Crude Palm Oil* (CPO) di Sumatera Utara memiliki kapasitas 30 ton/jam, dengan tahapan kritis yaitu *digester* dan *screw press*. Stasiun *digester* dan *screw press*, memiliki salah satu parameter yang berpengaruh terhadap terjadinya *oil losses* yang merupakan salah satu indikator kunci untuk menilai efisiensi proses ekstraksi minyak [2]. Kondisi saat ini, nilai *oil losses* pada stasiun *digester* dan *screw press* pada unit usaha masih berada pada kisaran 3–4% sedangkan standar operasional *oil losses* 2-3%. Bagi perusahaan *oil losses* yang tinggi berdampak langsung terhadap penurunan rendemen CPO yang dihasilkan, menyebabkan kerugian produksi, serta menurunkan efisiensi proses pengolahan pada tahapan selanjutnya. sehingga secara tidak langsung proses produksi belum berjalan secara efisien dan Perusahaan mengalami kerugian. Terjadinya kehilangan minyak dapat dipengaruhi oleh efisiensi mesin yang rendah, program pemeliharaan yang buruk, dan kebocoran (akibat kerusakan atau tangki yang meluap), serta penggunaan variabel proses yang tidak efektif [2]. Berdasarkan hasil observasi pada unit usaha, *Oil losses* pada proses *digester* dan *screw press* dapat terjadi karena pengaturan parameter operasi yang kurang tepat, seperti suhu *digester*, lama pengadukan, tekanan screw press, dan kecepatan putaran alat. *Oil Losses* yang terdapat pada stasiun digester dan *screw press* akan di tekan turun berada antara 2–3%, sebagai upaya meminimalisir *oil losses*.

Klasifikasi produk yang akurat sangat krusial dalam perdagangan internasional, dimana penerapan dan evaluasi beberapa algoritma untuk mengklasifikasikan produk pertanian dan pangan secara otomatis berdasarkan deskripsi teks dapat bersumber dari berbagai sumber [3]. Klasifikasi tingkat kehilangan minyak menjadi penting dalam menentukan langkah-langkah perbaikan yang efektif. Oleh karena itu diperlukan suatu pendekatan analitis yang mampu mengidentifikasi pola dan hubungan antar parameter terhadap kehilangan minyak. *Decision tree* (pohon keputusan) adalah algoritma klasifikasi yang secara otomatis menurunkan hierarki aturan pemisahan (partition rules) terhadap atribut target dari suatu himpunan data berukuran besar [4]. Alasan mengapa teknologi *decision tree* (pohon keputusan) begitu populer adalah karena pembentukannya tidak memerlukan pengetahuan domain (domain knowledge) khusus maupun pengaturan parameter, sehingga sangat cocok untuk penemuan pengetahuan eksploratif (exploratory knowledge discovery). *Decision tree* juga mampu menangani data berdimensi tinggi. Pengetahuan yang diperoleh dan direpresentasikan dalam bentuk pohon juga bersifat intuitif serta mudah dipahami [5]. Penggunaan *Decision tree* dapat dilihat pada berbagai obek. Seperti menganalisis data hasil tes fisik dan memetakan tingkat/indikator kesehatan fisik secara otomatis [6], menyaring (screening) dan memprediksi tingkat kelayakan/keberhasilan proyek wirausaha sejak dini [7], mendeteksi serangan False Data Injection (FDIA) pada smart grid [8]. Penggunaan Decision Tree dalam penelitian ini didasari oleh efisiensi serta transparansi model yang ditawarkannya. Secara visual, struktur pohon keputusan sangat intuitif dan mudah dipahami, sehingga proses pengambilan keputusannya dapat dilacak dengan jelas. Keunggulan lain dari metode ini terletak pada kepraktisan

 ISSN: 2830-0408

pengolahan datanya; Decision Tree tidak menuntut pra-pemrosesan yang rumit seperti normalisasi data, manipulasi variabel dummy, atau pembersihan nilai kosong. Selain itu, proses prediksinya relatif cepat karena memiliki kompleksitas komputasi yang efisien. Metode ini juga sangat fleksibel dalam memproses tipe data numerik maupun kategorikal, sanggup menangani kasus multi-output, dan beroperasi sebagai white box model yang kriteria keputusannya dapat dijelaskan secara transparan melalui logika Boolean. Ketahanan algoritma ini makin teruji karena mampu mempertahankan performanya dengan baik meskipun kondisi dataset tidak memenuhi asumsi-asumsi dasar secara sempurna [9].

Untuk menjawab fenomena di atas, penelitian ini difokuskan pada pemodelan klasifikasi level *oil losses* CPO pada stasiun digester dan screw press memanfaatkan metode *Decision Tree*. Analisis dilakukan terhadap data variabel proses berdasarkan atribut-atribut yang relevan. Hasil klasifikasi ini dapat membantu tim teknis pada unit usaha mengevaluasi letak kelemahan proses pengolahan, sehingga strategi perbaikan mesin maupun parameter operasional dapat dirumuskan secara lebih efektif.

2. Tinjauan Pustaka

A. Oil Losses

Kehilangan minyak (oil loss) merupakan parameter krusial yang dapat merepresentasikan tingkat profitabilitas suatu pabrik kelapa sawit. Marjin keuntungan pabrik kelapa sawit utamanya bergantung pada tingkat ekstraksi minyak dan inti sawit (kernel). Oleh karena itu, penurunan faktor ini dapat berdampak buruk pada kekuatan finansial pabrik [2] serta operasional pabrik kelapa sawit [10]. Proses penggilingan minyak kelapa sawit konvensional dikenal sebagai penyebab terjadinya kehilangan minyak (oil losses) dalam proses produksi minyak kelapa sawit[11], [12]. Sumber kehilangan minyak di pabrik kelapa sawit, yaitu tandan kosong setelah pemipilan, serat ampas perasan (press cake fibre) dari screw press, kondensat sterilisasi, lumpur (sludge) dari separator, dan tandan yang tidak terpipil (unstripped bunches / USB) setelah pemipilan. Selain itu, kehilangan minyak pada aliran limbah (waste streams) yang diwakili oleh minyak yang tertinggal atau minyak residu juga dapat berkontribusi pada penurunan pendapatan CPO [2]. Atau kehilangan dapat terjadi pada erjadi di berbagai stasiun pengolahan, termasuk stasiun sterilisasi, pemipilan (thresher), kempa (press station), dan unit klarifikasi [10].

