

# Analisis Algoritma C45 dan Regresi Linear untuk Memprediksi Hasil Panen Kelapa Sawit

Nurahman, Nindi Ernawati\*

Fakultas Ilmu Komputer, Program Studi Sistem Informasi, Universitas Darwan Ali, Sampit, Indonesia

Email: <sup>1</sup>nurahman@unda.ac.id, <sup>2,\*</sup>nindiernawati@gmail.com

Email Penulis Korespondensi: nindiernawati@gmail.com

Submitted: 21/08/2024; Accepted: 31/08/2024; Published: 31/08/2024

**Abstrak**—Indonesia, sebagai salah satu produsen utama kelapa sawit di dunia, memiliki sektor pertanian yang sangat berpengaruh terhadap perekonomian nasional, khususnya melalui ekspor minyak kelapa sawit. Prediksi hasil panen kelapa sawit menjadi krusial untuk meningkatkan efisiensi dalam perencanaan dan pengelolaan sumber daya. Penelitian ini dilakukan untuk membandingkan performa dua metode prediksi, yaitu Algoritma C45 dan Regresi Linear, dalam memprediksi hasil panen kelapa sawit. Rumusan masalah yang diangkat dalam penelitian ini meliputi: (1) Bagaimana perbandingan performa Algoritma C45 dan Regresi Linear dalam memprediksi hasil panen kelapa sawit? (2) Seberapa akurat prediksi yang dihasilkan oleh kedua algoritma tersebut berdasarkan data historis hasil panen? (3) Faktor-faktor apa saja yang mempengaruhi pilihan antara Algoritma C45 dan Regresi Linear untuk prediksi hasil panen kelapa sawit?. Data yang digunakan dalam penelitian ini merupakan data historis dari PT. Surya Inti Sawit Kahuripan, yang mencakup 106 blok data. Variabel yang dianalisis meliputi luas lahan, jumlah pohon, jumlah pohon per hektar, tahun tanam, jenis tanah, rencana penggunaan pupuk, serta hasil panen dalam ton. Analisis data dilakukan dengan menggunakan Algoritma C45, yang membentuk pohon keputusan berdasarkan data historis, serta metode Regresi Linear, yang menganalisis hubungan linear antara variabel-variabel independen dan variabel dependen. Akurasi prediksi diukur menggunakan Root Mean Squared Error (RMSE). Hasil penelitian menunjukkan bahwa Algoritma C45 memiliki nilai RMSE yang lebih rendah dibandingkan Regresi Linear, menunjukkan bahwa Algoritma C45 memberikan prediksi yang lebih akurat.

**Kata Kunci:** Kelapa Sawit; Algoritma C45; Algoritma Regresi Linear; Prediksi; Data Mining; RMSE

**Abstract**—Indonesia, as one of the main producers of palm oil in the world, has an agricultural sector that is very influential on the national economy, especially through palm oil exports. Prediction of oil palm yields is crucial to improve efficiency in planning and resource management. This study was conducted to compare the performance of two prediction methods, namely the C45 Algorithm and Linear Regression, in predicting oil palm yields. The formulation of the problems raised in this study includes: (1) How does the performance of the C45 Algorithm and Linear Regression compare in predicting oil palm yields? (2) How accurate are the predictions generated by the two algorithms based on historical data on crop yields? (3) What are the factors that influence the choice between C45 Algorithm and Linear Regression for oil palm yield prediction? The data used in this study is historical data from PT. Surya Inti Sawit Kahuripan, which includes 106 data blocks. The variables analyzed included land area, number of trees, number of trees per hectare, planting year, soil type, fertilizer use plan, and yield in tons. Data analysis was carried out using the C45 Algorithm, which forms a decision tree based on historical data, and the Linear Regression method, which analyzes the linear relationship between independent variables and dependent variables. Prediction accuracy is measured using Root Mean Squared Error (RMSE). The results show that the C45 Algorithm has a lower RMSE value than Linear Regression, indicating that the C45 Algorithm provides more accurate predictions.

**Keywords:** Palm Oil; C45 Algorithm; Linear Regression Algorithm; Predictions; Data Mining; RMSE

## 1. PENDAHULUAN

Indonesia merupakan negara agraris, dan sektor pertanian merupakan sektor yang penting untuk meningkatkan penerimaan devisa negara yang mampu menghadapi perkembangan perekonomian dunia khususnya yang berkaitan dengan krisis ekonomi. Upaya mengatasi krisis ekonomi dapat dilakukan melalui pembangunan pertanian dalam arti luas, antara lain pengembangan sektor tanaman pangan, peternakan, kehutanan, perikanan, dan perkebunan yang bertujuan untuk meningkatkan kesejahteraan masyarakat setempat[1]. Kelapa sawit merupakan salah satu dari bahan baku perkebunan dan merupakan bahan baku utama dan terpenting di Indonesia, dan produk utama terdiri dari minyak sawit mentah (CPO) dan inti sawit (KPO). Saat ini, kelapa sawit sangat diminati oleh masyarakat karena nilai ekonominya yang tinggi, seperti kemampuan untuk menghasilkan minyak sawit, bahan bakar, dan bahan lain[2].

Kelapa sawit juga salah satu tanaman penghasil minyak nabati yang memiliki produktivitas tinggi serta biaya produksi yang relatif rendah. Selain itu juga, tanaman kelapa sawit mempunyai ketahanan yang lebih baik terhadap serangan hama dan penyakit dibandingkan tanaman minyak nabati lainnya. Oleh karena itu, perkebunan kelapa sawit menjadi jenis usaha jangka panjang yang menjanjikan karena dapat menghasilkan produksi minyak secara maksimal dalam waktu yang relatif lama. Produksi kelapa sawit memberikan kontribusi yang sangat besar bagi perekonomian Negara-negara tropis[3], salah satunya negara Indonesia. Tanaman Kelapa sawit (*Elaeis guineensis* Jack) berasal dari Nigeria, Afrika Barat. Namun, ada sebagian pendapat yang justru menyatakan bahwa kelapa sawit berasal dari kawasan Amerika Selatan yaitu Brazil. Hal ini karena lebih banyak ditemukan spesies kelapa sawit di hutan Brazil dibandingkan dengan di Afrika. Pada kenyataannya tanaman kelapa sawit hidup subur di luar daerah asalnya, seperti Malaysia, Indonesia, Thailand, dan Papua Nugini[4]. Bahkan Mampu memberikan hasil produksi per hektarnya yang lebih tinggi. Meskipun industri kelapa sawit memberikan kontribusi ekonomi

yang signifikan, industri kelapa sawit juga menimbulkan berbagai masalah lingkungan dan sosial. Untuk meningkatkan produktivitasnya, diperlukan analisis produktivitas untuk mengidentifikasi komponen yang mempengaruhi pertumbuhan. Karena banyaknya faktor yang mempengaruhi, analisis faktor-faktor ini tidak mudah dilakukan. Lingkungan, faktor genetik, dan teknik budidaya adalah beberapa faktor yang mempengaruhi produktivitas kelapa sawit. Dengan mempertimbangkan faktor-faktor yang mempengaruhi produktivitas tanaman, analisis ini dapat membantu dalam meningkatkan produksi dan meningkatkan kemampuan analisis[5].

