

Prediksi Kadar Alpha Pinene Minyak Terpentin Berdasarkan Komposisi Bahan Baku Menggunakan Regresi Linear Berganda

Nisrina Nur Sa'idah¹, Masy Ari Ulinuha², Nur Cahyo Hendro Wibowo³, Iwan Gunawan⁴

^{1,2,3} Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Walisongo, Indonesia

⁴ Departemen Riset dan Inovasi, Perhutani Forestry Institute, Indonesia

¹nisrinanurs05@gmail.com, ²ulinuha@walisongo.ac.id, ³nur_cahyo@walisongo.ac.id, ³pefi@perhutani.co.id

Info Artikel

Riwayat Artikel:

Received 2025-06-23

Revised 2026-02-08

Accepted 2026-02-28

Corresponding Author:

Masy Ari Ulinuha

Email: ulinuha@walisongo.ac.id



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

Abstract – Turpentine oil is a product of pine resin distillation. This oil has high economic value, mainly due to its alpha pinene content, which determines the quality of turpentine oil products. Industry standards set the ideal alpha pinene content at $\geq 80\%$. However, some turpentine oil production at the Gondorukem and Terpentin Factory (PGT) produces levels below the standard, which can reduce market appeal. Therefore, this study focuses on developing a prediction model for alpha-pinene content based on the composition of pine resin raw materials. This study uses multiple linear regression with secondary data from Perum Perhutani, covering six input variables (Super Premium (SP), Premium (P), Quality IA (IA), Quality IB (IB), Quality IIA (IIA), and Quality IIB (IIB)) and one target variable (alpha pinene content). The research stages included data pre-processing, model design (manual and Python), prediction calculations, and evaluation. The results show that the model has a low prediction error rate: MAE 0.01312, MSE 0.00026, and RMSE 0.01622. However, the R-squared value is only 0.20854, meaning that the model only explains about 20.854% of the data variability. A low R-squared value indicates that the model has limitations in explaining the data variation comprehensively, so it serves as an initial (baseline) predictive model. Therefore, further model development is needed through the addition of upstream variables (genetic, environmental, handling) and downstream variables (processing) or a more complex modeling approach to improve prediction accuracy in the turpentine oil industry.

Keywords: Alpha Pinene; Turpentine Oil; Prediction; Multiple Linear Regression; Python

Abstrak – Minyak terpentin merupakan produk hasil penyulingan getah pinus. Minyak ini memiliki nilai ekonomi yang tinggi, terutama karena kandungan alpha pinenya yang menentukan kualitas produk minyak terpentin. Standar industri menetapkan kadar ideal alpha pinene adalah $\geq 80\%$. Namun, beberapa produksi minyak terpentin di Pabrik Gondorukem dan Terpentin (PGT) menghasilkan kadar di bawah standar, yang bisa mengurangi daya tarik pasar. Karena itu, penelitian ini berfokus pada pengembangan model prediksi kadar alpha pinene berdasarkan komposisi bahan baku getah pinus. Penelitian ini menggunakan metode regresi linear berganda dengan data sekunder dari Perum Perhutani, mencakup enam variabel input (Super Premium (SP), Premium (P), Mutu IA (IA), Mutu IB (IB), Mutu IIA (IIA), dan Mutu IIB (IIB)) dan satu variabel target (kadar alpha pinene). Tahapan penelitian meliputi pra-pemrosesan data, perancangan model (manual dan Python), perhitungan prediksi, dan evaluasi. Hasilnya, model menunjukkan tingkat kesalahan prediksi yang rendah: MAE 0,01312, MSE 0,00026, dan RMSE 0,01622. Namun, nilai R-squared hanya 0,20854, artinya model hanya menjelaskan sekitar 20,854% variabilitas data. Nilai R-squared yang rendah mengindikasikan bahwa model memiliki keterbatasan dalam menjelaskan variasi data secara menyeluruh sehingga bersifat sebagai model prediktif awal (baseline). Oleh karena itu, diperlukan pengembangan model lebih lanjut melalui penambahan variabel hulu (genetik, lingkungan, penanganan) dan hilir (proses pengolahan) atau pendekatan pemodelan yang lebih kompleks untuk meningkatkan akurasi prediksi pada industri minyak terpentin.

Kata Kunci: Alpha Pinene; Minyak Terpentin; Prediksi; Regresi Linear Berganda; Python

I. PENDAHULUAN

Minyak terpentin atau yang biasa dikenal sebagai "spirits of turpentine" merupakan cairan mudah menguap yang dihasilkan dari penyulingan getah pohon pinus [1]. Salah satu komponen utamanya adalah α -pinene, senyawa monoterpen yang memiliki nilai ekonomi tinggi karena penggunaannya yang luas dalam industri kimia, farmasi, hingga parfum [2]. Senyawa ini kerap dimanfaatkan sebagai bahan baku dalam pembuatan pelarut, antiseptik, kamfer, hingga produk kosmetik.

Di Indonesia, Pinus merkusi merupakan spesies pinus yang secara alami tumbuh di wilayah Sumatera bagian utara dan menjadi sumber utama produksi minyak terpentin [3]. Getah pinus tersebut menjadi sumber utama produksi minyak terpentin. Hutan pinus memberikan manfaat ganda baik bagi Perum Perhutani selaku pengelola maupun masyarakat petani yang tinggal di sekitar hutan [4]. Perum Perhutani, sebagai pengelola hutan pinus, memiliki delapan

pabrik gondorukem dan terpentin (PGT) yang tersebar di Jawa Tengah, Jawa Timur, dan Jawa Barat, dengan total kapasitas produksi mencapai 92.550 ton [3].

Minyak terpentin yang diproduksi di Indonesia umumnya memiliki kandungan α -pinene antara 57% hingga 86%, disertai senyawa lain seperti δ -karena sebesar 8-12%, serta monoterpen lainnya dalam kadar yang lebih rendah. [5]. Oleh karena itu alpha pinene adalah senyawa dominan yang sangat penting dalam menentukan kualitas produk. Standar industri menetapkan bahwa kandungan alpha pinene idealnya $\geq 80\%$ untuk memastikan kualitas yang optimal [3]. Namun, beberapa hasil produksi minyak terpentin di PGT menunjukkan kadar alpha pinene di bawah standar tersebut, sehingga berpotensi menurunkan minat pasar dan daya saing produk. Penurunan kualitas ini diperkirakan disebabkan oleh komposisi mutu getah pinus yang digunakan sebagai bahan baku. Oleh karena itu, diperlukan kajian untuk meningkatkan kadar alpha pinene hingga memenuhi standar, yang dapat dicapai dengan merancang model prediksi kadar alpha pinene berdasarkan komposisi bahan baku getah. Model ini diharapkan mampu mengidentifikasi variabel-variabel kunci yang memengaruhi kadar alpha pinene, sehingga langkah-langkah perbaikan dapat diimplementasikan untuk menjamin produksi minyak terpentin berkualitas tinggi.

Berdasarkan kajian terhadap penelitian-penelitian terdahulu, belum ditemukan penelitian yang secara spesifik membahas prediksi kadar alpha pinene minyak terpentin berdasarkan komposisi mutu getah menggunakan pendekatan regresi linear berganda. Penelitian terdahulu umumnya masih berfokus pada aspek proses produksi, analisis kimia, atau karakterisasi kualitas minyak terpentin, namun belum menitikberatkan pada pembangunan model prediktif dari sisi bahan bakunya. Oleh karena itu, penelitian ini diharapkan mampu mengisi celah tersebut dan menjadi kontribusi awal dalam pengembangan model prediksi berbasis data pada industri terpentin.