Proses pada digester dan screw press dapat dijelaskan sebagai berikut [13]:

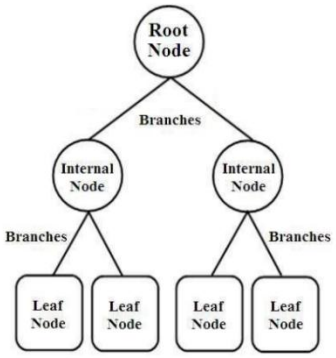
1. Proses di digester memerlukan injeksi uap (steam) dari ketel uap (boiler) untuk memanaskan dan melunakkan daging buah, yang dapat mempercepat waktu pelunakan dan mempermudah pelepasan sel-sel minyak dari mesokarp, serta pemisahan mesokarp dari biji (nut).
2. Selanjutnya, bahan-bahan dimasukkan ke dalam screw press untuk diproses secara mekanis melalui proses pengempaannya. Selama beroperasi,

B. Decision Tree

Pohon keputusan yang memprediksi variabel nominal atau numerik masing-masing disebut pohon klasifikasi (classification trees) dan pohon regresi (regression trees). Karakteristik penting dari pohon keputusan dalam konteks pembelajaran mesin (machine learning) adalah bahwa prediksinya merupakan hasil dari komputasi yang sederhana dan mudah diinterpretasikan (serangkaian pengujian yang relatif pendek). Oleh karena itu, pohon keputusan dikatakan dapat diinterpretasikan (interpretable) [14].

 ISSN: 2830-0408

Prosedur pembelajaran pohon keputusan melibatkan serangkaian langkah di mana data dibagi menjadi subset yang homogen, seperti yang ditunjukkan pada Gambar 1 yang merupakan titik awal dari pohon, mewakili seluruh set data. Algoritma mengidentifikasi fitur dan nilai ambang batas (threshold) yang menghasilkan pemisahan terbaik berdasarkan kriteria tertentu. Proses ini berlanjut secara rekursif, di mana setiap subset data dibagi lebih lanjut pada setiap *child node*. Hal ini terus berlanjut hingga kriteria penghentian tercapai, biasanya ketika simpul-simpul tersebut sudah murni (artinya, semua titik data dalam suatu simpul memiliki kelas yang sama) atau ketika kedalaman pohon yang telah ditentukan sebelumnya tercapai. Simpul tempat pohon berakhir, yang disebut *leaf node* atau *terminal node*, mewakili hasil atau label kelas. Keputusan untuk melakukan pemisahan pada setiap *node* dibuat menggunakan formulasi matematis seperti *information gain*, *Gini impurity*, atau *variance reduction*. [15], [16], [17].



Gambar 1. Contoh Pohon Keputusan
(Sumber:)[15]

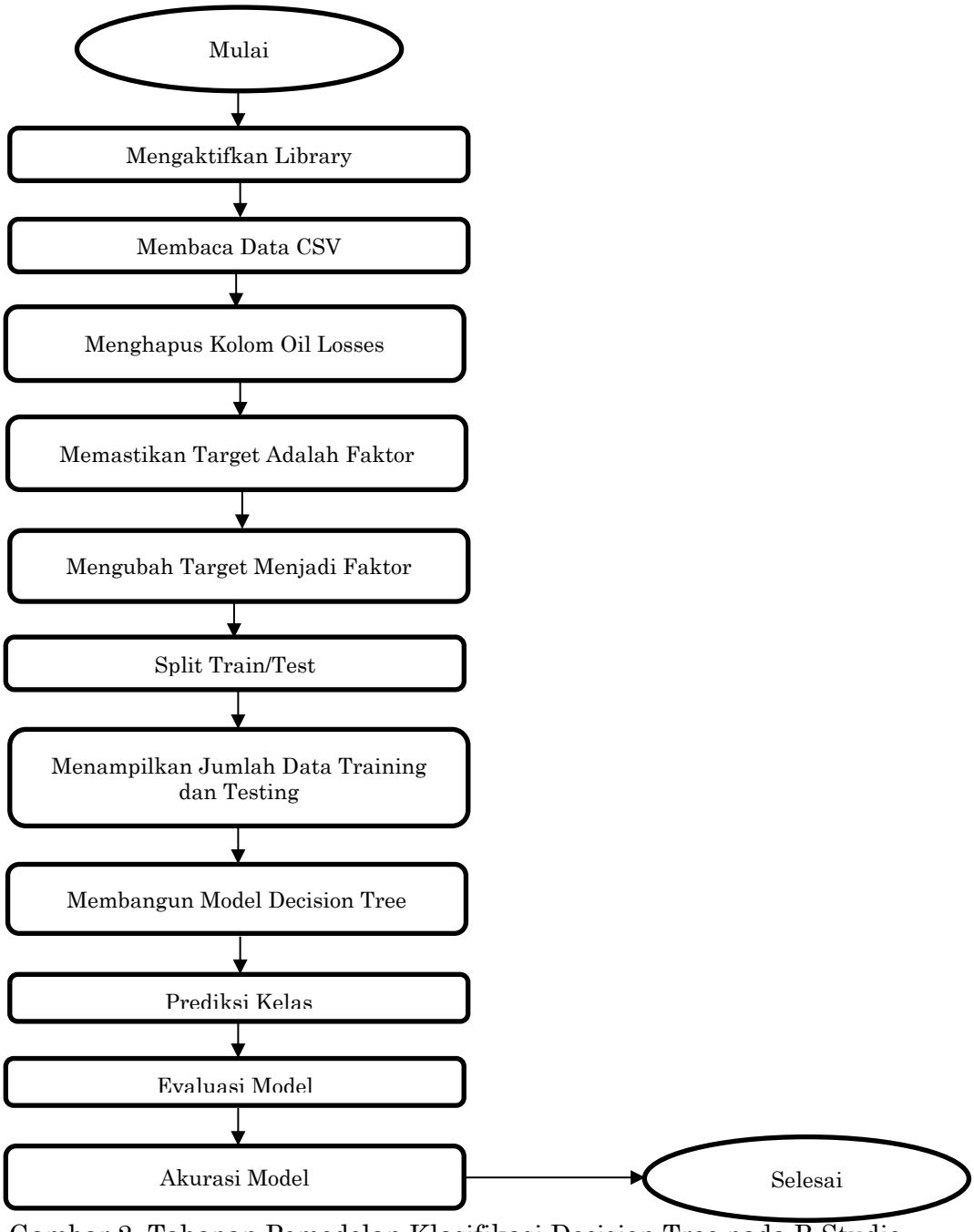
Decision Tree memiliki sejumlah keunggulan utama yang membuatnya banyak diterapkan dalam analisis data. Selain strukturnya yang relatif mudah dipahami, metode ini mampu ditransformasikan menjadi sekumpulan aturan operasional secara cepat. Decision Tree juga fleksibel dalam memproses luaran berbentuk kategorikal maupun numerik, meskipun atribut keluaran akhirnya tetap harus berupa variabel kategorikal. Kelebihan lainnya terletak pada kepraktisannya, di mana analisis dapat dilakukan tanpa memerlukan asumsi awal (a priori) terkait kualitas atau keakuratan hasil yang diperoleh.[9], [16].