Prediksi merupakan tujuan alami dari penelitian ilmiah, setidaknya sebagian besar. Namun, keduanya bergantung pada premis yang sering dilupakan atau dilewati, dengan istilah seperti predictor atau prediktif yang menjadi mode default dalam bidang pemodelan regresi tanpa memastikan bahwa semua fondasi yang relevan telah diterapkan dengan benar[6]. Ukuran sampel merupakan banyaknya sampel yang akan diambil dari suatu populasi. Jika jumlah populasinya kurang dari 100, sedangkan jika jumlah sampel diambil secara keseluruhan, tetapi jika populasinya lebih besar dari 100, maka bisa diambil 10-15% atau 20-25% dari jumlah populasinya. Maka disini peneliti akan menggunakan sampel sebesar 15% dari populasi yaitu populasi  $\times 15\% = \text{hasil}$ [7]. Perkebunan kelapa sawit memainkan peran yang sangat penting dalam ekonomi Indonesia. Namun, industri ini menghadapi banyak masalah, termasuk masalah lingkungan, sosial, dan ketidakpastian hasil panen. Untuk mengatasi masalah ini, diperlukan inovasi seperti penggunaan teknologi pertanian kontemporer, penciptaan varietas unggul, dan peningkatan praktik pertanian berkelanjutan. Penggunaan algoritma seperti regresi linier dan C45 dapat meningkatkan efisiensi produksi dan membantu memprediksi hasil panen. Peningkatan produktivitas perkebunan kelapa sawit juga dipengaruhi oleh peningkatan kualitas tenaga kerja melalui pelatihan dan pengembangan kapasitas[8].

Kelapa sawit telah terbukti menjadi komoditas strategis bagi Indonesia yang berkontribusi besar terhadap PDB dan devisa negara[9]. Produksi tinggi dan ketahanan tanaman terhadap hama dan penyakit membuat kelapa sawit menjadi sumber minyak nabati yang sangat potensial. Namun, pertumbuhan industri kelapa sawit harus sejalan dengan prinsip keberlanjutan, dengan memperhatikan aspek lingkungan dan sosial. Solusi optimal untuk meningkatkan produktivitas, efisiensi, dan keberlanjutan industri kelapa sawit diharapkan ditemukan melalui penelitian dan pengembangan yang terus-menerus[10]. Kelapa sawit memainkan peran penting dalam perekonomian nasional karena merupakan salah satu komoditas utama di Indonesia. Namun, di balik potensinya yang besar, industri ini menghadapi banyak masalah yang kompleks. Proses yang menyeluruh diperlukan untuk mencapai keseimbangan antara pertumbuhan ekonomi dan kelestarian lingkungan. Penggunaan algoritma prediksi hasil panen dan teknologi pertanian lainnya dapat meningkatkan efisiensi produksi. Industri kelapa sawit yang berkelanjutan juga bergantung pada peningkatan kualitas tenaga kerja dan pengembangan praktik pertanian yang berkelanjutan[11].

Peneliti mendapatkan inspirasi dari penelitian sebelumnya untuk memulai dan mengembangkan penelitian ini. Penelitian sebelumnya menggunakan berbagai metode dan algoritma untuk memprediksi hasil panen kelapa sawit, meningkatkan akurasi dan efisiensi. Penelitian tentang penggunaan algoritma pembelajaran mesin seperti regresi linear, C4.5, Random Forest, dan Support Vector Machine (SVM) sangat penting[12]. Peneliti melihat bagaimana masing-masing algoritma bekerja dengan berbagai dataset berdasarkan ukuran, kompleksitas, dan kualitas data. Penelitian ini menekankan bahwa pemilihan algoritma yang tepat berdasarkan karakteristik data dan tujuan prediksi sangat penting. Selain itu, penelitian ini mengusulkan penggunaan kombinasi algoritma untuk meningkatkan hasil prediksi[13].

Penggunaan algoritma seperti regresi linier dan C45 dapat membuat prediksi hasil panen lebih akurat, yang memungkinkan perencanaan produksi yang lebih efisien. Pelatihan dan pengembangan kapasitas tenaga kerja juga akan meningkatkan produktivitas perkebunan dan kualitas sumber daya manusia. Penelitian dan pengembangan yang berkelanjutan sangat penting untuk menemukan varietas unggul, metode budidaya baru, dan solusi kreatif untuk masalah yang dihadapi industri ini[14]. Oleh karena itu, kelapa sawit dapat terus mendukung perekonomian nasional sambil mempertahankan kelestarian lingkungan. Meningkatkan kesejahteraan petani, menjaga kelestarian lingkungan, dan meningkatkan kontribusi industri kelapa sawit terhadap perekonomian nasional adalah tujuan utama pengembangan industri kelapa sawit[15]. Untuk mencapai tujuan tersebut, diperlukan upaya komprehensif yang mencakup adopsi praktik pertanian berkelanjutan dan penerapan teknologi modern seperti algoritma untuk memprediksi hasil panen. Selain itu, peningkatan kualitas tenaga kerja melalui pelatihan dan pengembangan kapasitas akan mendukung implementasi inovasi baru. Penelitian dan pengembangan yang berkelanjutan juga sangat penting untuk menemukan solusi bagi tantangan yang dihadapi industri ini, sehingga kelapa sawit dapat tetap menjadi komoditas unggulan di Indonesia[16].