Penelitian ini menerapkan regresi linear berganda sebagai metode utama dalam membangun model prediksi kadar alpha pinene karena metode ini mampu menjelaskan hubungan linier antara beberapa variabel independen dan satu variabel dependen secara kuantitatif. Regresi linear berganda memungkinkan analisis kontribusi masing-masing variabel mutu getah terhadap perubahan kadar alpha pinene, sekaligus memberikan model yang bersifat interpretable dan mudah diterapkan pada industri [6]. Dalam penelitian ini, regresi linear digunakan untuk menganalisis hubungan linier antara beberapa variabel prediktor, yaitu komposisi mutu getah (SP, P, IA, IB, IIA, IIB), dengan satu variabel target, yaitu kadar alpha pinene. Pendekatan ini sesuai digunakan pada tahap awal pengembangan model prediksi berbasis data, khususnya ketika tujuan penelitian tidak hanya berfokus pada akurasi, tetapi juga pada pemahaman pengaruh variabel input terhadap kualitas produk.

Penelitian-penelitian sebelumnya menunjukkan bahwa regresi linear banyak digunakan untuk memprediksi berbagai fenomena di sektor pertanian, kesehatan, dan ekonomi. Namun, sebagian besar studi tersebut berfokus pada objek dan variabel yang berbeda, seperti prediksi produksi, kadar air, harga, atau kualitas produk di sektor non-terpentin. Hingga saat ini, belum terdapat penelitian yang secara khusus mengembangkan model prediksi kadar α -pinene minyak terpentin berbasis komposisi mutu bahan baku getah pinus. Oleh karena itu, penelitian ini memiliki kebaruan pada perbedaan objek kajian, perbedaan variabel prediktor, serta tujuan praktis yang berorientasi langsung pada industri pengolahan terpentin sebagai dasar pengambilan keputusan kualitas bahan baku.

Studi-studi sebelumnya menunjukkan penerapan regresi linear dalam berbagai bidang, Listina dkk mengembangkan model prediksi kadar air media tanam untuk budidaya tomat menggunakan regresi linear berganda, mencapai akurasi 83,80% dalam RMSE dan 72,02% dalam R-squared, yang dapat membantu petani dalam pengambilan keputusan penyiraman [7]. Di sektor lain, Jaelani membandingkan regresi linear dan *Long Short-Term Memory* (LSTM) dalam memprediksi produksi gula nasional, dengan regresi linear menunjukkan kesalahan sekitar 8%, yang memberikan wawasan tentang efisiensi metode dalam skala makro [8]. Dalam bidang kesehatan, penelitian yang dilakukan oleh Yanni merancang sistem prediksi jumlah obat keluar menggunakan regresi linear untuk menentukan jumlah pemesanan obat. Dalam penelitiannya menghasilkan kesalahan 12,42% (*Mean Absolute Percentage Error* - MAPE), yang menunjukkan relevansi metode ini dalam manajemen inventaris [9]. Penelitian yang dilakukan oleh Fajar juga memanfaatkan regresi linear dan *Autoregressive Integrated Moving Average* (ARIMA) untuk memprediksi harga rumah di Kabupaten Bogor, dengan model regresi linear menghasilkan kesalahan sekitar 2,66% [10]. Penelitiannya menyoroti aplikasinya dalam analisis pasar. Selain itu, penelitian Herdiana & Geraldine menerapkan model regresi linear untuk menganalisis kualitas hasil petik teh secara otomatis, hal ini menunjukkan potensi untuk mempercepat proses analisis kualitas di industri pertanian [11].

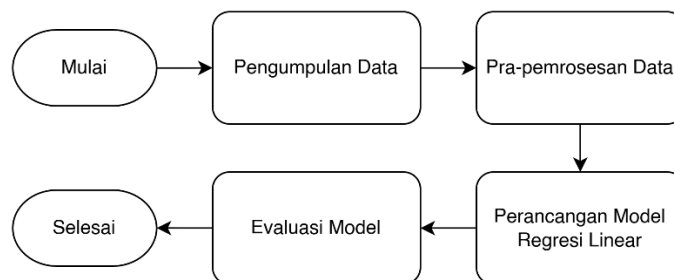
Meskipun regresi linear telah banyak diterapkan, penelitian ini memiliki fokus yang berbeda. Perbedaan utama terletak pada objek prediksi, yaitu kadar α -pinene minyak terpentin berbasis komposisi mutu bahan baku getah pinus. Penelitian ini tidak membandingkan berbagai algoritma, melainkan berfokus pada pengoptimalan regresi linear berganda serta evaluasi model secara komprehensif menggunakan *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan R-squared.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk merancang model prediksi kadar α -pinene dalam minyak terpentin menggunakan regresi linear berganda serta menganalisis pengaruh komposisi mutu getah terhadap kadar yang dihasilkan. Hasil penelitian diharapkan memberikan kontribusi ilmiah dan praktis bagi pengembangan industri terpentin dan pengelolaan bahan baku secara optimal.

II. METODE

A. Tahapan Penelitian

Penelitian ini merupakan penelitian kuantitatif yang menghasilkan temuan melalui prosedur statistik atau metode kuantifikasi lainnya. Dalam penelitian ini, data numerik yang digunakan berupa rekaman input produksi getah dengan berbagai klasifikasi mutu dan output kadar alpha pinene yang dihasilkan. Data ini kemudian diolah menggunakan algoritma regresi linear berganda untuk memprediksi kadar alpha pinene dalam minyak terpenin. Langkah-langkah dalam penelitian ini digambarkan melalui diagram alur pada Gambar 1.



Gambar 1. Alur Penelitian

Alur ini dimulai dengan pengumpulan data, diikuti oleh tahapan pra-pemrosesan data untuk memastikan kualitas data. Selanjutnya, data yang telah bersih digunakan untuk perancangan model regresi linear berganda. Tahap akhir adalah evaluasi model untuk menilai kinerja prediksinya. Diagram alur ini berfungsi sebagai peta jalan visual, memungkinkan pembaca untuk dengan cepat memahami pendekatan metodologis yang terstruktur dan sistematis yang diambil dalam studi ini.

B. Pengumpulan Data

Tahap pertama penelitian ini ialah menghimpun data yang diperlukan untuk membangun model prediksi. Sumbernya berupa data sekunder yang diperoleh dari Departemen Peneliti Produksi Industri & Pemasaran, Perhutani *Forestry Institute*. Dataset ini berisi 229 baris catatan produksi getah dari Januari 2023 sampai Mei 2024, mencakup mutu getah (SP, P, IA, IB, IIA, IIB) dan kadar alpha pinene. Informasi dikumpulkan melalui wawancara langsung dengan pihak institusi untuk memperjelas setiap variabel, ditambah penelaahan laporan dan arsip resmi. Kumpulan data tersebut kemudian dijadikan fondasi dalam pembangunan model regresi linear berganda.