Implementasi algoritma *Decision Tree* sangat bergantung pada media komputasi yang efisien dalam mengolah data dan menghasilkan visualisasi model. Dalam konteks ini, RStudio hadir sebagai lingkungan komputasi statistik yang menyediakan berbagai pustaka (*package*) khusus untuk mengeksekusi *Decision Tree*. Melalui RStudio, pemisahan *node* berdasarkan variabel parameter proses dapat dilakukan secara presisi, sekaligus menghasilkan diagram pohon keputusan yang intuitif dan mudah ditranslasikan menjadi aturan keputusan (*rules*) [18].

3. Metode Penelitian

Untuk mengklasifikasikan *level oil losses* pada produksi CPO, digunakan data numerik kontinu yang mencakup variabel proses seperti suhu, waktu operasional, tekanan, kecepatan pengadukan *digester* dan *screw press*, serta persentase oil losses. Sebanyak 117 data pengamatan berhasil dikumpulkan melalui logsheet manual sepanjang November 2025 sampai Maret 2026. Rangkaian data ini berfungsi sebagai dasar pemodelan untuk mengidentifikasi tingkat kehilangannya. Berikut tahapan pemodelan klasifikasi Decision Tree pada R Studio yang dapat dilihat pada gambar 2.

 ISSN: 2830-0408



Gambar 2. Tahapan Pemodelan Klasifikasi Decision Tree pada R Studio

Teknik pengumpulan data pada penelitian ini mencakup dua tahapan utama, yakni studi pustaka dan pengumpulan data operasional. Studi pustaka difokuskan pada pencarian referensi acuan yang mendukung metodologi penelitian, sedangkan pengumpulan data operasional dilakukan dengan mengompilasi rekam logsheet harian pabrik. Data parameter proses dan persentase oil losses pada unit Digester serta Screw Press inilah yang nantinya diolah untuk proses klasifikasi

 ISSN: 2830-0408

4. Hasil dan Pembahasan

Berdasarkan hasil pengamatan dan analisis yang dilakukan pada salah satu unit usaha pengolahan kelapa sawit di Sumatera Utara, diperoleh data berupa data numerik kontinu yang merepresentasikan *Oil Losses* (kehilangan minyak) pada proses produksi *Crude Palm Oil* (CPO), khususnya pada stasiun *Digester* dan *Screw press*. Data diperoleh dari pencatatan logsheet manual pada periode November 2025 sampai dengan Maret 2026 dengan total sebanyak 117 data pengamatan.

Data yang dikumpulkan digunakan sebagai dasar dalam proses analisis untuk melihat tingkat kehilangan minyak yang terjadi pada proses produksi. Berikut sampel data awal sebanyak 15 data pertama yang digunakan sebagai representasi awal sebelum dilakukan pengolahan dan analisis lebih mendalam.

Tabel 1. Logsheet Oil Losses Digester dan Screw press

Waktu Proses (menit)	Suhu (˚C)	Tekanan (Mbar)	Kecepatan Pengadukan (RPM)	Oil Losses	Status
46	91	35	25	2.65	Level 1
33	90	36	26	2.78	Level 1
31	92	35	26	2.88	Level 1
40	92	36	26	2.89	Level 1
31	95	35	26	2.90	Level 1
38	93	37	25	3.15	Level 2
43	92	35	26	3.10	Level 2
38	90	35	25	2.68	Level 1
39	91	37	25	2.98	Level 2
32	92	39	26	2.78	Level 1
32	93	38	25	2.75	Level 1
36	95	40	26	3.15	Level 2
36	95	38	26	3.06	Level 2
47	91	36	25	2.73	Level 1

Selanjutnya, file data Excel dikonversikan ke dalam format *comma separated values* (CSV) sebanyak 117 baris data (*records*), dengan empat atribut proses *Digester* dan *Screw press*, yaitu: waktu proses, suhu, tekanan, dan, kecepatan pengadukan.

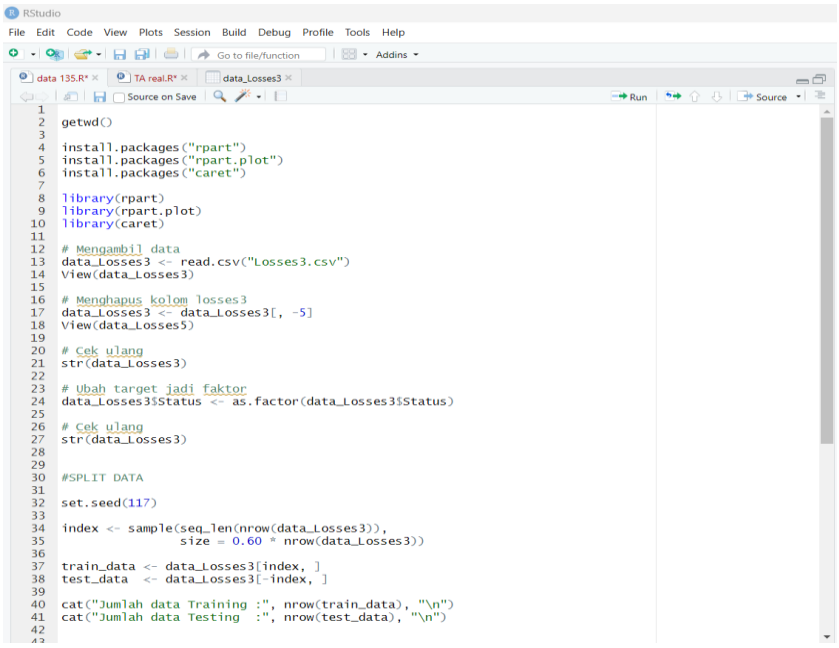
Kolom status berupa nilai losses yang diperoleh dari hasil proses *Digester* dan *Screw press* yang bersifat numerik (*kontinu*) kemudian dikonversikan menjadi data diskret (kategori). Lalu menghitung nilai rata-rata dari seluruh data Oil Losses sebesar 2,91. Nilai rata-rata ini selanjutnya dijadikan sebagai ambang batas untuk membedakan dua kategori level. Sampel dengan nilai *Oil Losses* ≥ 2,91 diklasifikasikan sebagai *Level* 2 yang merepresentasikan tingkat *Oil Losses* lebih tinggi, sedangkan sampel dengan nilai klorida < 2,91 dikategorikan sebagai Level 1 yang menunjukkan tingkat *Oil Losses* lebih rendah.