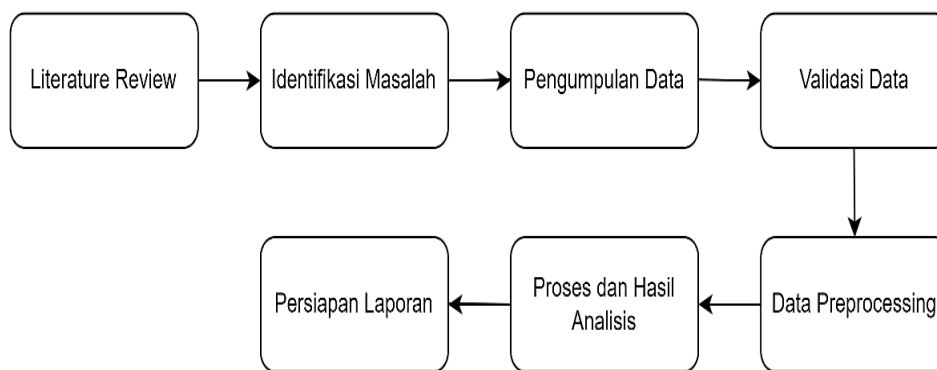
Tujuan dari penelitian ini adalah untuk menemukan kelebihan dan kekurangan dari algoritma regresi linear dan C45 dalam memprediksi hasil panen kelapa sawit. Ini akan dilakukan dengan menganalisis akurasi prediksi kedua algoritma berdasarkan data historis hasil panen dan faktor-faktor yang terkait. Tujuan utama penelitian ini adalah untuk memprediksi dengan akurat hasil panen kelapa sawit yang akan datang, yang akan membantu dalam pengambilan keputusan yang lebih baik tentang bagaimana mengelola perkebunan kelapa sawit[17]. Untuk mencapai tujuan tersebut, akan dilakukan perbandingan antara dua algoritma prediksi yang populer, yaitu C45 dan regresi linear. Selanjutnya, diharapkan akan ditemukan model prediksi terbaik dengan melakukan analisis data sejarah hasil panen dan komponen yang mempengaruhinya. Diharapkan hasil penelitian ini akan membantu meningkatkan efisiensi produksi kelapa sawit[18]. Tujuan penelitian ini adalah untuk mengetahui algoritma mana

yang lebih baik untuk memprediksi hasil panen kelapa sawit yang akan datang dengan menganalisis dan membandingkan kinerja regresi linear dan algoritma C45. Dengan memeriksa akurasi prediksi kedua algoritma berdasarkan data historis, penelitian ini diharapkan dapat menentukan algoritma mana yang lebih baik untuk memprediksi hasil panen yang akan datang. Pemahaman yang lebih baik tentang kebaikan dan keburukan masing-masing metode diharapkan akan memungkinkan pembuatan model prediksi yang lebih akurat dan dapat diandalkan untuk membantu proses pengambilan keputusan yang berkaitan dengan pengelolaan perkebunan kelapa sawit[19].

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Tahapan penelitian dilakukan agar penelitian ini lebih sistematis. Dalam penelitian ini meliputi beberapa Langkah yaitu Literatur Review, Identifikasi Masalah, Pengumpulan Data, Validasi Data, Proses dan Hasil Analisis, dan Persiapan Laporan. Dengan dilakukannya tahapan penelitian ini secara sistematis, peneliti dapat memastikan bahwa penelitian yang dilakukan sudah tersusun dengan baik dan dapat diandalkan. Berikut Gambar 1 menunjukkan tahapan yang dilakukan dalam penelitian ini.



Gambar 1. Tahapan Penelitian

Tujuan penelitian ini dicapai melalui serangkaian langkah yang meliputi literature review hingga persiapan laporan. Setiap tahap yang dirancang untuk menjawab pertanyaan penelitian yang telah dirumuskan sebelumnya.

a. Literature Review

Literature review juga dikenal dengan sebagai tinjauan Pustaka merupakan sebuah proses menyelidiki dan mengevaluasi literature yang relevan dengan topik penelitian. Literature review digunakan untuk mengumpulkan dan memperdalam teori yang berkaitan dengan perancangan dan pengolahan data mining. Penelitian ini menggunakan literature review sebagai bahan referensi jurnal dengan membaca jurnal dan artikel terdahulu. Jurnal yang diperdalam berkaitan dengan topik yang dibahas dan juga metode yang digunakan yang dimana metode yang digunakan pada penelitian ini adalah C45 dan Regresi Linear.

b. Identifikasi Masalah

Pada tahap ini atau Langkah awal dalam proses penelitian atau pemecahan masalah adalah dengan identifikasi masalah. Hak ini menentukan subjek atau masalah yang akan diteliti atau diselesaikan. Tahap identifikasi masalah dilaksanakan agar dapat merumuskan masalah sampai penelitian ini dapat terlaksana secara teratur dan tersusun baik.

c. Pengumpulan Data

Pada tahapan ini, peneliti mengumpulkan data yang bersumber langsung dari perusahaan PT. Surya Inti Sawit Kahuripan agar dapat lebih mempelajari permasalahan dalam melakukan penyusunan penelitian. Pengumpulan data yang dilakukan yaitu dengan metode sekunder yaitu melibatkan analisis literature atau data yang sudah ada untuk menjawab pertanyaan penelitian atau mencapai tujuan penelitian. Metode ini berbeda dengan metode primer, dimana peneliti harus mengumpulkan data secara langsung melalui penelitian mereka sendiri.

d. Validasi Data

Data yang telah didapatkan kemudian akan melewati fase pengecekan data. Pengecekan data dilakukan untuk memastikan tidak ada lagi kesalahan input data dengan cara meninjau kembali file atau dokumen yang didapatkan dari perusahaan. Hal ini dilakukan untuk lebih memastikan bahwa data yang diperoleh dapat dipercaya dan telah sesuai dengan karakteristik data yang dibutuhkan.

e. Data Preprocessing

Data preprocessing adalah sebuah pemrosesan sebuah data yang dilakukan sebelum melakukan perhitungan agar mudah di interpretasikan menggunakan sebuah mesin. Urutan tahapan dari data preprocessing yaitu cleaning data, transformasi data dan normalisasi data[12]. Cleaning data merupakan sebuah proses untuk mengidentifikasi serta mengoreksi atau menghapus data yang salah, kurang akurat, rusak, dan tidak relevan.