C. Pra-pemrosesan Data

Pra-pemrosesan data adalah langkah awal yang penting dalam proses pengolahan data guna menyiapkan data mentah agar layak dianalisis. Proses ini mencakup penanganan nilai kosong, data tidak konsisten, atau gangguan lain dalam data [12]. Dengan proses ini, kualitas data dapat ditingkatkan sehingga analisis yang dilakukan menjadi lebih tepat sasaran dan hasil model lebih dapat diandalkan. Dengan melalui tahapan ini, data menjadi lebih bersih, rapi, dan siap digunakan dalam proses pemodelan, sehingga hasil analisis yang diperoleh lebih dapat diandalkan. Pada penelitian ini, tahap pra-pemrosesan dimulai dengan pemilihan atribut yang dilakukan secara manual menggunakan Microsoft Excel. Dari seluruh atribut yang tersedia, hanya atribut yang relevan dengan proses prediksi kadar alpha pinene yang dipilih, yakni mutu getah SP, P, IA, IB, IIA, dan IIB sebagai variabel independen, serta kadar alpha pinene sebagai variabel dependen. Atribut-atribut lain yang tidak memiliki hubungan langsung terhadap proses prediksi dihapus guna mengurangi noise dan memastikan fokus pada data yang memiliki nilai prediktif terhadap kadar alpha pinene.

Setelah pemilihan atribut selesai, langkah berikutnya adalah pembersihan data. Tahap pembersihan data dilakukan menggunakan Microsoft Excel dan bahasa pemrograman Python. Langkah pertama dalam pembersihan data ini adalah mengonversikan nilai kadar alpha pinene pada dataset. Pada data asli, kadar alpha pinene dinyatakan dalam bentuk persentase (%), yang kemudian diubah menjadi format desimal. Penyeragaman format ini penting untuk memastikan semua nilai numerik dapat diolah secara konsisten dalam model regresi. Konversi ini dilakukan di Microsoft Excel sebelum data diimpor ke Python. Setelah data berhasil dikonversi, langkah selanjutnya adalah menangani missing value dalam dataset. Keberadaan nilai kosong dapat menyebabkan menurunnya keakuratan dan kualitas data jika tidak ditangani dengan baik [13]. Setelah identifikasi, nilai kosong tersebut diganti dengan nilai 0, mengindikasikan tidak adanya mutu getah yang masuk. Pendekatan ini dipilih karena dalam data bahan baku, ketiadaan data seringkali berarti kuantitas nol. Selain itu, karakter khusus seperti tanda "-" yang sering muncul sebagai penanda kosong juga diganti dengan nilai 0, memberikan interpretasi yang konsisten bahwa tidak ada mutu getah yang masuk pada data tersebut.

D. Perancangan Model Regresi Linear Berganda

Model regresi linear berganda ini dirancang untuk memprediksi kadar alpha pinene berdasarkan variabel input berupa mutu getah. Proses perancangan model melibatkan algoritma regresi linear berganda yang menghubungkan hubungan antara beberapa variabel independen (mutu getah) dan satu variabel dependen (kadar alpha pinene). Model yang dihasilkan diharapkan mampu memberikan prediksi yang sesuai berdasarkan pola data yang ada.

Proses perancangan model ini dimulai setelah data selesai melalui tahap pra-pemrosesan. Setelah proses pembersihan, data dipecah menjadi dua kelompok, yaitu data latih dan data uji. Proporsinya ditetapkan 80% untuk pelatihan dan 20% untuk pengujian. Bagian data latih dimanfaatkan untuk melatih model regresi linear berganda agar model mengenali pola hubungan antarkomponen variabel. Sisa 20% data dialokasikan sebagai data uji guna mengevaluasi kinerja dan kemampuan generalisasi model setelah pelatihan. Strategi pembagian ini bertujuan memastikan model tidak hanya optimal pada data pelatihan, tetapi juga tetap akurat saat dihadapkan pada data baru.

Pembangunan model dilakukan menggunakan algoritma regresi linear berganda. Model ini dibangun dan dilatih menggunakan data latih yang telah dipersiapkan. Algoritma regresi linear digunakan sebagai dasar pemodelan ini. Setelah proses pelatihan selesai, model menghasilkan parameter regresi berupa koefisien regresi dan intersep. Koefisien-koefisien menggambarkan besarnya perubahan nilai target (variabel dependen) akibat kenaikan satu satuan pada setiap variabel independen, dengan asumsi variabel lain tetap konstan. Sementara itu, intersep menunjukkan nilai awal prediksi variabel dependen ketika semua variabel independen bernilai nol, yang berfungsi sebagai titik potong garis regresi pada sumbu Y [14]. Setelah model terbentuk dan koefisien regresi diperoleh, model digunakan untuk melakukan prediksi pada data uji. Nilai hasil prediksi kemudian dibandingkan dengan nilai aktual dari data uji untuk menilai performa model secara objektif. Secara umum, persamaan regresi linear berganda dituliskan sebagai berikut:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (1)$$

Dalam persamaan tersebut, variabel Y merupakan variabel target atau variabel terikat yang ingin diprediksi. Sementara itu, X_1 , hingga X_n adalah variabel bebas yang memengaruhi nilai Y. Nilai b_0 adalah intersep atau konstanta, yang menyatakan nilai Y saat semua variabel X bernilai nol. Sedangkan b_1 hingga b_n adalah koefisien regresi yang menjelaskan pengaruh masing-masing variabel bebas terhadap nilai Y. Koefisien ini menunjukkan seberapa besar perubahan pada Y jika variabel X meningkat satu satuan, dengan asumsi variabel lainnya tidak berubah.

E. Evaluasi Model

Untuk mengukur seberapa baik model memprediksi kadar alpha pinene, dilakukan evaluasi menggunakan sejumlah metrik. Evaluasi ini bertujuan untuk menilai seberapa akurat model dalam memprediksi kadar alpha pinene. Beberapa metrik yang digunakan dalam penelitian ini meliputi MAE, MSE, RMSE, dan R-squared (R^2). Penggunaan berbagai metrik memberikan penilaian yang menyeluruh terhadap kinerja model.

MAE merepresentasikan rata-rata selisih absolut antara nilai prediksi dan nilai aktual tanpa mempertimbangkan arah selisihnya [15]. Rumus MAE adalah:

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (2)$$

Dalam rumus tersebut, y_i menunjukkan nilai aktual dari variabel target, \hat{y}_i merupakan hasil prediksi, dan n adalah jumlah data yang digunakan. Nilai MAE yang lebih kecil menandakan model semakin akurat dalam melakukan prediksi.

Selanjutnya, MSE menghitung rata-rata kuadrat selisih antara nilai aktual dan prediksi [15]. Rumus MSE dituliskan sebagai berikut:

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (3)$$

Dalam persamaan ini, y_i adalah nilai aktual dari variabel dependen, \hat{y}_i merupakan nilai hasil prediksi, n adalah jumlah total data. Karena MSE menggunakan kuadrat dari selisih, maka metrik ini lebih sensitif terhadap kesalahan yang besar dibandingkan dengan MAE. Nilai MSE yang kecil menunjukkan bahwa model memiliki performa prediksi yang baik secara keseluruhan. RMSE merupakan akar kuadrat dari MSE, yang mengembalikan nilai kesalahan ke dalam satuan yang sama dengan variabel aslinya [16]. Rumus RMSE sebagai berikut:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (4)$$

RMSE sering digunakan untuk menginterpretasikan besaran kesalahan prediksi karena satuannya sama dengan variabel yang diprediksi. Nilai RMSE yang kecil menandakan tingginya keakuratan prediksi [17].

Sementara itu, R-squared, atau koefisien determinasi, menunjukkan seberapa dekat data terhadap garis regresi. Nilai ini mengukur proporsi variansi variabel target yang mampu dijelaskan oleh prediktor [17]. Rumus R-squared dituliskan sebagai:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - y_{avg})^2} \quad (5)$$

Dalam rumus tersebut, y_i adalah nilai aktual dari variabel target, \hat{y}_i merupakan nilai prediksi, y_{avg} adalah rata-rata dari seluruh nilai aktual, dan n adalah jumlah data. Nilai R^2 berkisar 0 sampai 1, dimana semakin mendekati 1, artinya semakin banyak variasi data yang berhasil dijelaskan oleh model.