 ISSN: 2830-0408

Tabel 2. Dala Nilai Oil Losses

Oil Losses	Status
2.65	Level 1
2.78	Level 1
2.88	Level 1
2.89	Level 1
2.90	Level 1
3.15	Level 2
3.10	Level 2
2.68	Level 1
2.98	Level 2
2.78	Level 1
2.75	Level 1
3.15	Level 2
3.06	Level 2
2.73	Level 1
…	…
2.91	Rata-rata

Berikut merupakan tahapan-tahapan yang dilakukan dalam proses pembangunan model Decision Tree menggunakan bahasa pemrograman R pada perangkat lunak R Studio adalah sebagai berikut:



(a)

 ISSN: 2830-0408


```
38
39 # Membangun model Decision Tree
40 tree_model <- rpart(Status ~ ., data = train_data, method = "class")
41
42 # Menampilkan struktur Decision Tree
43 rpart.plot(tree_model, box.palette = "Red" )
44
45 # Prediksi kelas
46 prediksi <- predict(tree_model, test_data, type = "class")
47 prediksi
48
49 # Evaluasi model Confusion Matrix
50 library(caret)
51 cm <- confusionMatrix(prediksi, test_data$Status)
52 cm
53
54 # Tampilkan hasil
55 print(cm)
56
57 akurasi <- cm$overall["Accuracy"]
58 cat("\nAkurasi Model :", round(akurasi * 100, 2), "%\n")
59 print(paste("Akurasi Model:", round(akurasi * 100, 2), "%"))
60 ''
```

(b)

Gambar 3. (a) Program Dataset Awal, (b) Program Dataset Lanjutan

Berdasarkan gambar 2, pengolahan dataset kualitas CPO pada proses Digester & Screw press menggunakan aplikasi Rstudio mendapatkan hasil run sebagai berikut :

1. Mengaktifkan Libray

   Mengaktifkan library merupakan tahapan penting yang dilakukan untuk memanggil kumpulan fungsi, perintah, serta alat bantu tambahan agar dapat digunakan dalam lingkungan kerja (workspace).

```
library(rpart)
library(rpart.plot)
library(caret)
```

☑ caret	Classification and Regression Training	7.0-1	
☑ rpart	Recursive Partitioning and Regression Trees	4.1.27	
☑ rpart.plot	Plot 'rpart' Models: An Enhanced Version of	3.1.4	

Gambar 4. Mengaktifkan library

   Mengaktifan library caret (Classification and Regression Training), merupakan paket utama dalam R untuk machine learning yang digunakan untuk preprocessing data, membangun model, melakukan validasi silang, dan mengevaluasi performa model, library rpart (Recursive Partitioning and Regression Trees) untuk membangun model Decision Tree dan library rpart. plot digunakan untuk memvisualisasikan pohon keputusan yang telah dibangun dengan rpart, dalam bentuk diagram pohon yang mudah dipahami. Library ini digunakan sebagai alat bantu dalam proses klasifikasi dan visualisasi pohon keputusan.

2. Membaca Data Csv

   Membaca file Csv (Comma-Separated Values) dilakukan untuk mengambil data dari sumber eksternal dan memasukkannya ke dalam program, sehingga data tersebut siap untuk diproses dan dianalisis. Pengolahan data menggunakan software RStudio dengan bantuan tiga library, yaitu caret, rpart, dan rpart.plot. Library tersebut digunakan untuk membangun serta menampilkan model Decision Tree. Sehingga hasil pembacaan data pada program view, diperoleh dataset kualitas dalam bentuk file CSV dengan jumlah sebanyak 117 data, yang kemudian digunakan sebagai data input dalam proses analisis.

 ISSN: 2830-0408


```
> # Mengambil data
> data_Losses <- read.csv("Losses.csv")
> View(data_Losses)
```

	Waktu.Proses.menit.	Suhu...C...	Tekanan.Mbar.	Kecepatan.Pengadukan..RPM..	Oil.Losses	Status
1	46	91	35	25	2.65	Level 1
2	33	90	36	26	2.78	Level 1
3	31	92	35	26	2.88	Level 1
4	40	92	36	26	2.89	Level 1
5	31	95	35	26	2.90	Level 1
6	38	93	37	25	3.15	Level 2
7	43	92	35	26	3.10	Level 2
8	38	90	35	25	2.68	Level 1
9	39	91	37	25	2.98	Level 2
10	32	92	39	26	2.78	Level 1
11	32	93	38	25	2.75	Level 1
12	36	95	40	26	3.15	Level 2
13	36	95	38	26	3.06	Level 2
14	47	91	36	25	2.73	Level 1
15	30	91	38	25	2.93	Level 2

Gambar 5. data CSV

Parameter terdiri dari waktu proses, suhu, tekanan, dan kecepatan pengadukan, kemudian

a. Fungsi read.csv digunakan untuk membaca file CSV dari direktori lokal di folder (C:/KTA/ dengan nama file, Losses)
b. Fungsi View pada RStudio digunakan untuk menampilkan isi data dalam bentuk lembar kerja (spreadsheet) sehingga memudahkan pengguna dalam melihat dan mengecek data secara lebih jelas.

Tahap ini penting karena data tersebut menjadi input utama dalam proses analisis untuk melihat hubungan terhadap oil losses.

3. Menghapus Kolom *Oil Losses*

Menghapusan kolom "*Oil Losses*" dilakukan untuk menyederhanakan data dengan menghilangkan parameteryang tidak digunakan lagi dalam proses analisis.

```
> # Menghapus kolom Losses
> data_Losses <- data_Losses[, -5]
> View(data_Losses)
```

	Waktu.Proses.menit.	Suhu...C...	Tekanan.Mbar.	Kecepatan.Pengadukan..RPM..	Status
1	46	91	35	25	Level 1
2	33	90	36	26	Level 1
3	31	92	35	26	Level 1
4	40	92	36	26	Level 1
5	31	95	35	26	Level 1
6	38	93	37	25	Level 2
7	43	92	35	26	Level 2
8	38	90	35	25	Level 1
9	39	91	37	25	Level 2
10	32	92	39	26	Level 1
11	32	93	38	25	Level 1
12	36	95	40	26	Level 2
13	36	95	38	26	Level 2
14	47	91	36	25	Level 1
15	30	91	38	25	Level 2

Gambar 6. Hasil CSV Menghapus Kolom Oil Losses

Menghapusan kolom Oil Losses dilakukan karena data pada kolom tersebut tidak digunakan dalam proses analisis. informasi pada kolom tersebut tetap dimanfaatkan sebagai acuan dalam penentuan kategori level 1 dan level 2.

4. Cek Struktur ParameterTarget adalah Faktor untuk Klasifikasi

Memastikan parameter target memiliki tipe faktor dilakukan agar algoritma klasifikasi dapat mengenali nilai kategori sebagai label kelas yang digunakan dalam proses prediksi dan analisis.