Transformasi data membantu dalam menyesuaikan format dan skala data untuk dapat memenuhi persyaratan model analisis ketika dilakukan perhitungan. Dan untuk normalisasi data yaitu untuk mencegah fitur tertentu dalam mendominasi data analisis, normalisasi sangat penting untuk digunakan karena sangat berpengaruh pada fitur yang memiliki skala berbeda-beda.

f. Proses dan Hasil Analisis

Proses analisis data dimulai dengan pengumpulan data sekunder yang relevan, yang melibatkan pencarian, identifikasi, dan pemilihan data yang sesuai dengan tujuan penelitian. Setelah data dikumpulkan, langkah selanjutnya adalah evaluasi data untuk memastikan keakuratan, keandalan, dan relevansinya dengan pertanyaan penelitian. Hasil analisis menunjukkan pola-pola dan hubungan antara variabel-variabel yang diteliti dengan hasil panen kelapa sawit. Misalnya, analisis dapat menunjukkan apakah faktor-faktor seperti curah pupuk, jenis tanah, atau tahun tanam berpengaruh signifikan terhadap hasil panen. Selain itu, analisis juga dapat menunjukkan praktik manajemen pertanian yang efektif dalam meningkatkan produktivitas tanaman. Hasil analisis ini dapat memberikan wawasan yang berharga tentang metode manajemen pertanian.

g. Persiapan Laporan

Salah satu dokumen penting untuk merekam dan mengkomunikasikan informasi tentang produktivitas dan kualitas panen kelapa sawit adalah laporan hasil panen. Persiapan laporan dimulai dengan pengumpulan data hasil panen dari lapangan yang mencakup informasi seperti luas ha, jumlah pohon, pohon per ha, tahun tanam, tanah. Setelah itu, data tersebut diorganisir dan disajikan secara sistematis dalam laporan yang mencakup ringkasan dan informasi tentang kualitas.

## 2.2 Metode Pengumpulan data

Pada penelitian ini, metode pengumpulan data yang digunakan yaitu metode sekunder yang biasanya melibatkan pengumpulan dan analisis data yang relevan dengan topik penelitian tertentu. Metode ini melibatkan penggunaan data atau informasi yang sudah ada, biasanya dikumpulkan oleh pihak lain untuk tujuan penelitian, berbeda dengan metode primer yang mengumpulkan data sendiri. Metode sekunder memiliki beberapa keuntungan, yaitu lebih menghemat waktu dan biaya karena tidak perlu melakukan pengumpulan data sendiri. Namun, penggunaan metode ini perlu mempertimbangkan kelemahan nya, seperti jumlah data yang tersedia terbatas dan kemungkinan ketidakcocokan antara data yang ada dengan pertanyaan penelitian berbeda.

Pada penelitian ini menggunakan jenis data primer, yang dimana data tersebut dikumpulkan melalui cara wawancara atau observasi langsung ke lokasi kepada pihak-pihak yang bersangkutan mengenai hal-hal yang berhubungan dengan hasil panen kelapa sawit pada PT Surya Inti Sawit Kahuripan[20]. Data yang diperlukan sangat penting untuk berbagai keperluan, yang pertama untuk mengevaluasi efektivitas dan efisiensi Kinerja Tanaman, produktivitas Tanaman, waktu panen dan kualitas tanaman. Yang kedua membuat rancangan panen dengan menggunakan data primer dapat membuat rencana panen yang lebih akurat dengan mempertimbangkan factor-faktor seperti umur Tanaman, kondisi tanah dan per kg/pohon. Yang ketiga menggunakan data primer meningkatkan hasil tanaman dengan mengidentifikasi faktor-faktor yang mempengaruhi hasil tanaman dan melakukan upaya perbaikan untuk meningkatkan hasil panen. Yang terakhir melakukan penelitian terhadap berbagai aspek tanaman kelapa sawit seperti pengaruh ke pemupukan, irigasi dan pengendalian hama[21]. Pada penelitian ini menggunakan 2 metode yaitu sebagai berikut:

### 2.2.1 Algoritma C45

C45 merupakan sebuah algoritma yang biasanya menggunakan pohon keputusan dari atas ke bawah. Atribut tertinggi disebut akar dan atribut terbawah disebut daun. Keuntungan dari Teknik ini adalah mudah dipahami oleh pengguna akhir dan efektif dalam menganalisis sejumlah besar fitur dari data saat ini[12]. Algoritma C4.5 adalah algoritma klasifikasi yang paling banyak digunakan dalam pemrosesan data dan pengajaran mesin. Beberapa kelebihan algoritma ini termasuk kemampuan untuk menangani atribut kontinu dan diskrit serta kemampuan untuk menangani data pelatihan dengan nilai atribut yang berbeda.

Algoritma C4.5 dapat digunakan untuk analisis data, pengambilan keputusan, dan sistem pakar. Ini adalah teknik klasifikasi yang efektif dan mudah dipahami yang menggunakan gain informasi untuk memilih atribut terbaik dalam membagi data dan menggunakan pengukuran kinerja seperti RMSE, SE, dan AE untuk mengevaluasi model. Tahapan perhitungan algoritma C45 yaitu sebagai berikut:

1. Untuk seluruh data pada node akar, hitung nilai RMSE, SE, dan AE. Nilai-nilai ini menunjukkan kesalahan prediksi rata-rata pada awal proses.
2. Menghitung gain informasi untuk setiap atribut dalam data. Gain informasi adalah ukuran seberapa banyak atribut tersebut membantu mengurangi ketidakpastian klasifikasi.
3. Memilih atribut terbaik untuk memisahkan data ke dalam kelas yang berbeda, pilih atribut dengan gain informasi tertinggi.
4. Hitung nilai RMSE, SE, dan AE untuk setiap kelompok tanaman. Nilai-nilai ini menunjukkan kinerja klasifikasi untuk setiap kelompok data. Sampai salah satu kondisi berikut terpenuhi: semua atribut telah digunakan, prosedur diulang secara rekursif. Target kinerja terpenuhi (misalnya, RMSE minimum tercapai). Hanya ada satu kelas data dalam subtree[22].

**2.2.2 Regresi Linear**

Regresi Linear adalah teknik statistik yang digunakan untuk memprediksi nilai variabel dependen (target) yang didasarkan pada satu atau lebih variabel independen (prediktor). Untuk melakukan prediksi ini, hubungan linier antara variabel independen dan variabel dependen dimodelkan, dan kemudian nilai variabel dependen diprediksi dengan menggunakan model regresi linear. Namun, penting untuk memahami asumsi dan keterbatasan regresi linear sebelum menggunakannya untuk menganalisis data atau membangun model prediksi karena regresi linear adalah teknik statistik yang kuat dan serbaguna yang dapat digunakan untuk berbagai tujuan. Untuk model regresi linear sederhana yang hanya menggunakan satu variabel independen, rumus umumnya adalah sebagai berikut:

$$Y = a + b X \tag{1}$$

Dalam analisis regresi, persamaan garis lurus dapat digunakan untuk menunjukkan hubungan antara dua variabel: variabel dependen (Y) dan variabel independen (X). Dua elemen penting dalam persamaan ini adalah intercept (a) dan koefisien regresi (b). Intercept (a) menunjukkan nilai Y ketika nilai X sama dengan nol, atau di mana garis regresi memotong sumbu Y. Sementara itu, koefisien regresi (b) menunjukkan kemiringan garis regresi, yang menunjukkan seberapa besar perubahan nilai Y yang diharapkan terjadi untuk setiap satuan perubahan pada nilai X.

