

Implementasi Data Mining Memprediksi Penjualan Crude Palm Oil Berdasarkan Kapasitas Tangki Menggunakan Multiple Linear Regression

Ana Komaria Baskara*, Alwis Nazir, Muhammad Irsyad, Yusra, Fitri Insani

Fakultas Sains dan Teknologi, Program Studi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ^{1,*}11950121672@students.uin-suska.ac.id, ²alwis.nazir@uin-suska.ac.id, ³muhammad.irsyad@uin-suska.ac.id, ⁴yusra@uin-suska.ac.id, ⁵fitri.insani@uin-suska.ac.id

Email Penulis Korespondensi: 11950121672@students.uin-suska.ac.id

Submitted: 05/02/2023; Accepted: 26/03/2023; Published: 31/03/2023

Abstrak—Data mining adalah suatu proses penggalian informasi dari data yang dapat digunakan untuk meningkatkan bisnis, pengembangan produk, dan pengambilan keputusan lainnya. Salah satu penerapan Data mining yaitu pada PT. Kerry Sawit Indonesia yang merupakan perusahaan agribisnis Wilmar Grup yang bergerak dibidang pengolahan minyak kelapa sawit mentah atau Crude Palm Oil (CPO). Penjualan Crude Palm Oil (CPO) merupakan hal yang sangat penting bagi perusahaan perkebunan kelapa sawit. Dalam rangka meningkatkan efisiensi dan profitabilitas, perusahaan perkebunan kelapa sawit dapat melakukan prediksi penjualan CPO untuk mengoptimalkan penjualan dan persediaan CPO. Salah satu metode yang dapat digunakan untuk melakukan prediksi penjualan CPO adalah dengan menggunakan teknik data mining. Pada penelitian ini, teknik data mining yang digunakan adalah multiple linear regression. Multiple linear regression digunakan untuk mengetahui hubungan antara variabel kapasitas tangki dengan penjualan CPO. Data yang digunakan dalam penelitian ini adalah data produksi CPO, penjualan CPO dan kapasitas tangki yang diperoleh dari perusahaan perkebunan kelapa sawit selama lima tahun terakhir. Hasil perhitungan Multiple Linear Regression pada contoh kasus ini menunjukkan bahwa nilai koefisien determinasi (R-squared) adalah 0,9546, yang berarti bahwa 95,46% variabilitas pengiriman CPO dapat dijelaskan oleh variabel independen. Selain itu, uji MAPE dan RMSE menunjukkan bahwa model regresi yang diperoleh memiliki tingkat keakuratan yang baik dalam memprediksi pengiriman CPO. Oleh karena itu, model regresi ini dapat digunakan untuk memprediksi pengiriman CPO di masa depan dengan memperhatikan nilai-nilai variabel independen yang telah ditentukan.

Kata Kunci: Data Mining; Prediksi; Crude Palm Oil (CPO); Penjualan; Multiple Linear Regression

Abstract—Data mining is a process of discovering information from data that can be used to improve business, product development, and other decision-making processes. One application of data mining is in PT. Kerry Sawit Indonesia, which is an agribusiness company in the Wilmar Group that deals with processing crude palm oil (CPO). Sales of CPO are crucial for palm oil plantation companies. To increase efficiency and profitability, palm oil plantation companies can predict CPO sales to optimize sales and CPO inventory. One method that can be used to predict CPO sales is through data mining techniques. In this study, the data mining technique used is multiple linear regression. Multiple linear regression is used to determine the relationship between the tank capacity variable and CPO sales. The data used in this study are CPO production data, CPO sales data, and tank capacity data obtained from palm oil plantation companies over the last five years. The results of the Multiple Linear Regression calculation in this case study show that the coefficient of determination (R-squared) value is 0.9546, indicating that 95.46% of the CPO delivery variability can be explained by the independent variables. Additionally, the MAPE and RMSE tests show that the regression model obtained has good accuracy in predicting CPO deliveries. Therefore, this regression model can be used to predict CPO deliveries in the future, considering the predetermined independent variable values.