 ISSN: 2830-0408


```
> # Cek ulang
> str(data_Losses3)
'data.frame':   117 obs. of  5 variables:
 $ Waktu.proses: int  46 33 31 40 31 38 43 38 39 32 ...
 $ Suhu        : int  91 90 92 92 95 93 92 90 91 92 ...
 $ Tekanan     : int  35 36 35 36 35 37 35 35 37 39 ...
 $ Kecepatan   : int  25 26 26 26 26 25 26 25 25 26 ...
 $ Status      : Factor w/ 2 levels "Level-dua","Level-satu": 2 2 2 2 2 1 1 2 1 2 ...
```

Gambar 7. Cek Struktur ParameterTarget adalah Faktor untuk Klasifikasi

Pada cek struktur, perlakukan terdiri dari:

a. Melakukan perintah str(data_Losses), digunakan untuk melihat struktur data secara ringkas.
b. Mengetahui bahwa dataset memiliki 117 data dan 5 variabel yang terdiri dari : Waktu Proses, Suhu, Tekanan, dan Kecepatan Pengadukan sudah bertipe numerik (integer), sehingga sudah sesuai untuk analisis.
c. Parameter Status masih bertipe karakter (chr), padahal parameter ini merupakan target klasifikasi. perlu dilakukan perubahan tipe data menjadi faktor agar bisa digunakan dalam metode Decision Tree.

5. Split train/test (Memisahkan Data menjadi Dua Bagian)

Pemisahan data menjadi data latih (train) dan data uji (test) dilakukan agar model dapat dibentuk menggunakan data latih, kemudian kinerjanya dievaluasi menggunakan data uji untuk mengetahui tingkat kemampuan model dalam melakukan prediksi.

```
> set.seed(117)
> index <- sample(seq_len(nrow(data_Losses3)),
+                 size = 0.60 * nrow(data_Losses3))
> train_data <- data_Losses3[index, ]
> test_data  <- data_Losses3[-index, ]
> cat("Jumlah data Training :", nrow(train_data), "\n")
```

Gambar 8. Hail run Split train/test

Split train/test terdiri dari :

a. Fungsi set.seed (117) untuk memastikan bahwa hasil pengambilan data tetap konsisten. Dengan pembagiannya 60% Training dan 40% sebagai data testing
b. Hasil pengambilan Indeks tersebut disimpan dalam parameterindex, kemudian digunakan untuk memisahkan dataset menjadi train_data dan test_data. Tahap ini penting sebelum ingin membangun model Decision Tree agar model dapat Training dan testing secara adil.

Dalam memilih proposi pembagian data dilakukan berdasarkan akurasi tertinggi, dari hasil tes akurasi proposi 0,60 memiliki tingkat akurasi yaitu 57% yang paling tinggi dibandingkan dengan proposi yang lain seperti 0,65 menghasilkan akurasi 0,41%, proporsi 0,70 sebesar 36%, proporsi 0,75 sebesar 46%, proporsi 0,80 sebesar 41%, proporsi dari hasil tersebut terlihat bahwa tidak semua penambahan jumlah data training akan meningkatkan akurasi model. Jika jumlah data training terlalu banyak, maka data testing akan menjadi lebih sedikit. Akibatnya, proses pengujian model menjadi kurang optimal karena data yang digunakan untuk mengevaluasi model tidak cukup mewakili kondisi sebenarnya. Menurut An (2021), pada dataset yang berukuran kecil, jumlah data testing yang terlalu sedikit dapat menyebabkan hasil evaluasi model menjadi kurang stabil dan kurang dapat dipercaya. Oleh karena itu, diperlukan pembagian data yang seimbang agar model dapat belajar dengan baik dari data training sekaligus memperoleh hasil evaluasi yang lebih representatif dari data testing.

 ISSN: 2830-0408

Penggunaan proporsi 60:40, dianggap paling optimal karena memberikan keseimbangan antara kemampuan model dalam belajar dari data training dan keakuratan dalam proses evaluasi menggunakan data testing.

Tabel 1. Data *Split train/test*

Proposi	Akurasi
0,6	0,57
0,65	0,41
0,70	0,3611
0,75	0,4667
0,80	0,4167

6. Menapilkan Jumlah Data training dan testing
   Menampilkan jumlah data training dan testing dilakukan untuk memastikan pembagian data telah sesuai dengan proporsi yang ditentukan sebelum proses pemodelan dilakukan.

```
> cat("Jumlah data Training :", nrow(train_data), "\n")
Jumlah data Training : 70
> cat("Jumlah data Testing  :", nrow(test_data), "\n")
Jumlah data Testing  : 47
```

Gambar 9. Menampilkan Jumlah Data dalam Training dan Testing

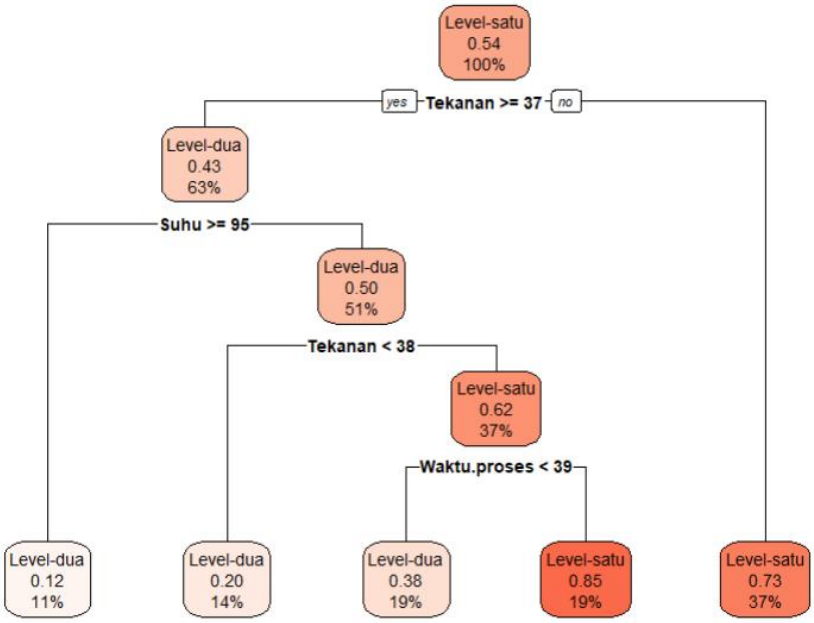
   Hasil dari RStudio, diketahui bahwa jumlah data training sebanyak 70 data, sedangkan data testing sebanyak 47 data. Jumlah tersebut sesuai dengan proporsi pembagian 60% untuk data latih dan 40% untuk data uji dari total 117 data yang digunakan dalam penelitian.
7. Membangun model Decision Tree
   Pembangunan model Decision Tree dilakukan untuk menghasilkan model yang dapat mengelompokkan data ke dalam kategori tertentu berdasarkan pola yang diperoleh dari data latih.

```
> # Membangun model Decision Tree
> tree_model <- rpart(Status ~ ., data = train_data, method = "class")
```