Dengan kata lain, koefisien regresi (b) menghitung kekuatan dan arah hubungan antara variabel yang disebutkan sebagai X dan Y. Jika koefisien regresi bernilai positif, maka nilai X akan meningkat bersamaan dengan nilai Y. Sebaliknya, jika koefisien regresi negatif, maka nilai Y akan turun jika nilai X meningkat. Model regresi linear sederhana ini memungkinkan kita untuk memprediksi nilai Y berdasarkan nilai X yang diketahui. Kita dapat menghitung nilai Y yang diprediksi untuk setiap nilai X tertentu dengan menggunakan nilai intercept (a) dan koefisien regresi (b)[23].

Regresi linear dapat digunakan untuk memprediksi nilai variabel dependen yang didasarkan pada satu atau lebih variabel independen; variabel dependen harus kontinu, artinya mereka harus memiliki nilai yang dapat diukur dalam skala interval atau rasio. Selain itu, regresi linear dapat digunakan untuk menganalisis cara variabel-variabel dalam dataset berhubungan satu sama lain. Regresi linear adalah teknik statistik yang sederhana dan efektif untuk memprediksi nilai variabel respons berdasarkan variabel independen. Algoritma ini menggunakan optimasi untuk menemukan nilai parameter model yang optimal dan menggunakan pengukuran kinerja seperti RMSE, SE, dan AE untuk mengevaluasi akurasi model. Tahapan perhitungan algoritma Regresi Linear yaitu sebagai berikut:

1. Nilai prediksi variabel respons untuk setiap titik data instruksi dihitung menggunakan model regresi linear. Kemudian, hitung kesalahan prediksi, yang dihitung sebagai selisih antara nilai prediksi dan nilai variabel respons yang sebenarnya.
2. Untuk menemukan nilai parameter model yang meminimalkan fungsi error, lakukan iterasi optimasi. Fungsi error rata-rata yang digunakan adalah Mean Squared Error (MSE).
3. Hitung nilai kinerja model pada data pelatihan dan pengujian. Untuk mengevaluasi akurasi model dalam memprediksi nilai variabel respons, metrik kinerja yang umum digunakan adalah RMSE (*Root Mean Square Error*), SE (*Standard Error*), dan AE (*Absolute Error*).

**3. HASIL DAN PEMBAHASAN**

**3.1 Hasil**

Penelitian ini dilakukan berdasarkan tahapan yang ditentukan dalam metode penelitian. Tahap penelitian dilakukan mulai dari penelitian literatur review hingga publikasi hasil laporan. Sumber data yang digunakan dalam penelitian berasal dari laporan statistik tanaman perusahaan. Untuk itu, komunikasi peneliti yang dilakukan sebagai bagian dari pengumpulan data menghasilkan total 106 blok, data penelitian sebagai sampel dalam penelitian ini yang diolah dan dianalisis untuk mengidentifikasi pola dari kumpulan data tersebut. Untuk menemukan pola dan variabel yang mempengaruhi hasil panen, berbagai metode statistik dan algoritma pembelajaran mesin digunakan untuk menganalisis data yang dikumpulkan dari 106 blok perkebunan kelapa sawit. Sebelum diproses lebih lanjut, pengolahan data memastikan kualitas data terbaik dengan melakukan prosedur pra-pemrosesan seperti pembersihan, penanganan nilai yang tidak ada, normalisasi, dan transformasi.

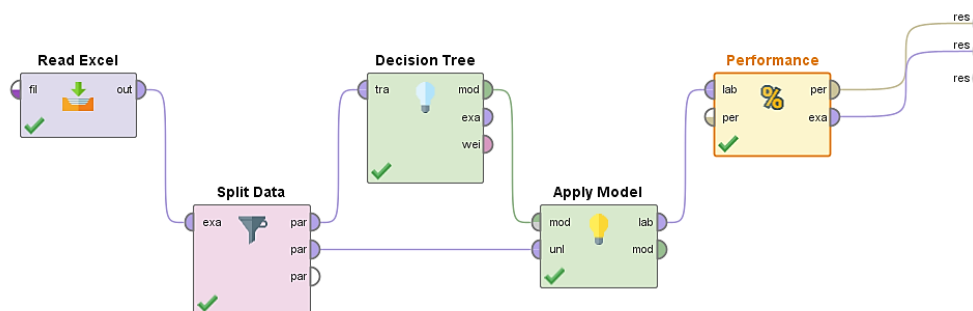
**Tabel 1.** Dataset

No	Blok	Luas (Ha)	Jumlah Pohon	Pohon Per ha	Tahun Tanam	Tanah	Renc	Pupuk	Kg/phn	Ton	Kondisi
1	C18	14,53	1184	50	2009	Mineral	Feb-Mar	NPK 13:6:27:4	2,25 0	2,65 0	Sedikit
2	C18	14,53	1184	50	2009	Mineral	Mei-Jun	NPK 13:6:27:4	2,25 0	2,65 0	Sedikit
3	C18	14,53	1184	50	2009	Mineral	Sep-Okt	NPK 13:6:27:4	2,25 0	2,65 0	Sedikit

4	D18	30,53	3303	76	2009	Mineral	Feb-Mar	NPK 13:6:27:4	2,25 0	7,45 0	Sedang
5	D18	30,53	3303	76	2009	Mineral	Mei-Jun	NPK 13:6:27:4	2,25 0	7,45 0	Sedang
..	.....	.....	.....	.....	.....	.....	.....	.....	...	...	....
104	F42	15,74	1835	135	2005	Pasir	Feb-Mar	NPK 13:6:27:4 +B	2,45 2	5,67 0	Sedang
105	F42	15,74	1835	135	2005	Pasir	Mei-Jun	NPK 13:6:27:4 +B	2,45 2	5,67 0	Sedang
106	F42	15,74	1835	135	2005	Pasir	Sep-Okt	NPK 13:6:27:4 +B	2,45 2	5,67 0	Sedang

### 3.1.1 Hasil proses analisis performance C45

Berdasarkan data historis dan faktor-faktor terkait, algoritma C45 digunakan untuk membuat sebuah model prediktif hasil panen kelapa sawit. Algoritma ini dapat mengatasi masalah klasifikasi dan regresi, serta menangani atribut kontinu dan diskrit. Hasil perhitungan menunjukkan bahwa algoritma C45 memiliki tingkat akurasi yang tinggi dalam memprediksi hasil panen. Algoritma C45 mengevaluasi modelnya dengan menggunakan berbagai metrik kesalahan seperti AE, RMSE, dan SE. Untuk proses algoritma C45 dapat dilihat pada gambar 2 di bawah ini.