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Dataset Penelitian

Dataset yang digunakan dalam penelitian ini adalah acuan dalam membangun model prediksi kadar alpha pinene dalam minyak terpentin. Data ini bersumber dari laporan produksi milik Perum Perhutani, yang mencakup periode Januari 2023 hingga Mei 2024. Dataset ini terdiri dari variabel independen berupa komposisi bahan baku getah pinus, yang dikategorikan sebagai SP (Super Premium), P (Premium), IA (Mutu IA), IB (Mutu IB), IIA (Mutu IIA), dan IIB (Mutu IIB), yang diukur dalam ton. Variabel dependennya adalah kadar alpha pinene (Y) yang dihasilkan dari proses produksi minyak terpentin, dinyatakan dalam persentase. Tabel 1 menyajikan cuplikan data yang digunakan dalam penelitian ini

TABEL 1
DATA PENGUMPULAN BAHAN BAKU DAN KADAR ALPHA PINENE

Mutu Getah (ton)						Kadar AP (Y)
SP (X1)	P (X2)	IA (X3)	IB (X4)	IIA (X5)	IIB (X6)	
0	0	67,5	231,487	1,013	0	77,72%
-	-	158,758	412,013	6,729	-	77,90%
0	0	110,23	275,722	4,048	0	79,18%
-	-	104,878	323,044	7,078	-	77,71%
0	0	126,68	319,695	3,625	0	78,71%
-	-	165,000	433,495	1,505	-	79,68%
-	-	50,43	332,07	35,00	2,50	76,74%
-	-	60,00	319,74	157,50	2,76	78,13%
-	-	60,000	229,230	175,000	0,77	78,12%
0	0	75,652	268,48	165,001	0,867	78,71%

Tabel 1 menunjukkan representasi dari data masukan mutu getah (komposisi berbagai kualitas getah dalam ton) dan keluaran yang sesuai (kandungan alpha pinene). Dengan data ini, memungkinkan untuk dapat memahami sifat variabel dan format data, yang merupakan dasar untuk memahami analisis dan model selanjutnya. Tabel ini juga secara implisit menunjukkan keberadaan nilai-nilai yang memerlukan pra-pemrosesan.

B. Pra-pemrosesan Data

Sebelum data digunakan untuk pembangunan model, tahap pra-pemrosesan data sangat penting untuk memastikan kualitas data yang baik dan siap diolah. Proses ini bertujuan untuk meningkatkan keakuratan dan efisiensi analisis dengan memperbaiki atau menyaring data yang tidak relevan, tidak lengkap, atau mengandung kesalahan. Dalam proses pra-pemrosesan data, terdapat beberapa tahapan yang dilakukan, yaitu pemilihan atribut dan pembersihan data.

Pada tahap ini, dilakukan pemilihan atribut secara manual menggunakan Microsoft Excel untuk memastikan hanya informasi yang relevan yang digunakan dalam pembangunan model. Dari seluruh atribut yang tersedia dalam

dataset, hanya atribut mutu getah (SP, P, IA, IB, IIA, IIB) yang dipilih sebagai variabel independen, dan kadar alpha pinene sebagai variabel dependen. Atribut-atribut ini dipilih karena memiliki keterkaitan langsung dengan komposisi mutu getah yang berperan dalam menentukan kadar alpha pinene. Atribut lain yang tidak berkaitan langsung dengan prediksi dihapus dari dataset untuk mengurangi noise dan memastikan fokus pada data yang memiliki nilai prediktif terhadap kadar alpha pinene.

Gambar 2 menyajikan data mentah rekapan input getah yang digunakan dalam penelitian. Berdasarkan data asli tersebut, beberapa atribut seperti nomor, periode, satuan, jumlah, terpentin (ton), rendemen (%), total gondorukem (ton), XB, X, WW, Total Rend (%), dan keterangan, tidak digunakan dalam proses pemodelan karena tidak berkontribusi langsung terhadap prediksi kadar alpha pinene. Oleh karena itu, proses seleksi atribut dilakukan dengan mempertahankan variabel SP, P, IA, IB, IIA, dan IIB sebagai variabel prediktor dalam perancangan model. Hasil pemilihan atribut disajikan pada Gambar 3.

No	ERIOD	SAY	GETAH INPUT										PDT Rejojangan										Ket
			SP	P	IA	IB	IIA	IIB	Jumlah	ipentn (%)	rendemen (%)	alpha P (%)	Total	XB	X	WW	Total Rend (%)						
1	JAN I	TON	0	0	67,5	231,487	1,013	0	300,00	14,446	13,82%	77,72%	28,36	0	32,88	185,28	68,54%						
JAN II	TON	-	-	158,758	412,013	6,729	-	577,80	81,085	14,04%	77,90%	439,52	-	180,400	389,120	68,85%							
FEB I	TON	0	0	110,23	275,722	4,048	0	380,00	56,943	14,50%	79,18%	20,92	0	86,88	184,44	68,86%							
FEB II	TON	-	-	104,878	323,044	7,078	-	436,00	65,807	14,97%	77,71%	31,96	0	32,880	220,800	68,95%							
MAR I	TON	0	0	126,68	319,695	3,625	0	450,00	71,699	15,31%	78,71%	20,64	0	95,76	216,68	67,80%							
MAR II	TON	-	-	165,000	433,495	1,505	-	600,00	85,040	15,04%	79,68%	425,76	-	107,840	307,820	68,80%							
APR I	TON	-	-	50,43	332,07	35,00	2,50	420,00	64,39	15,23%	76,74%	239,52	-	28,88	272,04	68,64%							
APR II	TON	-	-	60,000	229,230	175,000	0,77	540,00	83,27	15,43%	78,12%	348,24	-	43,44	341,44	67,05%							
MAY I	TON	0	0	78,952	338,481	855,000	0,827	590,00	78,781	14,42%	79,71%	562,78	0	89,880	220,44	68,68%							
MAY II	TON	-	-	88,905	297,844	52,750	5,20	526,00	61,982	15,50%	79,44%	374,88	0	110,04	261,84	68,80%							
JUN I	TON	0	0	60,000	319,74	157,50	2,76	450,00	83,31	18,23%	78,13%	443,76	-	188,44	244,72	68,84%							
JUN II	TON	-	-	132,82	359,89	532,29	32,41	667,50	106,80	17,26%	77,78%	479,28	-	85,82	327,36	68,95%							
JUL I	TON	-	-	5,60	162,38	432,08	28,47	637,50	89,47	17,07%	80,30%	447,20	-	166,72	300,48	68,85%							
JUL II	TON	-	-	12,40	196,48	277,68	32,76	472,50	76,53	16,20%	77,91%	341,28	-	132,24	209,04	68,84%							
AUG I	TON	-	-	66,88	250,60	432,22	21,8	780,00	87,08	16,28%	79,42%	567,84	###	238,16	290,04	69,05%							
AUG II	TON	-	-	64,05	170,87	338,28	19,21	625,50	86,35	17,05%	79,01%	458,60	###	142,61	288,04	69,71%							
SEP I	TON	-	-	177,88	588,95	1.082,28	54,01	1.082,50	37,46	16,86%	79,64%	1.388,24	###	486,96	853,20	69,50%							
SEP II	TON	-	-	36,82	317,82	476,91	10,62	840,00	138,08	16,44%	77,88%	618,72	-	253,82	266,80	68,05%							
OCT I	TON	-	-	62,80	322,72	472,91	2,67	860,00	86,29	16,90%	78,85%	1.168,00	-	460,20	460,20	68,84%							
NOV I	TON	-	-	123,72	281,80	361,65	0,33	705,00	117,74	16,70%	78,84%	821,04	5,62	172,08	343,44	69,61%							
NOV II	TON	-	-	43,24	393,94	305,68	0,26	670,50	16,48	16,70%	77,00%	514,20	-	144,24	370,56	69,50%							
DES I	TON	-	-	13,89	126,41	87,49	5,53	0,08	232,50	37,13	19,37%	77,88%	170,64	-	55,20	18,44	68,36%						
DES II	TON	-	-	7,88	91,58	72,89	0,45	-	172,80	25,95	14,86%	79,08%	125,28	-	5,82	18,76	67,38%						