Gambar 10. Hasil run Membagun Model Decision Tree

   Dalam pemodelan ini, ParameterStatus berperan sebagai ParameterTarget yang diprediksi berdasarkan seluruh variable model yang terdapat dalam data training dilakukan menggunakan (dataTrain) method = "class", yang menunjukkan bahwa model yang dibangun bertujuan untuk di klasifikasi. Hal ini sesuai dengan tujuan penelitian, yaitu mengelompokkan data ke dalam beberapa kategori level oil losses. Hasil model yang terbentuk akan disimpan dalam objek tree_model dan selanjutnya digunakan untuk melakukan prediksi pada data testing dan mengevaluasi kinerja model.
8. Menampilkan struktur Decision Tree
   Pembuatan model Decsion Tree dilakukan dengan membentuk struktur pohon keputusan yang membagi data berdasarkan atribut tertentu secara bertahap. Setiap cabang menunjukkan hasil dari suatu kondisi, sedangkan bagian akhir pohon menunjukkan kelas atau kategori hasil prediksi.

 ISSN: 2830-0408



Gambar 11. Hasil run struktur Decision Tree

Berdasarkan hasil pengolahan 117 data menggunakan metode Decision Tree, diperoleh pohon keputusan yang menunjukkan hubungan antara parameter tekanan, suhu, dan waktu proses terhadap klasifikasi status oil losses.

Pada node akar terdapat nilai Level-satu, 0,54, dan 100%. angka 100% menunjukkan bahwa seluruh data yang digunakan dalam pembentukan model berjumlah 117 data berada pada node tersebut. Nilai 0,54 menunjukkan bahwa sekitar 54% data termasuk ke dalam kelas Level-satu, sehingga node awal diklasifikasikan sebagai Level-satu. Warna oranye pada node menunjukkan bahwa kelas Level-satu lebih dominan dibandingkan Level-dua.

Pemisahan pertama dilakukan berdasarkan parameter Tekanan ≥ 37 mbar. Data yang memiliki tekanan kurang dari 37 mbar langsung diklasifikasikan sebagai Level-satu dengan nilai persentase data 0,73 dan mencakup sekitar 37% dari total data. Nilai tersebut menunjukkan bahwa sekitar 73% data pada kelompok ini termasuk ke dalam kelas Level-satu sehingga warna node menjadi lebih gelap karena tingkat dominasinya tinggi. Sementara itu, sekitar 63% data memiliki tekanan ≥ 37 mbar akan masuk ke proses pemisahan berikutnya menggunakan parameter suhu. Pada kelompok ini, node diklasifikasikan sebagai Level-dua dengan nilai Persentase data 0,43. Hal ini menunjukkan bahwa proporsi data Level-satu hanya sekitar 43%, sehingga mayoritas data pada kelompok tersebut termasuk Level-dua.

Pemisahan selanjutnya dilakukan berdasarkan Suhu ≥ 95°C. Data yang memiliki suhu ≥ 95°C langsung diklasifikasikan sebagai Level-dua dengan nilai persentase data 0,12 dan mencakup sekitar 11% dari total data. Nilai ini menunjukkan bahwa sekitar 88% data pada kelompok tersebut termasuk ke dalam kelas Level-dua sehingga node berwarna sangat pucat.

Untuk data dengan suhu < 95°C, model kembali melakukan pemisahan menggunakan parameter Tekanan < 38 mbar. Pada kondisi tekanan < 38 mbar, data langsung diklasifikasikan sebagai Level-dua dengan persentase data 0,20 dan mencakup sekitar 14% dari total data. Artinya, sekitar 80% data pada kelompok ini termasuk Level-dua. Apabila tekanan ≥ 38 mbar, maka data masuk ke node

 ISSN: 2830-0408

berikutnya yang diklasifikasikan sebagai Level-satu dengan persentase data 0,62 dan mencakup sekitar 37% dari total data. Hal ini menunjukkan bahwa sekitar 62% data pada kelompok tersebut termasuk Level-satu. Pada kelompok terakhir tersebut dilakukan pemisahan menggunakan parameter waktu Proses < 39 menit. Jika waktu proses kurang dari 39 menit, data diklasifikasikan sebagai Level-dua dengan Persentase data 0,38 dan mencakup sekitar 19% dari total data. Sebaliknya, jika waktu proses ≥ 39 menit, data diklasifikasikan sebagai Level-satu dengan Persentase data 0,85 dan mencakup sekitar 19% dari total data. Nilai Persentase data yang tinggi menunjukkan bahwa sekitar 85% data pada kelompok ini termasuk Level-satu sehingga node memiliki warna merah paling gelap.

Secara keseluruhan, pohon keputusan menunjukkan bahwa tekanan merupakan parameter yang paling dominan dalam menentukan klasifikasi status oil losses, diikuti oleh suhu dan waktu proses. Hal ini terlihat dari posisi tekanan sebagai akar pohon keputusan yang digunakan pada pemisahan pertama data. Hal ini didukung juga oleh penelitian terdahulu, yang menyatakan bahwa tekanan memiliki peran dominan dalam peningkatan oil losses pada *screw press* [19], [20].

9. Prediksi Kelas (Berdasarkan Data testing)

Prediksi kelas berdasarkan data testing bertujuan untuk mengetahui kemampuan model dalam mengelompokkan data baru yang belum pernah digunakan sebelumnya ke dalam kategori yang sesuai.

```
> # Prediksi kelas
> prediksi <- predict(tree_model, test_data, type = "class")
> prediksi
```

Gambar 12. Hasil run Prediksi Kelas

fungsi predict digunakan untuk memprediksi kelas Status pada data uji (dataTest) berdasarkan model yang tersimpan di objek tree_model. Parameter type = "class" menunjukkan hasil yang ingin ditampilkan adalah berupa kelas atau kategori, Jadi output yang dihasilkan langsung berupa label seperti Level 1 atau Level 2, sesuai dengan parametertarget yang digunakan. yang kemudian dapat digunakan untuk evaluasi akurasi model.