Gambar 2. Alur proses analisis C45

Dari proses analisis performance diatas, mendapatkan hasil Mean Squared Error (MSE) dengan menggunakan software rapid miner yaitu 0.207

```
PerformanceVector:
absolute_error: 0.085 +/- 0.189
root_mean_squared_error: 0.207 +/- 0.000
squared_error: 0.043 +/- 0.173
```

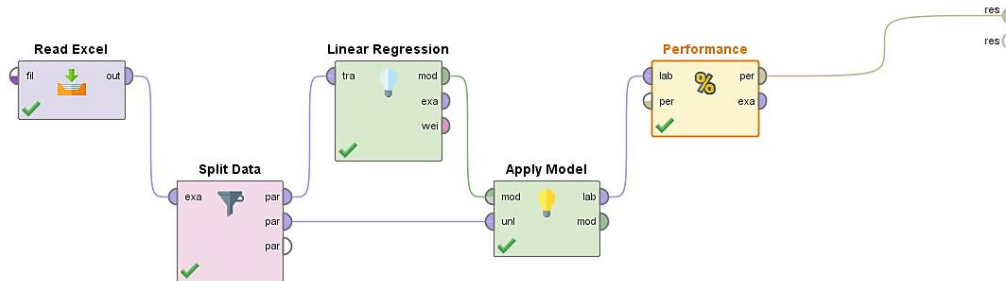
Gambar 3. Performance Vector C45

Penjelasan mengenai 3 performance vector pada gambar 3 diatas yaitu evaluasi kinerja model C4.5 yang dilakukan menunjukkan hasil yang cukup baik berdasarkan metrik error yang digunakan. Absolute error sebesar 0.085 mengindikasikan bahwa secara rata-rata, prediksi model hanya meleset sebesar 0.085 unit dari nilai aktual. Sementara itu, Root Mean Squared Error (RMSE) sebesar 0.207 dan Squared Error sebesar 0.043 juga memberikan konfirmasi akan akurasi model. Nilai RMSE yang relatif rendah menunjukkan bahwa model mampu menghasilkan prediksi yang cukup dekat dengan nilai sebenarnya. Meskipun demikian, perlu diingat bahwa interpretasi nilai-nilai error ini sangat bergantung pada konteks permasalahan dan skala data yang digunakan.

### 3.1.2 Hasil proses analisis performance regresi

Berdasarkan data yang sudah dilakukan proses transformasi, regresi digunakan untuk memprediksi nilai suatu variabel (variabel dependen) didasarkan pada nilai variabel independen lainnya dalam regresi. jenis regresi yang paling sederhana adalah regresi linear, yang mana hubungan antara variabel dependen dan variabel independen diasumsikan linear. Berikut penjelasan proses dalam melakukan perhitungan menggunakan software rapidminer.

1. Read Excel digunakan untuk membaca file excel yang sudah di transformasi ke dalam operator untuk dapat mengubahnya menjadi format yang dibutuhkan.
2. Split Data dapat membagi data menjadi dua bagian yaitu data training dan data testing[24]. Dataset training digunakan untuk melatih model backpropagation dan dataset testing digunakan dalam mengevaluasi performa model. Pada perhitungan algoritma regresi menggunakan data training sebesar 0.8 dan data testing sebesar 0.2.
3. Linear Regression merupakan model regresi linear yang digunakan untuk memprediksi churn pelanggan.
4. Apply Model membantu penerapan model regresi linear yang telah dilatih diterapkan pada dataset testing.
5. Performance digunakan untuk menghitung kinerja model regresi linear pada dataset testing.



**Gambar 4.** Alur proses analisis Regresi Linear

Dari proses analisis performance diatas, mendapatkan hasil Mean Squared Error (MSE) dengan menggunakan software rapid miner yaitu 1.230

```
PerformanceVector:
root_mean_squared_error: 1.230 +/- 0.000
absolute_error: 1.020 +/- 0.687
squared_error: 1.512 +/- 2.099
```

**Gambar 5.** Performanc vector Regresi Linear

Penjelasan mengenai 3 performance vector pada regresi linear pada gambar 5 diatas bahwa evaluasi kinerja model regresi linear yang dilakukan menunjukkan hasil yang dapat diinterpretasikan. Berdasarkan metrik error yang digunakan, dapat disimpulkan bahwa model memiliki tingkat akurasi yang cukup baik. Nilai absolute error sebesar 1.020 mengindikasikan bahwa secara rata-rata, prediksi model meleset sekitar 1.020 unit dari nilai aktual. Sementara itu, nilai Root Mean Squared Error (RMSE) sebesar 1.230 dan Squared Error sebesar 1.512 juga memberikan informasi serupa. Nilai RMSE yang relatif rendah menunjukkan bahwa model mampu menghasilkan prediksi yang cukup dekat dengan nilai sebenarnya. Meskipun demikian, penting untuk diingat bahwa interpretasi nilai-nilai error ini sangat bergantung pada konteks permasalahan dan skala data yang digunakan.

## 2.2 Pembahasan

Pada bagian pembahasan ini memberikan gambaran mengenai hasil panen kelapa sawit pada perkebunan sawit PT Surya Inti Sawit Kahuripan. Pembahasan juga menghasilkan rekomendasi untuk penelitian lanjutan. Peneliti menyarankan metode atau pendekatan yang dapat digunakan dalam penelitian mendatang jika temuan menunjukkan bahwa variabel tertentu yang belum banyak diteliti perlu dipelajari lebih lanjut. Ini memberikan jalan yang jelas bagi peneliti lain di bidang yang sama dan mendorong kemajuan ilmu pengetahuan.

Bagian penting dari laporan penelitian adalah pembahasan. Bagian ini peneliti menyampaikan isi-isi penelitian, membangun hubungan dengan literatur yang ada, dan memberikan saran praktis. Analisis dan interpretasi mendalam dari temuan penelitian membantu mengaitkan temuan dengan bidang ilmu yang lebih luas dan menunjukkan relevansinya. Pada analisis ini, terdapat beberapa variabel yang digunakan untuk memberikan pemahaman tentang karakteristik yang ada pada perkebunan kelapa sawit, karakteristik pada penelitian ini yaitu blok, luas(ha), jumlah pohon, pohon per ha, tahun tanam, tanah, renc, pupuk, kg/pohon, ton dan kondisi. Dari hasil yang telah dilakukan dapat diketahui bahwa hasil perhitungannya sebagai berikut:

### 3.2.1 Perbandingan Hasil Performance C45 dan Regresi Linear

Berdasarkan hasil proses analisis menggunakan rapid miner, performa kedua algoritma yaitu C45 dan Regresi Linear mendapatkan tiga metrik yaitu absolute error, root mean squared error dan squared error.