Gambar 2. Data Asli Rekapan Input Getah

SP	P	IA	IB	IIA	IIB	Alpha P (%)
0	0	67,5	231,487	1,013	0	77,72%
-	-	158,758	412,013	6,729	-	77,90%
0	0	110,23	275,722	4,048	0	79,18%
-	-	104,878	323,044	7,078	-	77,71%
0	0	126,68	319,695	3,625	0	78,71%
-	-	165,000	433,495	1,505	-	79,68%
-	-	50,43	332,07	35,00	2,50	76,74%
-	-	60,000	319,74	157,50	2,76	78,13%
-	-	60,000	229,230	175,000	0,77	78,12%
0	0	75,652	268,48	165,001	0,867	78,71%

Gambar 3. Hasil Pemilihan Atribut Dataset

Setelah atribut yang relevan berhasil dipilih, selanjutnya melakukan pembersihan data. Proses ini dilakukan menggunakan Microsoft Excel dan bahasa pemrograman Python. Pada tahap pembersihan data ini, yang dilakukan pertama adalah mengonversikan nilai kadar alpha pinene pada dataset. Pada data asli, kadar alpha pinene dinyatakan dalam bentuk persentase (%), yang kemudian diubah menjadi format desimal. Penyeragaman format ini penting untuk memastikan semua nilai numerik dapat diolah secara konsisten dalam model regresi. Konversi ini dilakukan di Microsoft Excel sebelum data diimpor ke Python. Hasil konversi disajikan pada Gambar 4.

SP	P	IA	IB	IIA	IIB	Alpha P (%)
0	0	67,5	231,487	1,013	0	0,7772
-	-	158,758	412,013	6,729	-	0,779
0	0	110,23	275,722	4,048	0	0,7918
-	-	104,878	323,044	7,078	-	0,7771
0	0	126,68	319,695	3,625	0	0,7871
-	-	165,000	433,495	1,505	-	0,7968
-	-	50,43	332,07	35,00	2,50	0,7674
-	-	60,000	319,74	157,50	2,76	0,7813
-	-	60,000	229,230	175,000	0,77	0,7812
0	0	75,652	268,48	165,001	0,867	0,7871

Gambar 4. Hasil Konversi Kadar Alpha Pinene

Setelah data berhasil dikonversi ke dalam desimal, langkah selanjutnya adalah mengimpor dataset yang akan digunakan ke Python menggunakan pustaka Pandas. Setelah dataset berhasil diimpor, selanjutnya adalah menangani *missing value* pada dataset. Untuk mendeteksi keberadaan *missing value*, digunakan fungsi `isna()` yang akan menunjukkan kolom mana saja yang memiliki nilai kosong. Kemudian, fungsi `sum()` digunakan untuk menghitung

jumlah *missing value* pada setiap kolom. Hasilnya menunjukkan adanya *missing value* pada kolom mutu getah IIA sebanyak 1 dan mutu getah IIB sebanyak 22 data, sebagaimana ditunjukkan pada Gambar 5(a). Nilai-nilai ini kemudian diganti dengan 0 menggunakan fungsi `fillna()`, mengindikasikan tidak adanya mutu getah pada entri tersebut. Hasil dari penggantian *missing value* dengan nilai 0 ditampilkan pada Gambar 5(b).

```
# Menghitung missing values
df_getah.isna().sum()
```

	0
SP	0
P	0
IA	0
IB	0
IIA	1
IIB	22
Alpha P (%)	0

(a)

```
# Menghitung missing values
df_getah.isna().sum()
```

	0
SP	0
P	0
IA	0
IB	0
IIA	0
IIB	0
Alpha P (%)	0

(b)

Gambar 5. Menangani missing value pada data (a) Data mengandung missing value (b) Hasil penanganan missing value

Tahap selanjutnya dalam pembersihan data adalah mengganti karakter khusus seperti tanda "-" pada data dengan nilai 0. Hal ini dilakukan untuk memastikan bahwa dataset tidak mengandung simbol atau karakter yang dapat mengganggu proses analisis data. Karakter "-" sering muncul dalam dataset sebagai penanda kosong atau tidak ada data, sehingga menggantinya dengan nilai 0 memberikan arti bahwa tidak ada mutu getah yang tercatat pada entri tersebut. Proses penggantian dilakukan menggunakan fungsi `replace()` pada pustaka Pandas. Hasil pembersihan data setelah penggantian karakter khusus disajikan pada Gambar 6.

```
# Mengubah nilai - dengan 0
df_getah.replace(['-'], 0, inplace=True)
df_getah.head()
```

	SP	P	IA	IB	IIA	IIB	Alpha P (%)
0	0	0	67.5	231.487	1.013	0	0.7772
1	0	0	158.758	412.013	6.729	0	0.7790
2	0	0	110.23	275.722	4.048	0	0.7918
3	0	0	104.878	323.044	7.078	0	0.7771
4	0	0	126.68	319.695	3.625	0	0.7871

Gambar 6. Mengganti Karakter “-” dengan 0

Gambar 6 menunjukkan bahwa data yang sebelumnya mengandung karakter “-” telah berhasil dikonversi menjadi nilai nol, yang artinya tidak ada mutu getah yang masuk. Selain itu, nilai kosong pun telah berhasil ditangani yang ditampilkan pada Gambar 5(b). Oleh karena itu, setelah semua langkah pembersihan data dijalankan, dataset menjadi seragam, lengkap, dan siap untuk dianalisis.

C. Perancangan Model Regresi Linear Berganda

Pada tahap ini, proses perancangan model dimulai dengan membagi dataset menjadi dua subset, yaitu data latih (80%) dan data uji (20%). Langkah pertama dari proses ini adalah memisahkan variabel bebas (X) yakni mutu getah SP, P, IA, IB, IIA, dan IIB dari variabel terikat (y) berupa kadar alpha pinene. Porsi 80% dialokasikan untuk melatih model agar dapat mengenali pola hubungan antarkomponen data, sedangkan 20% sisanya disimpan untuk menilai kinerja model pada sampel yang belum pernah dilihatnya. Dari proses pembagian ini diperoleh 183 data latih dan 46 data uji. Pembagian dataset dilakukan secara acak dengan *random seed* tetap untuk memastikan reproduibilitas hasil. Ilustrasi proses pembagian dataset disajikan pada Gambar 7.

```
# Persiapan Fitur dan Target
X = df_getah[[' SP ', ' P ', ' IA ', ' IB ', ' IIA ', ' IIB ']] # Variabel independent (fitur)
y = df_getah[' Alpha P (%) ' ] # Variabel dependent (target)

# Pembagian Dataset
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)
```

Gambar 7. Code Pembagian Dataset

Merujuk potongan kode pada Gambar 7, fungsi `train_test_split` dari pustaka `scikit-learn` digunakan untuk memisahkan data. Parameter `test_size=0.2` menandakan 20% data dijadikan set uji, sedangkan `random_state=0` digunakan agar pembagian data bersifat tetap ketika kode dijalankan ulang. Fungsi ini menghasilkan empat keluaran `X_train` dan `y_train` untuk pelatihan, serta `X_test` dan `y_test` untuk pengujian.