10. Evaluasi model Confusion Matrix

Menampilkan hasil confusion matrix bertujuan untuk mengetahui tingkat ketepatan hasil klasifikasi dengan melihat jumlah data yang terprediksi benar maupun salah pada setiap kategori secara rinci.

```
Confusion Matrix and Statistics

            Reference
Prediction   Level-dua Level-satu
  Level-dua         13         10
  Level-satu        10         14

               Accuracy : 0.5745
                 95% CI : (0.4218, 0.7174)
    No Information Rate : 0.5106
    P-Value [Acc > NIR] : 0.2332

                  Kappa : 0.1486

 Mcnemar's Test P-Value : 1.0000

            Sensitivity : 0.5652
            Specificity : 0.5833
         Pos Pred Value : 0.5652
         Neg Pred Value : 0.5833
             Prevalence : 0.4894
         Detection Rate : 0.2766
   Detection Prevalence : 0.4894
      Balanced Accuracy : 0.5743

       'Positive' Class : Level-dua
```

Gambar 13. Hasil run Confusion Matrix

 ISSN: 2830-0408

11. Akurasi Mode

Hasil akurasi model merupakan salah satu metrik evaluasi yang menunjukkan tingkat keberhasilan model dalam melakukan prediksi.

```
Akurasi Model : 57.45 %
> print(paste("Akurasi Model:", round(akurasi * 100, 2), "%"))
[1] "Akurasi Model: 57.45 %"
```

Gambar 14. Hasil Run Akurasi Model

Evaluasi perfoma model dilakukan menggunakan confusion matrix untuk melihat seberapa jauh model mampu membaca pola dari data uji [21], [22]. Dari hasil evaluasi yang diperoleh adalah

a. Akurasi Model. Nilai Akurasi sebesar 57,45% menunjukkan bahwa model mampu mengklasifikasikan data dengan benar sebanyak 57,45% dari seluruh data pengujian. Artinya, dari setiap 100 data yang diprediksi, sekitar 57 data berhasil diprediksi dengan benar oleh model.
b. *Confidence Interval* (CI) digunakan untuk menunjukkan perkiraan rentang nilai akurasi model jika diterapkan pada data yang lebih banyak. Nilai CI sebesar 42,18%–71,74% menunjukkan bahwa akurasi model diperkirakan berada pada rentang tersebut.
c. *No Information Rate* (NIR) digunakan sebagai pembanding untuk mengetahui apakah model lebih baik dibandingkan tebakan sederhana berdasarkan kelas yang paling banyak muncul. Nilai NIR sebesar 51,06% berarti jika semua data ditebak sebagai kelas yang paling dominan, maka tingkat ketepatannya sekitar 51,06%. Karena akurasi model hanya 57,45%, maka peningkatan yang diberikan model masih belum terlalu besar.
d. *P-Value* [Acc > NIR] digunakan untuk melihat apakah akurasi model benar-benar lebih baik dibandingkan NIR. Nilai 0,2332 lebih besar dari 0,05 sehingga belum dapat disimpulkan bahwa model secara statistik lebih baik daripada metode tebakan sederhana berdasarkan kelas mayoritas
e. *Kappa* digunakan untuk mengukur seberapa baik hasil prediksi model dibandingkan kondisi sebenarnya. Nilai Kappa sebesar 0,1486 menunjukkan bahwa tingkat kecocokan antara hasil prediksi dan data aktual masih rendah. Hal ini menunjukkan bahwa kemampuan model dalam membedakan Level-satu dan Level-dua masih perlu ditingkatkan.
f. *McNemar's Test* digunakan untuk melihat apakah model lebih sering salah pada salah satu kelas tertentu. Nilai 1,0000 menunjukkan bahwa kesalahan prediksi model pada kelas Level-satu dan Level-dua relatif seimbang. Dengan kata lain, model tidak cenderung lebih banyak salah pada salah satu kelas.
g. *Sensitivity* (Recall) digunakan untuk mengukur kemampuan model dalam mengenali data Level-dua dengan benar. Nilai 56,52% menunjukkan bahwa dari seluruh data yang benar-benar termasuk Level-dua, sekitar 56,52% berhasil dikenali dengan benar oleh model.
h. *Specificity* digunakan untuk mengukur kemampuan model dalam mengenali data Level-satu dengan benar. Nilai 58,33% menunjukkan bahwa dari seluruh data yang benar-benar termasuk Level-satu, sekitar 58,33% berhasil dikenali dengan benar oleh model.
i. *Negative Predictive Value* (NPV) digunakan untuk melihat ketepatan prediksi ketika model menyatakan suatu data termasuk Level-satu. Nilai 58,33%

 ISSN: 2830-0408

menunjukkan bahwa ketika model memprediksi data sebagai Level-satu, maka kemungkinan prediksi tersebut benar adalah sebesar 58,33%.

j. *Positive Predictive Value* (PPV) atau Precision digunakan untuk melihat ketepatan prediksi ketika model menyatakan suatu data termasuk Level-dua. Nilai 56,52% menunjukkan bahwa ketika model memprediksi data sebagai Level-dua, maka kemungkinan prediksi tersebut benar adalah sebesar 56,52%.
k. *Prevalence* digunakan untuk menunjukkan jumlah data Level-dua dalam data pengujian.Nilai 48,94% menunjukkan bahwa sekitar 48,94% dari seluruh data pengujian merupakan data Level-dua.
l. *Detection Rate* digunakan untuk menunjukkan seberapa banyak data Level-dua yang berhasil ditemukan dengan benar oleh model. Nilai 27,66% menunjukkan bahwa model berhasil mengenali sekitar 27,66% data Level-dua dari seluruh data pengujian.
m. *Detection Prevalence* digunakan untuk menunjukkan seberapa banyak data yang diprediksi sebagai Level-dua oleh model.Nilai 48,94% menunjukkan bahwa sekitar 48,94% dari seluruh data diprediksi sebagai Level-dua.
n. *Balanced Accuracy* digunakan untuk melihat kemampuan model dalam mengenali kedua kelas (Level-satu dan Level-dua) secara seimbang.

Berdasarkan hasil *Confusion Matrix*, model *Decision Tree* memperoleh *accuracy* sebesar 57,45%, yang menunjukkan bahwa model mampu mengklasifikasikan 27 dari 47 data pengujian dengan benar. Nilai *Sensitivity* sebesar 56,52% dan *Specificity* sebesar 58,33% menunjukkan bahwa kemampuan model dalam mengenali kelas Level-satu dan Level-dua relatif seimbang. Namun, nilai Kappa sebesar 0,1486 menunjukkan bahwa tingkat kesesuaian antara hasil prediksi model dan data aktual masih rendah. Oleh karena itu, model telah mampu mengenali pola dalam data, tetapi performanya masih perlu ditingkatkan agar hasil prediksi menjadi lebih akurat dan lebih andal dalam mengklasifikasikan tingkat oil losses.

5. Kesimpulan

Berdasarkan hasil pohon keputusan Parameter proses yang mempengaruhi tingkatan oil losses pada proses Digester dan Screw press adalah tekanan, suhu, dan waktu proses. Ketiga parameter tersebut mempengaruhi proses. Oleh karena itu, pengendalian tekanan, suhu, dan waktu proses secara optimal perlu dilakukan untuk membantu menekan tingkat *oil losses* pada proses produksi CPO. Bredasarkan struktur pohon keputusan, parameter yang paling mempengaruhi tingkatan oil losses adalah tekanan. Hal ini ditunjukkan oleh posisi tekanan sebagai node akar (*root node*) atau percabangan pertama pada pohon keputusan. Setelah tekanan, parameter yang turut mempengaruhi adalah suhu dan waktu proses. dari ketiga parameter yang harus lebih diperhatikan yaitu tekanan.