Nilai AE menunjukkan selisih absolut rata-rata antara nilai aktual dan nilai prediksi model. Nilai AE yang lebih rendah menunjukkan bahwa model memprediksi nilai aktual dengan lebih baik. Pada kondisi ini, C4.5 memiliki nilai AE 0,085 lebih rendah daripada Regresi, yang menunjukkan bahwa C4.5 secara umum menghasilkan prediksi yang lebih dekat dengan nilai aktual daripada Regresi. Deviasi standar untuk C4.5 lebih

rendah, yaitu 0.189, sedangkan regresi sebesar 0,687, yang menunjukkan bahwa prediksi C4.5 lebih konsisten daripada Regresi.

Nilai RMSE adalah akar kuadrat dari error kuadrat rata-rata, seperti AE, dan nilai RMSE yang lebih rendah, yaitu 0,207 daripada Regresi, menunjukkan bahwa C4.5 secara keseluruhan menghasilkan prediksi yang lebih akurat daripada Regresi. Nilai deviasi standar tidak tersedia karena hanya 0.000, tetapi C4.5 tetap lebih unggul berdasarkan nilai RMSE.

Nilai SE yang lebih kecil, menunjukkan seberapa baik model memprediksi nilai aktual. Nilai SE menunjukkan selisih rata-rata kuadrat antara nilai aktual dan nilai prediksi model. Nilai SE C4.5 lebih rendah 0,043 dibandingkan Regresi 1,512, yang menunjukkan bahwa C4.5 secara rata-rata menghasilkan prediksi yang lebih dekat dengan nilai aktual dibandingkan Regresi. Deviasi standar C4.5 juga lebih rendah, 0,173, yang menunjukkan bahwa prediksi C4.5 lebih stabil dan konsisten dibandingkan Regresi, yang memiliki deviasi standar yang tinggi, 2.099, yang menunjukkan variabilitas yang lebih besar.

Pada semua metrik yang dianalisis (AE, RMSE, dan SE), C4.5 menunjukkan performa yang lebih baik dibandingkan dengan Regresi Linear. Deviasi standar yang lebih kecil pada C4.5 menunjukkan bahwa prediksi dari model ini lebih konsisten dibandingkan dengan Regresi Linear, sementara Regresi Linear menghasilkan nilai error yang lebih besar dan deviasi standar yang lebih besar, yang menunjukkan bahwa model ini kurang akurat dan prediksinya kurang konsisten dibandingkan dengan C4.5.

Dari analisis ini, dapat disimpulkan bahwa, jika dibandingkan dengan regresi linear, algoritma C4.5 adalah pilihan yang lebih baik untuk memprediksi hasil panen kelapa sawit dalam penelitian ini. Memiliki nilai squared error yang kecil menunjukkan bahwa model prediksi akurat. Ini adalah metrik yang digunakan untuk menghitung kuadrat selisih antara nilai prediksi model dan nilai aktual. Untuk nilai performance pada squared error regresi linear yaitu 1.512.

## 4. KESIMPULAN

Berdasarkan hasil penelitian prediksi dengan membandingkan algoritma C45 dan Regresi Linear yang telah dilakukan dalam meneliti hasil panen kelapa sawit PT Surya Inti Sawit Kahuripan, memperoleh hasil sebagai berikut. Perbandingan analisis performance yang dihasilkan menggunakan algoritma C45 dan Regresi Linear menghasilkan nilai Absolute Error C45 sebesar 0,085 dengan deviasi sebesar 0,189 dan nilai Absolute Error Regresi Linear sebesar 1,020 dengan deviasi sebesar 0,687. Nilai Root Mean Squared Error C45 sebesar 0,207 dengan deviasi sebesar 0,000 dan nilai Root Mean Squared Error Regresi Linear sebesar 1,230 dengan deviasi sebesar 0,000. Nilai Squared Error C4 sebesar 0,043 dengan deviasi sebesar 0,173 dan nilai Squared Error Regresi Linear sebesar 1,512 dengan deviasi sebesar 2,099. Dari hasil perbandingan nilai performane tersebut, algoritma dengan nilai RMSE yang lebih kecil adalah algoritma terbaik dan memiliki tingkat keakuratan yang lebih baik. Jadi, metode C45 lebih disarankan dalam melakukan prediksi hasil panen kelapa sawit pada PT Surya Inti Sawit Kahuripan karena memiliki angka error yang lebih kecil dibandingkan algoritma Regresi Linear. Kesimpulan dari faktor-faktor yang dipertimbangkan saat memilih algoritma terbaik bergantung pada analisis dataset dan kebutuhan prediksi. Menguji kedua algoritma dan membandingkan hasilnya dalam hal akurasi, interpretabilitas, dan kecepatan akan membantu dalam memahami model yang lebih sesuai untuk prediksi hasil panen kelapa sawit. C4.5 lebih baik dalam melakukan analisis hasil panen kelapa sawit karena memiliki tingkat nilai error lebih rendah dan lebih akurat dibandingkan dengan regresi linear.

## REFERENCES

- [1] A. Cynthia, S. Siregar, P. S. Veni, and R. Ingesti, "ANALISIS KUALIFIKASI TENAGA KERJA PANEN TERHADAP CAPAIAN NORMA PANEN KELAPA SAWIT (*Elaies guineensis* Jacq)," 2024.
- [2] D. Suryani, A. Yulianti, E. L. Maghfiroh, and J. Alber, "Klasifikasi Kualitas Produk Kelapa Sawit Menggunakan Metode Naïve Bayes Quality Classification of Palm Oil Products Using Naïve Bayes Method," 2022. [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [3] Abubakar Ahmed and Ishak Yusoff Mohd, "Gambaran Umum Peran Petani Kecil dalam Sistem Produksi Kelapa Sawit dalam Perubahan Iklim," *NsifatElingkungan danPolusiTteknologi Jurnal Ilmiah Triwulanan Internasional*, pp. 2055–2071, 2022, [Online]. Available: [www.neptjournal.com](http://www.neptjournal.com)
- [4] Darmono, "PEMBERDAYAAN PEKEBUN KELAPA SAWIT SWADAYA MENUJU PEMBANGUNAN PERKEBUNAN BERKELANJUTAN DI KABUPATEN BERAU," vol. 5, no. 1, 2021.
- [5] Situmorang Sebastian Redi, "PENERAPAN DATA MINING UNTUK MEMREDIKSI JUMLAH PRODUKSI KELAPA SAWIT MENGGUNAKAN METODE REGRESI LINEAR BERGANDA (Studi Kasus : PT. SGSR Mujur Sinar Gunung Sawit Raya)," 2023.
- [6] Afonso Jose *et al.*, "Model regresi tidak secara inheren bersifat prediktif. Tinjauan edukatif tentang mengapa penelitian olahraga dan kedokteran olahraga harus lebih berhati-hati dengan istilah "prediktor" (Versi 1)," 2024.
- [7] Lestari Indah Hidayah Nur, "PENGARUH KOMPENSASI DAN BUDAYA ORGANISASI TERHADAP PRODUKTIVITAS DENGAN MEDIASI MOTIVASI KERJA (Studi Kasus Pada PT. Hutan Rindang Banua)," *JBS (Jurnal Berbasis Sosial) Pendidikan IPS STKIP Al Maksum*, vol. 1, no. 2, 2020.
- [8] N. Rusmilawati and P. T. Prasetyaningrum, "Penerapan Data Mining Dalam Prediksi Hasil Produksi Kelapa Sawit PT Borneo Ketapang Indah Menggunaka Metode Linier Regression," 2021.