Setelah proses pembagian dataset selesai, langkah berikutnya adalah membangun model regresi linear berganda dengan menggunakan data latih yang telah disiapkan. Model ini dikembangkan menggunakan algoritma regresi linear dari pustaka `Scikit-learn`. Proses diawali dengan membuat objek model `LinearRegression()` dan menyimpannya ke dalam variabel `reg`. Selanjutnya, model dilatih dengan menjalankan fungsi `fit()` menggunakan `X_train` sebagai input variabel bebas dan `y_train` sebagai output target. Setelah proses pelatihan selesai, model akan menghasilkan parameter berupa koefisien regresi untuk masing-masing variabel dan nilai intersep sebagai konstanta. Implementasi proses pelatihan model disajikan pada Gambar 8.

```
#Training Algoritma
reg = LinearRegression()
reg.fit(X_train, y_train)
coef = reg.coef_
intercept = reg.intercept_
print(coef)
print(intercept)
```

Gambar 8. Code Pelatihan Model Regresi Linear Berganda

Setelah proses pelatihan selesai, diperoleh parameter regresi berupa koefisien regresi (`coef_`) dan intersep (`intercept_`). Nilai intersep yang diperoleh adalah 0,8172103. Nilai ini menunjukkan bahwa ketika semua variabel independen (SP, P, IA, IB, IIA, dan IIB) bernilai nol, kadar alpha pinene yang diprediksi oleh model adalah 0,8172103. Dengan kata lain, nilai ini merepresentasikan titik awal prediksi sebelum ada pengaruh dari variabel-variabel independen. Selain itu, diperoleh koefisien regresi untuk masing-masing variabel independen sebagai berikut: SP= -0,0003138; P= -0,0000122; IA= -0,0000030; IB= -0,0000241; IIA= -0,0001369; IIB= -0,0006079.

Koefisien-koefisien ini menunjukkan bagaimana setiap variabel independen berkontribusi terhadap perubahan kadar alpha pinene yang diprediksi. Seluruh koefisien bernilai negatif, yang berarti bahwa peningkatan nilai pada variabel SP, P, IA, IB, IIA, dan IIB akan menyebabkan penurunan kadar alpha pinene yang diprediksi oleh model, dengan asumsi variabel lainnya tetap konstan. Variabel IIB memiliki dampak paling besar terhadap kadar alpha pinene, dengan nilai koefisien sebesar -0,0006079. Ini menunjukkan bahwa peningkatan satu unit pada variabel IIB akan menyebabkan penurunan kadar alpha pinene sebesar 0,0006079 unit. Sebaliknya, variabel IA memiliki dampak paling kecil, dengan nilai koefisien sebesar -0,0000030, yang berarti perubahan dalam variabel ini hanya sedikit memengaruhi kadar alpha pinene yang diprediksi. Koefisien P (-0,0000122) juga menunjukkan pengaruh yang sangat kecil, artinya perubahan pada variabel ini tidak terlalu berdampak terhadap kadar alpha pinene dalam model ini. Variabel SP (-0,0003138) memiliki pengaruh yang lebih besar dibanding P, IA, dan IB, IIA, dan IIB, menunjukkan bahwa semakin tinggi nilai SP, semakin rendah kadar alpha pinene yang diprediksi oleh model.

Namun, karena nilai koefisien yang diperoleh relatif kecil, hal ini menunjukkan bahwa hubungan antara variabel-variabel independen dengan kadar alpha pinene tidak terlalu kuat. Ini berarti setiap perubahan pada variabel independen hanya memberikan dampak yang kecil terhadap kadar alpha pinene yang diprediksi. Hal ini dapat

mengindikasikan bahwa faktor-faktor lain yang tidak dimasukkan dalam model mungkin memiliki pengaruh yang lebih besar terhadap variasi kadar alpha pinene.

D. Perhitungan Nilai Prediksi

Setelah persamaan regresi linear berganda diperoleh, langkah selanjutnya adalah melakukan perhitungan nilai prediksi kadar alpha pinene menggunakan data uji. Data uji ini merupakan 20% dari total dataset, atau berjumlah 46 data. Untuk memberikan gambaran proses prediksi, sebuah contoh perhitungan disajikan menggunakan salah satu data uji: $X_1 = 0$; $X_2 = 16,72$; $X_3 = 354,73$; $X_4 = 41,25$; $X_5 = 0$; $X_6 = 0$.

Dengan menggunakan persamaan (2.1), persamaan regresi linear berganda yang diperoleh dalam penelitian ini, yaitu:

$$Y = 0,8172103 - 0,0003138X_1 - 0,0000122X_2 - 0,0000030X_3 - 0,0000241X_4 - 0,0001369X_5 - 0,0006079X_6$$

Maka nilai prediksi kadar alpha pinene (Y) untuk data tersebut dihitung sebagai berikut :

$$Y = 0,8172103 - (0,0003138 * 0) - (0,0000122 * 16,72) - (0,0000030 * 354,73) - (0,0000241 * 41,25) - (0,0001369 * 0) - (0,0006079 * 0)$$

$$Y = 0,8150$$

Jadi, kadar alpha pinene yang diprediksi untuk data tersebut adalah 81,50%. Selain contoh tersebut, Tabel 2 berikut menampilkan cuplikan hasil prediksi dari sebagian data uji yang telah dihitung menggunakan model regresi linear berganda yang telah dibentuk:

TABEL 2
 HASIL PREDIKSI KADAR ALPHA PINENE

No	Mutu Getah (ton)						Y_Test (Kadar AP Aktual)	Y_Pred (Kadar AP Prediksi)
	X1 (SP)	X2 (P)	X3 (IA)	X4 (IB)	X5 (IIA)	X6 (IIB)		
1	0	16,72	354,73	41,25	0	0	0,833	0,8150
2	0	0	81,6	753,557	1,243	0	0,8079	0,7987
3	0	23,304	109,942	331,039	0	0	0,836	0,8086
4	0	0	470,96	19,5	0	0	0,820515	0,8153
5	0	0	7,365	286,3	0	0	0,7967	0,8103
...
46	0	0	147,9	494,7	0	0	0,7814	0,8049

E. Evaluasi Model

Setelah model regresi linear berhasil dibuat dan menghasilkan prediksi, langkah selanjutnya adalah mengevaluasi performa model untuk mengetahui seberapa baik model dalam memprediksi kadar alpha pinene. Evaluasi model dilakukan menggunakan pustaka metrics dari sklearn.metrics dalam Python, dengan membandingkan nilai aktual (y_{test}) dengan nilai prediksi (y_{pred}). Implementasi perhitungan metrik evaluasi disajikan pada Gambar 9, sedangkan hasil perhitungan metrik evaluasi model ditampilkan pada Gambar 10.

```
#Evaluasi Model
print('Mean Absolute Error: {:.5f}'.format(metrics.mean_absolute_error(y_test, y_pred)))
print('Mean Squared Error: {:.5f}'.format(metrics.mean_squared_error(y_test, y_pred)))
print('Root Mean Squared Error: {:.5f}'.format(np.sqrt(metrics.mean_squared_error(y_test, y_pred))))
print('R2 Score: {:.5f}'.format(metrics.r2_score(y_test, y_pred)))
```

Gambar 9. Code Evaluasi Model

```
Mean Absolute Error: 0.01312
Mean Squared Error: 0.00026
Root Mean Squared Error: 0.01622
R2 Score: 0.20854
```

Gambar 10. Hasil Evaluasi Model

Dari hasil evaluasi yang disajikan pada Gambar 10, dapat dianalisis seberapa baik model bekerja berdasarkan empat metrik utama. Berikut penjabaran masing-masing metrik dan implikasinya terhadap kualitas model.