Keywords: Data Mining; Prediction; Crude Palm Oil (CPO); Despatches; Multiple Linear Regression

1. PENDAHULUAN

PT. Kerry Sawit Indonesia merupakan perusahaan agribisnis Wilmar Grup yang bergerak di bidang pengolahan minyak kelapa sawit mentah atau Crude Palm Oil (CPO). CPO atau minyak kelapa sawit mentah merupakan jenis minyak yang diperoleh dari sebuah pohon kelapa sawit. Biasanya digunakan untuk berbagai produk makanan, minyak masakan, makanan ringan, sabun, lilin, kosmetik, dan juga bisa digunakan sebagai sumber biofuel atau biodiesel [1]. CPO merupakan sumber yang kaya beta-karotena, dan tinggi antioksidan. Kebanyakan minyak kelapa sawit diproduksi di Asia, Afrika dan Amerika Selatan karena pohon kelapa sawit membutuhkan suhu hangat, sinar matahari dan curah hujan yang tinggi untuk memaksimalkan produksi [2].

Proses pengolahan produksi Crude Palm Oil pada PT. Kerry Sawit Indonesia dilakukan setiap hari yang masuk ke tangki penyimpanan (storage tank). Tangki penyimpanan sebagai tempat penyimpanan sementara sebelum penjualan dilakukan. PT. Kerry Sawit Indonesia memiliki 2 tangki penyimpanan masing-masing memiliki kapasitas 1.200 ton, apabila tangki penyimpanan hampir penuh akan terjadi danger [3]. Penjualan CPO bergantung pada ketersediaan transporter atau mobil yang mengangkut CPO ke pelabuhan. Jika terjadi over kapasitas [4] maka akan dijual terlebih dahulu minimal 200 ton sekitar 12 mobil truck tangki dengan kapasitas 18-25 ton per unit tangki untuk mengamankan level storage tank.

Menurut Adi Baskara sebagai Supervisor Proses pada PT. Kerry Sawit Indonesia seiring berjalannya waktu dan produksi maka semakin banyak minyak kelapa sawit mentah atau Crude Palm Oil yang dihasilkan dan harus segera dijual atau despatches. Sudah memenuhi kapasitas maksimal maka pabrik harus stop produksi. Hal ini

menjadi sebuah masalah serius dalam memprediksi Crude Palm Oil berdasarkan kapasitas tangki [5] agar tidak terjadi danger pada storage tank. Data mining dapat dijadikan sebagai solusi untuk menyelesaikan permasalahan danger pada storage tank.

Data mining merupakan suatu proses penggalian informasi dari data yang besar dan kompleks menggunakan algoritma statistik dan teknik lainnya [6]. Sehingga dapat ditemukan pola dan hubungan yang sebelumnya tidak diketahui dalam data yang dapat digunakan untuk meningkatkan bisnis, pengembangan produk, dan pengambilan keputusan lainnya. Data mining dibagi menjadi beberapa teknik atau metode yaitu Estimation (estimasi), Forecasting (prediksi), Classification (klasifikasi), Clustering (pengelompokan), Association (asosiasi). Data mining sering digunakan untuk prediksi dengan menggunakan beberapa algoritma, pada penelitian ini algoritma yang akan digunakan untuk mengelola data adalah Multiple Linear Regression [7].