Guna meningkatkan ketepatan model klasifikasi pada penelitian selanjutnya, dapat digunakan himpunan data yang lebih komprehensif dengan rentang waktu pengamatan yang lebih panjang. Di samping itu, pengembangan model selanjutnya disarankan untuk mengintegrasikan variabel kualitatif dan teknis tambahan, seperti mutu Tandan Buah Segar (TBS), kadar air, tingkat kematangan, hingga kondisi efisiensi mekanis peralata

 ISSN: 2830-0408

Pustaka

[1] N. Yuselin, E. Rosyidi, and E. Adetya, "Efforts to reduce oil losses at press stations from 6.91% to 6% in crude palm oil (CPO) processing at PT. ABC," in *Proceedings ASTRAtech International Conference*, 2025, pp. 64–74.
[2] U. K. H. M. Nadzim, R. Yunus, R. Omar, and B. Y. Lim, "Factors contributing to oil losses in crude palm oil production process in Malaysia: A review," *International Journal of Biomass and Renewables*, vol. 9, no. 1, pp. 10–24, 2020.
[3] I. Marra De Artiñano, F. Riottini Depetris, and C. Volpe Martincus, "Automatic Product Classification in International Trade: Machine Learning and Large Language Models," *Rev International Economics*, vol. 34, no. 1, pp. 3–19, Feb. 2026, doi: 10.1111/roie.70009.
[4] X. Li and C. Claramunt, "A Spatial Entropy-Based Decision Tree for Classification of Geographical Information," *Transactions in GIS*, vol. 10, no. 3, pp. 451–467, May 2006, doi: 10.1111/j.1467-9671.2006.01006.x.
[5] S. Li, D. Xu, Y. Liu, R. Wang, and J. Zhang, "Identification Method of Influencing Factors of Hospital Catering Service Satisfaction Based on Decision Tree Algorithm," *Applied Bionics and Biomechanics*, vol. 2022, pp. 1–11, Apr. 2022, doi: 10.1155/2022/6293908.
[6] L. Chen and M. Miao, "Application of Decision Tree Intelligent Algorithm in Data Analysis of Physical Health Test," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–7, Jan. 2022, doi: 10.1155/2022/8584377.
[7] Y. M. Wang and L. Xue, "Application of Decision Tree Algorithm in Early Entrepreneurial Project Screening," *Scientific Programming*, vol. 2022, pp. 1–9, Mar. 2022, doi: 10.1155/2022/3584196.
[8] Z. Haider, D. Sabir, L. Khan, and Z. Ullah, "A Comparative Study of Decision Tree, RNNs and CNNs for Detection of False Data Injection Attacks in Cyber-Physical Power Systems," *IET Cyber-Phy Sys Theory & Ap*, vol. 11, no. 1, p. e70047, Jan. 2026, doi: 10.1049/cps2.70047.
[9] B. Gupta, A. Rawat, A. Jain, A. Arora, and N. Dhami, "Analysis of various decision tree algorithms for classification in data mining," *International Journal of Computer Applications*, vol. 163, no. 8, pp. 15–19, 2017.
[10] G. R. Jannah, N. Annisa, M. S. T. Situmorang, R. A. Rangkuti, and others, "Analysis of Oil Losses in Press Cake Using Fault Tree Analysis (FTA) in Crude Palm Oil (CPO) Processing Industry," in *Talenta Conference Series: Energy and Engineering (EE)*, 2026, pp. 616–622.
[11] D. Irwansyah, C. I. Erliana, and W. M. Manurung, "Analisis Kehilangan Minyak (oil losses) pada Crude Palm Oil dengan Metode Statistical Process Control," in *Seminar Nasional Teknik Industri 2019*, Teknik Industri Universitas Malikussaleh, 2019.
[12] J. Yardani, A. Ulimaz, P. W. Astuti, and others, "Analisis Kehilangan Minyak (Oil Losses) Di Empty Bunch Menggunakan Alat Bunch Press Di PT. PQRS," *Jurnal Teknologi Pertanian Andalas*, vol. 29, no. 1, pp. 62–70, 2025.
[13] Zulfatri Aini, T. R. S. A. Tengku, and Sri Basriati, "Steam requirements and mass balance in digesters and screw presses at palm oil mill," *JEMMME*, vol. 9, no. 2, pp. 109–120, Feb. 2025, doi: 10.22219/jemmme.v9i2.37043.
[14] H. Blockeel, L. Devos, B. Frénay, G. Nanfack, and S. Nijssen, "Decision trees: from efficient prediction to responsible AI," *Frontiers in artificial intelligence*, vol. 6, p. 1124553, 2023.
[15] I. D. Mienye and N. Jere, "A Survey of Decision Trees: Concepts, Algorithms, and Applications," *IEEE Access*, vol. 12, pp. 86716–86727, 2024, doi: 10.1109/ACCESS.2024.3416838.
[16] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of applied science and technology trends*, vol. 2, no. 01, pp. 20–28, 2021.
[17] N. Fayza, N. Svensons, S. A. Fatmawati, K. Erviona, and others, "Analisis Perbandingan Algoritma Klasifikasi Decision Tree, K-Nearest Neighbors, Naive Bayes, dan Random Forest pada Data Pemilihan Legislatif KPU Menggunakan Kurva ROC," 2025.
[18] M. Katamaneni, G. Guttikonda, and M. Suneetha, "Implementing of decision tree algorithm using R-studio and java," in *2018 Fourth International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)*, IEEE, 2018, pp. 1–5.
[19] S. Rizal and L. Rahmawati, "Analysis Of Oil Losses at Press Station In PT. Palmina Utama," *Agrisains: Jurnal Budidaya Tanaman Perkebunan Politeknik Hasnur*, vol. 6, no. 02, pp. 37–41, 2021.
[20] R. Rosmiati, D. Kurnia, and S. Lumbantoruan, "Perhitungan Oil Losses Pada Ampas Press Di Proses Pengepresan Unit Screw Press PT. XYZ," *Jurnal Vokasi Teknik (JuVoTek)*, vol. 2, no. 1, pp. 1–5, 2024.
[21] M. Fadli and R. A. Saputra, "Klasifikasi dan evaluasi performa model Random Forest untuk prediksi stroke," *Jurnal Teknik*, vol. 12, no. 2, 2023.
[22] N. K. Negoro, E. Utami, and A. Yaqin, "Klasifikasi deteksi penggunaan masker menggunakan metode convolutional neural network," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 2, pp. 664–674, 2023.