1) *Mean Absolute Error (MAE)*: nilai MAE sebesar 0,01312 menunjukkan bahwa rata-rata selisih absolut antara hasil prediksi dan nilai aktual kadar alpha pinene adalah sekitar 0,01312. Nilai ini tergolong rendah, yang berarti model memiliki tingkat kesalahan yang cukup kecil saat memprediksi kadar alpha pinene.

2) *Mean Squared Error (MSE)*: nilai MSE sebesar 0,00026 mengindikasikan bahwa rata-rata kesalahan kuadrat dalam prediksi kadar alpha pinene juga rendah. Karena MSE menghitung kuadrat dari selisih prediksi dan aktual, metrik ini lebih peka terhadap kesalahan besar. Oleh karena itu, nilai yang kecil menunjukkan bahwa model memiliki kinerja yang baik secara keseluruhan.

3) *Root Mean Squared Error (RMSE)*: nilai RMSE sebesar 0,01622 menunjukkan bahwa rata-rata kesalahan model dalam memprediksi kadar alpha pinene masih dalam batas yang dapat diterima. Karena RMSE memiliki satuan yang sama dengan data aslinya, metrik ini sering digunakan untuk memahami skala kesalahan model secara lebih intuitif dibandingkan MSE.

4) *R-Squared (R^2)*: nilai R^2 sebesar 0,20854 menunjukkan bahwa model hanya mampu menjelaskan sekitar 20,854% variasi dari data kadar alpha pinene. Nilai ini cukup rendah, yang mengindikasikan bahwa masih terdapat faktor lain yang mempengaruhi kadar alpha pinene namun belum terakomodasi dalam model regresi linear yang digunakan.

Nilai R-squared (R^2) sebesar 0,20854 menunjukkan bahwa model regresi linear berganda hanya mampu menjelaskan sekitar 20,85% variasi data kadar α -pinene. Nilai ini tergolong rendah dan mengindikasikan bahwa selain variabel komposisi mutu getah yang digunakan dalam penelitian ini, terdapat faktor-faktor lain yang turut memengaruhi kadar α -pinene namun belum terakomodasi dalam model. Temuan ini menunjukkan bahwa hubungan antara komposisi bahan baku dan kadar α -pinene bersifat kompleks dan tidak sepenuhnya dapat direpresentasikan oleh model linier sederhana.

Kadar alpha pinene dalam minyak terpentin dipengaruhi oleh berbagai faktor, baik di tingkat hulu (pada saat pengambilan bahan baku) maupun di tingkat hilir (pada saat pengolahan di Pabrik Gondorukem dan Terpentin - PGT) [18]. Pada tingkat hulu, kadar α -pinene dalam minyak terpentin dipengaruhi oleh faktor genetik dan lingkungan tanaman pinus. Variasi genetik antar pohon, perbedaan spesies atau klon pinus, serta kondisi lingkungan seperti ketinggian tempat, jenis tanah, dan iklim diketahui berkontribusi terhadap perbedaan komposisi kimia getah [19]. Selain itu, waktu dan metode penyadapan getah juga berpengaruh, karena proses penguapan senyawa volatil seperti α -pinene dapat terjadi apabila getah tidak segera diproses atau disimpan dalam kondisi yang tidak optimal [20].

Sementara itu, pada tahap hilir mutu getah yang diklasifikasikan ke dalam beberapa kategori, seperti Super Premium (SP), Premium (P), IA, IB, IIA, dan IIB, menunjukkan variasi kadar α -pinene yang berbeda. Namun, proses pengolahan di Pabrik Gondorukem dan Terpentin (PGT), seperti suhu dan tekanan distilasi, teknik pemanasan, serta durasi penyimpanan sebelum pengolahan, juga berpotensi memengaruhi kadar α -pinene yang dihasilkan. Ketidaktepatan pengaturan parameter proses dapat menyebabkan degradasi atau penguapan α -pinene, sehingga menurunkan kualitas produk akhir.

Dengan demikian, rendahnya nilai R^2 pada penelitian ini dapat dipahami sebagai konsekuensi dari keterbatasan variabel yang digunakan, yang hanya merepresentasikan komposisi mutu bahan baku getah. Penelitian ini tidak bertujuan untuk memodelkan seluruh faktor hulu dan hilir secara menyeluruh, melainkan sebagai studi awal dalam pengembangan model prediksi kadar α -pinene berbasis data mutu bahan baku. Untuk meningkatkan akurasi model, penelitian selanjutnya dapat mempertimbangkan penambahan variabel yang merepresentasikan faktor lingkungan dan parameter proses, serta penerapan metode pemodelan yang lebih kompleks, seperti regresi nonlinier atau pendekatan pembelajaran mesin.

IV. SIMPULAN

Berdasarkan penelitian yang dilakukan untuk memprediksi kadar α -pinene pada minyak terpentin menggunakan regresi linear berganda, dapat ditarik beberapa kesimpulan. Pertama, model regresi linear berganda berhasil dikembangkan untuk memprediksi kadar alpha pinene berdasarkan komposisi mutu bahan baku getah pinus, yaitu SP, P, IA, IB, IIA, dan IIB. Implementasi model dilakukan menggunakan bahasa pemrograman Python dengan bantuan pustaka komputasi ilmiah dan analisis data, sehingga proses pengolahan data, pelatihan model, dan evaluasi dapat dilakukan secara terstruktur. Pendekatan ini menunjukkan bahwa regresi linear dapat digunakan sebagai model prediksi awal (*baseline model*) yang bersifat interpretable dan aplikatif dalam lingkungan industri. Kedua, hasil pemodelan menunjukkan bahwa seluruh variabel bebas memiliki nilai koefisien regresi negatif dengan pengaruh yang relatif kecil terhadap kadar α -pinene. Hal ini mengindikasikan bahwa peningkatan jumlah getah pada setiap kategori mutu tidak secara langsung meningkatkan kadar α -pinene secara signifikan. Secara teoritis, kondisi ini dapat terjadi karena kadar α -pinene tidak hanya dipengaruhi oleh komposisi mutu bahan baku, tetapi juga oleh faktor-faktor lain

yang belum dimodelkan secara komputasional, seperti faktor genetik pohon pinus, kondisi lingkungan tumbuh (iklim, ketinggian, dan jenis tanah), waktu dan metode penyadapan getah, serta parameter proses distilasi pada tahap pengolahan. Ketiga, evaluasi kinerja model yang diimplementasikan dalam Python menunjukkan nilai MAE sebesar 0,01312, MSE sebesar 0,00026, dan RMSE sebesar 0,01622 yang mengindikasikan tingkat kesalahan prediksi relatif kecil. Namun, nilai R-squared (R^2) sebesar 0,20854 menunjukkan bahwa model hanya mampu menjelaskan sekitar 20,85% variasi kadar α -pinene. Nilai ini menunjukkan keterbatasan model regresi linear dalam merepresentasikan hubungan yang kompleks antara komposisi bahan baku dan kadar α -pinene, sehingga model masih memiliki keterbatasan dalam menjelaskan pola data, meskipun performa prediksi dasarnya cukup stabil. Sebagai implikasi, penelitian ini memberikan kontribusi di bidang Informatika melalui penerapan pemodelan prediktif berbasis Python untuk mendukung pengambilan keputusan pada industri pengolahan terpentin. Model yang dikembangkan dapat digunakan sebagai alat bantu analisis awal berbasis data. Untuk meningkatkan akurasi dan kemampuan penjelasan model, penelitian selanjutnya disarankan untuk mengintegrasikan variabel hulu dan hilir ke dalam sistem pemodelan, serta mengeksplorasi pendekatan komputasional yang lebih kompleks, seperti regresi nonlinier atau metode pembelajaran mesin.