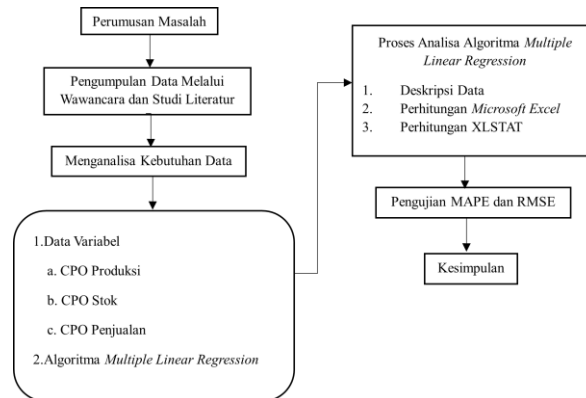
Berikut beberapa penelitian tentang minyak kelapa sawit: Data Mining Untuk Memprediksi Penjualan Buah Sawit Menggunakan Metode Multi Regresi pada PT. Karya Hevea Indonesia dengan melakukan penelitian menggunakan metode algoritma multi regresi linear. Hasil yang diperoleh untuk mempermudah dalam memprediksi penjualan buah kelapa sawit dengan menggunakan software Visual Studio Code [8]. Penelitian lain Penerapan Data Mining Dalam Prediksi Hasil Produksi Kelapa Sawit PT Borneo Ketapang Indah Menggunakan Metode Linier Regression dengan melakukan penelitian menggunakan metode Linear Regression. Diimplementasikan menggunakan tools RapidMiner. Hasil yang diperoleh untuk produksi kelapa sawit yaitu variable sebesar 0,646 untuk variable luas tanah dan variable curah hujan sebesar 0,530 [9].

Berdasarkan penjelasan di atas, penerapan data mining dalam memprediksi crude palm oil menggunakan metode Multiple Linear Regression dan pengimplementasian menggunakan Microsoft Excel untuk pengujian menggunakan MAPE (Mean Absolute Percentage Error) dan RMSE (Root Mean Squared Error). Tujuan penelitian ini adalah untuk memprediksi penjualan crude palm oil berdasarkan kapasitas tangki penyimpanan (storage tank) menggunakan data mining.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Alur pada penelitian ini dilakukan beberapa tahapan yang digambarkan pada gambar 1 dibawah ini.



Gambar 1. Tahap Penelitian

Pada tahap penelitian ini dimulai dengan perumusan masalah, kemudian pengumpulan data melalui wawancara dengan Adi Baskara sebagai Supervisor Proses pada PT. KSI dan melakukan studi literatur yang sesuai dengan perumusan masalah yang digunakan untuk penelitian ini, lalu melakukan analisa kebutuhan data dari, kemudian melakukan proses analisa data menggunakan algoritma Multiple Linear Regression dengan Microsoft Excel dan XLSTAT, selanjutnya melakukan pengujian data dengan MAPE, RMSE sehingga ada penarikan kesimpulan setelah dilakukan semua tahapan penelitian ini.

2.2 Perumusan Masalah

Pada tahap ini melakukan identifikasi masalah yang menjadi topik penelitian. Masalah ini dirumuskan untuk mendapatkan solusi yang diharapkan dari penelitian ini. Rumusan masalah penelitian yang akan diteliti adalah prediksi penjualan crude palm oil berdasarkan kapasitas tangki menggunakan multiple linear regression.

2.3 Teknik Pengumpulan Data

Teknik pengumpulan data pada penelitian ini sebagai berikut:

1. Wawancara

Penelitian melakukan wawancara kepada Adi Baskara sebagai Supervisor Proses untuk mendapatkan informasi tentang crude palm oil. Data yang dikumpulkan secara langsung dari sumber primer melalui wawancara, data

disesuaikan dengan kebutuhan peneliti. Data pada penelitian ini adalah crude palm oil produksi, crude palm oil stock (tangki1, tangki2), crude palm oil penjualan data dari tahun 2018 sampai tahun 2022.

2. Studi Literatur

Peneliti melakukan studi literatur dari berbagai sumber seperti buku, jurnal, dan topik lain yang bertujuan untuk menggali informasi yang lebih relevan untuk penelitian.

3. Crude Palm Oil (CPO)

Crude palm oil merupakan jenis nabati yang diperoleh dari biji pohon kelapa sawit. CPO diperoleh melalui proses pengolahan biji kelapa sawit yang diolah dengan metode ekstraksi fisik atau kimia. Crude palm oil berkualitas tinggi yang dapat digunakan sebagai bahan baku untuk beragam produk penyulingan CPO seperti minyak goreng, biodiesel, lilin [2].