- [9] Noviana Githa and Ardiani Fani, "ANALISIS PENDAPATAN PETANI KELAPA SAWIT SEBELUM DAN SELAMA COVID-19 (STUDI KASUS: KABUPATEN PADANG LAWAS UTARA) ANALYSIS OF OIL PALM FARMERS INCOME BEFORE AND DURING COVID-19 (CASE STUDY: NORTH PADANG LAWAS DISTRICT)," 2020.
- [10] Aswan Novita, Fadhillah Yusra, Siregar Hasan Noor Muhammad, Puspita Rina Nurkhasanah, Mahaji Toga, and Harahap Utari Angelia, "Analisis Regresi Linier dalam Estimasi Produktivitas Kelapa Sawit(*Elaeis guineensis*) Di Kabupaten Tapanuli Selatan," *INNOVATIVE: Journal Of Social Science Research*, vol. 4, no. 1, pp. 10264–10276, 2024.
- [11] N. Ambar Suharyanti, K. Mizuno, and A. Sodri, "Pengaruh defisit air pada periode berbunga pada produktivitas kelapa sawit di lahan gambut," in *E3S Web of Conferences*, EDP Sciences, Nov. 2020. doi: 10.1051/e3sconf/202021105005.
- [12] D. Sri Rahayu, J. Afifah, and S. Intan, "SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat Classification of Diabetes Mellitus Using C4.5 Algorithm, Support Vector Machine (SVM) and Linear Regression Klasifikasi Penyakit Diabetes Melitus Menggunakan Algoritma C4.5, Support Vector Machine (SVM) dan Regresi Linear," 2023. [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas>
- [13] A. Agung and Putri, "Penerapan Data Mining Untuk Mengestimasi Laju Data Mining Usage To Estimate Civil Growth In Denpasar," vol. 6, no. 1, pp. 37–44, 2023.
- [14] V. Arinal and M. Azhari, "Penerapan Regresi Linear Untuk Prediksi Harga Beras Di Indonesia," *Jurnal Sains dan Teknologi*, vol. 5, no. 1, p. pp, 2023, doi: 10.55338/saintek.v5i1.1417.
- [15] C. E. Simbolon, "Penerapan Algoritma Regresi Linier Sederhana Dalam Memprediksi Keuntungan dan Kerugian Kelapa Sawit Pt. Sri Ulina Ersada Karina," 2021.
- [16] A. Prasetyo, Salahuddin, and Amirullah, "Prediksi Produksi Kelapa Sawit Menggunakan Metode Regresi Linier Berganda," *Multimedia & Jaringan*, vol. 6, no. 2, 2021.
- [17] W. Ananda, M. Safii, and M. Fauzan, "Prediksi Jumlah Hasil Panen Sawit Menggunakan Algoritma Naive Bayes," 2021.
- [18] D. Nurhasanah, D. Ayu Lestari, S. Simatupang, S. Informasi, and S. Royal Kisaran, "PEMILIHAN KUALITAS PRODUK KELAPA SAWIT MENGGUNAKAN METODE NAÏVE BAYES DI LABUHANBATU SELATAN," *Jurnal Teknologi Komputer dan Sistem Informasi Februari*, no. 1, pp. 24–31, 2023, [Online]. Available: <http://jurnal.goretanpena.com/index.php/teknisi>
- [19] M. Marsella and W. Susanti, "Penerapan Algoritma Naïve Bayes untuk Klasifikasi Spare Part Industri Kelapa Sawit," *Jurnal Mahasiswa Aplikasi Teknologi Komputer dan Informasi*, vol. 4, no. 3, pp. 88–99, 2022.
- [20] Yusuf Yusril, "ANALISIS PERBANDINGAN SISTEM PENDUKUNG KEPUTUSAN MENGGUNAKAN METODE SAW DAN WP DALAM PENILAIAN KINERJA TENAGA KONTRAK," 2022.
- [21] A. Bravo, T. Tursina, and H. Sastypratiwi, "Penerapan Metode Naive Bayes Untuk Penentuan Bibit Kelapa Sawit Berdasarkan Kondisi Daerah Tanam dan Perawatan Tanaman," *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 11, no. 1, p. 101, Jan. 2023, doi: 10.26418/justin.v11i1.52277.
- [22] I. Ubaedi and Y. M. Djaksana, "OPTIMASI ALGORITMA C4.5 MENGGUNAKAN METODE FORWARD SELECTION DAN STRATIFIED SAMPLING UNTUK PREDIKSI KELAYAKAN KREDIT," *Sistem Informasi* /, vol. 9, no. 1, 2022.
- [23] Juwita Ratna Ayu, Pratama Rizky Adi, Mudzakir Al Tohirin, and Sukmawati Emilia Cici, "PERBANDINGAN ALGORITMA C4.5 DAN REGRESI LINEAR DALAM MEMPREDIKSI KETERLAMBATAN PEMBAYARAN UANG KULIAH," *Konferensi Nasional Penelitian dan Pengabdian (KNPP) Ke-3*, 2023.
- [24] I. N. G. Neyun, W. Sulandari, and I. Slamet, "Hibrida Autoregresiive Integrated Moving Average dan Fuzzy Time Series Cheng untuk Prediksi Harga Saham Autoregresiive Integrated Moving Average and Fuzzy Time Series Cheng Hybrid for Predicting Stock Price," *Jurnal Bumigora Information Technology (BITE)*, vol. 5, no. 2, pp. 139–150, 2023, doi: 10.30812/bite/v5i1.2972.