UCAPAN TERIMAKASIH

Ucapan terima kasih disampaikan kepada Bapak Dr. Masy Ari Ulinuha, S.T., M.T. dan Bapak Nur Cahyo Hendro Wibowo, M.Kom. selaku dosen pembimbing atas bimbingan, arahan, dan motivasi selama proses penelitian ini. Ucapan terima kasih juga disampaikan kepada Bapak Iwan Gunawan, S.Hut., M.Si. dari Perhutani Forestry Institute atas bantuan data dan informasi yang diberikan. Selain itu, apresiasi disampaikan kepada Universitas Islam Negeri Walisongo Semarang atas fasilitas dan dukungan yang diberikan selama penelitian berlangsung. Semoga penelitian ini dapat memberikan manfaat dan menjadi bahan pertimbangan bagi penelitian selanjutnya.

DAFTAR PUSTAKA

- [1] J. Siregar, "Ekstraksi Minyak Terpentin Dari Getah Pinus," Teknik Kimia, Universitas Sumatera Utara, Sumatera Utara, 2020.
- [2] N. Wijayanti, K. Rakainsa, O. Pratama, and C. Sumawi, "Transformasi Alfa Pinena Dari Minyak Terpentin Melalui Reaksi Metoksilasi," *Pemanfaatan Sumber Daya Alam Indonesia: Ketahanan Pangan, Energi Dan Material Maju*, pp. 95–123, 2021.
- [3] I. Gunawan, "Kajian Kandungan Alpha-Pinene Minyak Terpentin Di Pgt Rejowinangun- Trenggalek," Cepu, 2024.
- [4] A. Aris, "Produktivitas Kerja Penyadapan Getah Pinus Di Kph Mamasa Tengah Sulawesi Barat," Program Studi Kehutanan, Universitas Hassanuddin, Makassar, 2021.
- [5] S. Rosalinda and R. Sumirat, "Analisis Minyak Terpentin Hasil Penyulingan di Pabrik Gondorukem dan Terpentin Sindangwangi," *Metana: Media Komunikasi Rekayasa Proses dan Teknologi Tepat Guna*, vol. 20, no. 1, pp. 49–56, 2024.
- [6] I. Muthahharah and Inayanti Fatwa, "Analisis Regresi Linear Berganda Untuk Media Pembelajaran Daring Terhadap Prestasi Belajar Mahasiswa di STKIP Pembangunan," *Jurnal MSA (Matematika dan Statistika serta Aplikasinya)*, vol. 10, no. 1, pp. 53–60, Jun. 2022, doi: 10.24252/msa.v10i1.25145.
- [7] S. A. Listina, R. M. Sampurno, D. Ciptaningtyas, and A. Thoriq, "Model Prediksi Kadar Air Media Tanam Menggunakan Regresi Linear Berganda (Studi Kasus Kebun Tomat Beef di Serenity Farm Mitra Habibi Garden)," *Jurnal Teknotan*, vol. 16, no. 3, p. 133, Dec. 2022, doi: 10.24198/jt.vol16n3.1.
- [8] T. Jaelani, "Machine Learning untuk Prediksi Produksi Gula Nasional," *JMPM (Jurnal Material dan Proses Manufaktur)*, vol. 6, no. 1, Aug. 2022, doi: 10.18196/jmpm.v6i1.14897.
- [9] I. Yanni, "Sistem Prediksi Jumlah Obat Keluar Menggunakan Metode Regresi Linier untuk Menentukan Jumlah Pemesanan Obat," Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, 2020.
- [10] M. A. Fajar, "Implementasi Algoritma Linear Regression dan Autoregressive Integrated Moving Average dalam Memprediksi Harga Rumah (Studi Kasus Rumah di Kabupaten Bogor)," Teknik Informatika, UIN Syarif Hidayatullah, Jakarta, 2023.
- [11] Y. Herdiana and W. A. Geraldine, "Penerapan Machine Learning Dengan Model Linear Regression Terhadap Analisis Kualitas Hasil Petik The Di Pt. Perkebunan Nusantara VIII Kebun Sedep," *COMPUTING | Jurnal Informatika*, vol. 9, no. 01, pp. 1–9, Jun. 2022, doi: 10.55222/computing.v9i01.855.
- [12] A. Prasetyo, S. Salahuddin, and A. Amirullah, "Prediksi Produksi Kelapa Sawit Menggunakan Metode Regresi Linier Berganda," *Jurnal Infomedia*, vol. 6, no. 2, p. 76, Dec. 2021, doi: 10.30811/jim.v6i2.2343.
- [13] N. D. Irawan, W. Wijono, and O. Setyawati, "Perbaikan Missing value Menggunakan Pendekatan Korelasi Pada Metode K-Nearest Neighbor," *JURNAL INFOTEL*, vol. 9, no. 3, Aug. 2017, doi: 10.20895/infotel.v9i3.286.
- [14] D. Kurniawan, *Regresi Linier*, 3rd ed., vol. 2. Statistic, 2008.
- [15] I. M. S. Aji, "Prediksi Data Sensor Kelembapan Tanah yang Rusak Menggunakan Algoritma K-Mean dan Support Vector Regression," Program Studi Teknik Informatika, Universitas Islam Negeri Syarif Hidayatullah, Jakarta, 2023.
- [16] I. Id, *MACHINE LEARNING : Teori, Studi Kasus dan Implementasi Menggunakan Python*, 1st ed., vol. 978-632-255-092–6. UR PRESS, 2021.
- [17] F. H. Hamdanah and D. Fitrihanah, "Analisis Performansi Algoritma Linear Regression dengan Generalized Linear Model untuk Prediksi Penjualan pada Usaha Mikro, Kecil, dan Menengah," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 10, no. 1, p. 23, Apr. 2021, doi: 10.23887/janapati.v10i1.31035.
- [18] Tim Peneliti Prodisar, "Kajian Kadar Alphaspinen Minyak Terpentin Di PGT Rejowinangun Trenggalek," 2024.
- [19] Q. Luan, S. Diao, H. Sun, X. Ding, and J. Jiang, "Prediction and Comparisons of Turpentine Content in Slash Pine at Different Slope Positions Using Near-Infrared Spectroscopy," *Plants*, vol. 11, no. 7, p. 914, Mar. 2022, doi: 10.3390/plants11070914.
- [20] J. Liang *et al.*, "Influence of a novel resin tapping method on the resin yield and turpentine chemical composition of Masson pine (*Pinus massoniana*)," *Ind Crops Prod*, vol. 204, p. 117377, Nov. 2023, doi: 10.1016/j.indcrop.2023.117377.