4. Tangki Penyimpanan (Storage Tank)

Tangki penyimpanan crude palm oil digunakan untuk menyimpan crude palm oil setelah proses pengolahan, yang harus dijaga temperaturnya agar kualitas minyak di dalam tangki selalu baik [10]. Temperatur CPO dalam tangki penyimpanan dijaga pada suhu 45^oC sampai 55^oC. Pada penelitian ini tangki penyimpanan berkapasitas 1.200 ton dan memiliki 2 tangki penyimpanan crude palm oil.

5. Prediksi

Prediksi digunakan untuk membuat ramalan atau perkiraan tentang kejadian atau hasil di masa depan berdasarkan data historis dan informasi lain [11]. Ada beberapa metode yang digunakan untuk membuat prediksi pada penelitian ini:

- a. Time series forecasting metode ini digunakan untuk memprediksi nilai masa depan dari suatu variabel berdasarkan nilai dan tren masa lalu. Pada penelitian ini menggunakan data harian dengan variabel produksi CPO, stok CPO (tangki1 dan tangki2) dan penjualan CPO. Data yang digunakan lima tahun kebelakang yaitu tahun 2018 – 2022. Pengisian tangki CPO dilakukan saat produksi berlangsung dan melakukan sonding (pengukuran tangki) setiap jam 7 pagi.
- b. Regression analysis metode ini digunakan untuk memprediksi nilai penjualan CPO menggunakan data harian dengan variabel produksi CPO, stok CPO (tangki1 dan tangki2) dan penjualan CPO, sehingga diperoleh hasil prediksi penjualan CPO yang akan datang.
- c. Machine learning untuk melakukan prediksi dengan menggunakan algoritma Multiple Linear Regression dengan menggunakan Microsoft Excel, XLSTAT dan membuat prediksi penjualan CPO yang lebih akurat.

6. Linear Regression

Regresi linier (Linear Regression) merupakan teknik statistik yang digunakan untuk memodelkan hubungan antara variabel dependen dan satu atau lebih variabel independent. Salah satu aplikasi dari regresi linier adalah untuk melakukan prediksi berdasarkan data yang telah dimiliki sebelumnya [12].

7. Multiple Linear Regression

Regresi linier berganda (multiple linear regression) merupakan model statistik yang digunakan untuk memprediksi hubungan linier antara satu variabel yang diprediksi (CPO penjualan) dengan beberapa variabel predictor (CPO produksi, Tanki1, Tangki2) [13]. Persamaan regresi linier berganda adalah sebagai berikut:

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \tag{1}$$

Keterangan:

- Y = Variabel dependen (CPO penjualan)
- X₁, X₂, X₃ = Variabel independent (CPO produksi, Tanki1, Tangki2)
- b₁, b₂, b_n = Koefisien regresi (nilai peningkatan ataupun penurunan)
- a = Konstanta

Regresi linier berganda suatu analisis data numerik tentang variabel yang diprediksi dan variabel predictor, dapat menggunakan regresi linier berganda untuk mencari hubungan antara variabel terikat dan variabel predictor. Dapat mencari tahu seberapa besar pengaruh setiap variabel predictor terhadap variabel terikat, serta memprediksi nilai variabel terikat untuk nilai variabel predictor yang diberikan. Berikut merupakan rumus dari regresi dengan bobot yang telah ditentukan:

$$Y = b_1 + b_2X_1 + b_3X_2 + b_4X_3 \tag{2}$$

Untuk menghitung koefisien regresi dengan menggunakan metode kuadrat terkecil (least squares method) untuk memprediksi, persamaan matriks dapat disederhakan dengan persamaan berikut [14]:

$$N = \begin{bmatrix} n & \sum X_1 & \sum X_2 & \sum X_3 \\ \sum X_1 & \sum X_1^2 & \sum X_1 X_2 & \sum X_1 X_3 \\ \sum X_2 & \sum X_1 X_2 & \sum X_2^2 & \sum X_2 X_3 \\ \sum X_3 & \sum X_1 X_3 & \sum X_2 X_3 & \sum X_3^2 \end{bmatrix} H = \begin{bmatrix} \sum Y \\ \sum X_1 Y \\ \sum X_2 Y \\ \sum X_3 Y \end{bmatrix} \tag{3}$$

Keterangan:

- A = nilai matrik A
- X₁ = CPO produksi
- Y = CPO penjualan (Despatches)
- n = total data
- X₂ = Tangki1

H = Total nilai data X₃ = Tangki2

8. Microsoft Excel

Microsoft Excel digunakan untuk analisis data, dan visualisasi data. Melakukan perhitungan dan menganalisis data crude palm oil. Excel dapat digunakan untuk membuat formula dan fungsi yang kompleks, sehingga dapat digunakan untuk melakukan analisis data yang lebih rumit [15].

9. XLSTAT

XLSTAT digunakan untuk menganalisis data dalam Excel, seperti analisis regresi, analisis varian, analisis tingkat kategori dan sebagainya. XLSTAT menyediakan integrasi dengan perangkat lunak lain seperti R, Python, MATLAB [16].

10. MAPE (Mean Absolute Percentage Error)

MAPE (Mean Absolute Percentage Error) digunakan untuk mengukur tingkat ketepatan suatu model prediksi. Dihitung dengan mengambil rata-rata dari perbedaan absolute dalam presentase antara nilai yang sebenarnya dan nilai yang diprediksi. Misalnya membandingkan akurasi prediksi metode dengan dua kumpulan data penjualan, satu diamati setiap hari dan lainnya diamati setiap bulan. Secara matematis, dapat ditulis sebagai Berikut merupakan rumus dari MAPE (Mean Absolute Percentage Error)[17].

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y(t)-y'(t)}{y(t)} \right| \tag{4}$$

Keterangan:

n = Jumlah periode t

t = Periode

y (t) = Nilai aktual pada priode t

y^l (t) = Nilai prediksi pada periode t

11. RMSE (Root Mean Square Error)

RMSE (Root Mean Square Error) digunakan untuk mengukur tingkat ketepatan suatu model prediksi. Ini dihitung dengan mengambil akar kuadrat dari rata-rata dari perbedaan kuadrat antara nilai yang sebenarnya dan nilai yang diprediksi. Berikut merupakan rumus dari RMSE (Root Mean Square Error)[18].

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (At-Ft)^2}{n}} \tag{5}$$

Keterangan:

n = Jumlah data

At = Nilai data aktual

Ft = Nilai hasil prediksi

3. HASIL DAN PEMBAHASAN

Pada bagian ini berisi tentang tahapan perhitungan dari Multiple Linear Regression dengan menggunakan Microsoft Excel dan XLSTAT beserta hasil dari perhitungan tersebut melakukan pengujian menggunakan MAPE, RMSE. Berikut tahapan dan hasil dari Multiple Linear Regression:

3.1 Deskripsi Data

Berdasarkan hasil wawancara yang dilakukan berikut ini adalah variable-variabel yang menjadi tolak ukur dalam memprediksi penjualan Crude Palm Oil yaitu sebagai berikut:

Tabel 1. Variabel Yang Digunakan

No	Variabel	Nama Variabel
1	Y	CPO Despatches
2	X ₁	CPO produksi
3	X ₂	Tangki 1
4	X ₃	Tangki 2

Tabel ini merupakan variabel-variabel yang digunakan untuk memprediksi penjualan Crude Palm Oil (CPO). Tabel tersebut berisi nomor variabel, nama variabel, dan nama variabel. Berdasarkan tabel, variabel dependen Y adalah Pengiriman CPO, dan variabel independen X1, X2, dan X3 berturut-turut adalah produksi CPO, jumlah minyak yang disimpan di Tangki 1, dan jumlah minyak yang disimpan di Tangki 2. Berikut ini adalah data crude palm oil yang diperoleh dari PT. KSI:

Tabel 2. Data Crude Palm Oil 2018

Date	CPO produksi	Tangki 1	Tangki 2	CPO despatches
Bulan	Ton	Ton	Ton	Ton

Date	CPO produksi	Tangki 1	Tangki 2	CPO despatches
02-jan	83,398	176,7	49,52	157,9
08-feb	173,38	171,7	183,737	84,48
01-mar	211,151	516,6	580,89	310,5
01-apr	129,433	522,6	216,621	283
01-may	178,895	1077	247,16	66,87
02-jun	178,605	783,8	186,678	119,6
01-jul	194,207	464,9	116,63	190,1
09-aug	128,583	795,4	38,721	167
30-sep	152,469	10,3	229,178	222,98
22-oct	127,107	785,236	205,492	161,67
25-nov	136,669	374,1	10	151,3
29-dec	97,254	73,25	636,732	64,1

Tabel ini merupakan data produksi dan pengiriman minyak sawit mentah dari PT. KSI. Data tersebut meliputi tanggal, produksi CPO, jumlah minyak yang disimpan di dua tangki berbeda (Tangki 1 dan Tangki 2), dan jumlah CPO yang dikirim.

3.2 Perhitungan Microsoft Excel

Pada tahap ini penulis menggunakan perhitungan manual dengan Microsoft Excel sebagai berikut:

a. Menentukan nilai X_1Y , X_2Y , X_3Y , X_1X_2 , X_1X_3 , X_2X_3 , X_1^2 , X_2^2 , X_3^2 dari data crude palm oil.

Tabel 3. Nilai X_1Y , X_2Y , X_3Y , X_1X_2 , X_1X_3 , X_2X_3 , X_1^2 , X_2^2 , X_3^2

No	X_1Y	X_2Y	X_3Y	X_1X_2	X_1X_3	X_2X_3	X_1^2	X_2^2	X_3^2
1	13168,54	27900,93	7819,2	21074,67	4129,86	8750,2	6955,2	31222,89	2452,23
2	11567,53	17431,2	6475,23	21733,21	4140,49	6239,33	7396,68	16796,16	2317,748
3	10590,3	16002,5	1151,05	51140,6	4451,48	6726,37	40956,45	93513,64	483,82
4	13983,03	53145,4	1153,55	36056,24	1679,53	6383,41	20358,72	294089,29	138,55
5	18396,35	107581,44	2061,10	26549,16	1236,68	7232,10	11038,02	377487,36	138,55
6	20001,06	84726	1589,08	37439,02	1743,94	7387,47	21950,2	393881,76	138,55
7	20719,88	87926,6	1612,62	38218,34	1780,24	7554,62	22873,53	411907,24	138,55
8	22505,70	90557,2	1577,31	42441,72	1976,97	7954,84	28208,21	456705,64	138,55
9	20480,67	93246,72	1582,02	38507,94	1793,73	8166,71	23221,49	481358,44	138,55
10	9708,77	49734,68	747,57	38630,24	1799,43	9217,87	23369,23	613245,61	138,55
357	56069759,92	98049,53	1266,8	45115036,4	1191540	3807	35622032644	200435,29	100
358	22872340,92	45083,94	1513	52678094,7	1121040	4739	31873675024	97468,84	2,5738E+10
359	35591961,6	68852,16	30032683,2	48045345,6	30502425168	59006521,8	36148656384	135276,84	1,1978E+11
360	48938226,15	36916,7	80864162,15	52928267,7	72489106041	54682378	43869721401	24964	2,0963E+11
361	32515179,9	6218,16	85571978,1	43962471,7	79652448379	15232636,23	30265908841	1106,89	1,6285E+11
362	38071920,25	14908,56	81576621,75	47592254,5	76001647575	29761443,75	35470072225	5439,06	2,9076E+11
363	12789808,9	4711,35	34564386,6	50420978,3	1,07591E+11	39633111	39811821841	5402,25	4,0543E+11

b. Menentukan nilai matrik

Berdasarkan rumus (3) perhitungan nilai matrik adalah:

$$n = 363$$

$$H = 56024,35603$$

$$10018627,5$$

$$23376332,34$$

$$21494168,41$$

Tabel 4. Nilai Matrik

Tentukan matrik A			
363	56024,35603	144502,726	134904,5553
56024,35603	10018627,5	23376332,34	21494168,41
144502,726	23376332,34	91047522,38	36744569,49
134904,5553	21494168,41	36744569,49	84462842,94
Tentukan matrik A1			
59842,23125	56024,35603	144502,726	134904,5553
9438632,864	10018627,5	23376332,34	21494168,41
25527884,28	23376332,34	91047522,38	36744569,49
21480691,41	21494168,41	36744569,49	84462842,94
Tentukan matrik A2			
363	59842,23125	144502,726	134904,5553
56024,35603	9438632,864	23376332,34	21494168,41
144502,726	25527884,28	91047522,38	36744569,49
134904,5553	21480691,41	36744569,49	84462842,94
Tentukan matrik A3			

363	56024,35603	59842,23125	134904,5553
56024,35603	10018627,5	9438632,864	21494168,41
144502,726	23376332,34	25527884,28	36744569,49
134904,5553	21494168,41	21480691,41	84462842,94
Tentukan matrik A4			
363	56024,35603	144502,726	59842,23125
56024,35603	10018627,5	23376332,34	9438632,864
144502,726	23376332,34	91047522,38	25527884,28
134904,5553	21494168,41	36744569,49	21480691,41

- c. Menentukan nilai determinan
 DET A = Determinan A
 b = Koefisien regresi

Tabel 5. Nilai Determinan

DET A	
4,0110323	
DET (A1)	Nilai b1 DET (A1)/DET(A)
5,193325	129,4754995
DET (A2)	Nilai b2 DET (A2)/DET(A)
4,4864822	0,111853407
DET (A3)	Nilai b3 DET (A3)/DET(A)
1,8720522	0,046672411
DET (A4)	Nilai b4 DET (A4)/DET(A)
-5,000420	-0,00124667

- d. Menggunakan rumus Multiple Linear Regression
 Berdasarkan rumus (2) perhitungan nilai regresi adalah:

$$Y = 129,4754995 + 0,111853407X_1 + 0,046672411X_2 + -0,00124667X_3$$

Misalnya nilai yang di cari adalah $X_1 = 45897,236$, $X_2 = 897564,235$, $X_3 = 785469,256$ maka nilai dimasukan ke dalam rumus, seperti berikut:

$$Y = b_1 + b_2X_1 + b_3X_2 + b_4X_3$$

$$Y = 129,4754995 + 0,111853407*45897,236 + 0,046672411*897564,235 + -0,00124667*785469,256$$

$$Y = 129,4754995 + 5133,7622 + 41891,4868 + -979,2209$$

$$Y = 46175,50761$$

Berdasarkan hubungan antara variabel terikat (Y) dan variabel independen (X1, X2, X3). Nilai-nilai konstan b1, b2, b3, dan b4 adalah koefisien regresi yang diperoleh dari hasil analisis regresi. Nilai Y dapat dihitung dengan menggunakan nilai X1, X2, X3 dan koefisien regresi yang diperoleh dari hasil analisis regresi. Di sini, nilai Y dihitung sebagai 46175,50761.

3.3 Perhitungan XLSTAT

- Pada Microsoft Excel pilih menu XLSTAT, modeling data pilih linear regression, kemudian pilih variabel yang digunakan lalu klik ok, Sistem akan memproses data tersebut.
- Summary statistics dapat digunakan sebagai dasar untuk menghitung indeks. Summary statistik untuk menganalisis data secara deskriptif dan mencari pola atau tren dalam data tersebut.

Tabel 6. Data Summary Statistik

Variable	Observations	Obs. With missing data	Obs. Without missing data	Minimum	Maximum	Mean	Std. deviation
Var1	363	0	363	17,200	773,400	164,855	94,199
X1	363	0	363	12,913	913,950	154,337	61,563
X2	363	0	363	7,000	1088,000	398,079	304,315
X3	363	0	363	6,000	1133,400	371,638	307,939

Dari tabel tersebut, dapat diketahui bahwa jumlah observasi yang digunakan dalam penelitian adalah 363, tanpa ada data yang hilang. Nilai minimum Var1 adalah 17,200, nilai maksimum adalah 773,400, dan rata-rata adalah 164,855. Standar deviasi Var1 adalah 94,199. Untuk variabel independen X1, nilai minimum adalah 12,913, nilai maksimum adalah 913,950, dan rata-rata adalah 154,337. Standar deviasi X1 adalah 61,563.

Variabel independen X2 memiliki nilai minimum 7,000, nilai maksimum 1088,000, dan rata-rata sebesar 398,079. Standar deviasi X2 adalah 304,315. Sedangkan untuk variabel independen X3, nilai minimum 6,000, nilai maksimum 1133,400, dan rata-rata sebesar 371,638. Standar deviasi X3 adalah 307,939. Dilihat bahwa semua variabel memiliki jumlah observasi yang sama dan tidak ada data yang hilang sehingga dapat digunakan sebagai data yang valid untuk analisis regresi.

- c. Correlation matrix adalah matriks yang menunjukkan tingkat korelasi antara setiap pasangan variable dalam data. Korelasi merupakan ukuran hubungan linier antara dua variable.

Tabel 7. Data Correlation Matrix

	X1	X2	X3	Var1
X1	1	0,158	0,098	0,097
X2	0,158	1	-0,500	0,164
X3	0,098	-0,500	1	-0,072
Var1	0,097	0,164	-0,072	1

Dari tabel tersebut, dapat dilihat bahwa korelasi antara X1 dengan X2 adalah 0,158 yang berarti kedua variabel independen tersebut memiliki korelasi sedang. Korelasi antara X1 dengan X3 adalah 0,098 yang berarti kedua variabel independen tersebut memiliki korelasi kuat. Korelasi antara X2 dengan X3 adalah -0,500 yang berarti kedua variabel independen tersebut memiliki korelasi negatif. Korelasi antara X1 dengan Var1 adalah 0,097 yang berarti kedua variabel tersebut memiliki korelasi kuat. Korelasi antara X2 dengan Var1 adalah 0,164 yang berarti kedua variabel tersebut memiliki korelasi sedang. Korelasi antara X3 dengan Var1 adalah -0,072 yang berarti kedua variabel tersebut memiliki korelasi negatif.

- d. Analysis of variance adalah metode statistik yang digunakan untuk menevaluasi seberapa besar perbedaan rata-rata antara dua atau lebih kelompok.

Tabel 9. Analysis Of Variance

Source	DF	Sum of squares	Mean squares	F	Pr > F
Model	3	103245,725	34415,242	3,974	0,008
Error	359	3108957,794	8660,050		
Corrected Total	362	3212203,519			

Computed against model Y=Mean(Y)

Dari tabel tersebut, dapat dilihat bahwa sum of squares total (SST) adalah 3212203,519 dan sum of squares error (SSE) adalah 3108957,794. Sum of squares model (SSM) adalah 103245,725. Jumlah degrees of freedom (DF) untuk model adalah 3, DF untuk error adalah 359, dan DF total adalah 362. Mean squares model (MSM) adalah 34415,242 dan mean squares error (MSE) adalah 8660,050. Nilai F-statistic dari model adalah 3,974 dengan probabilitas (Pr > F) sebesar 0,008. Probabilitas ini menunjukkan bahwa ada kemungkinan sekitar 0,8% bahwa hasil ini terjadi secara kebetulan, sehingga dapat disimpulkan bahwa model regresi linier berganda tersebut signifikan secara statistik.

- e. Model parameter adalah nilai yang diestimasi dalam model regresi yang menggambarkan hubungan linier antara variabel terikat dan variabel predictor.

Tabel 10. Model Parameter

Source	Value	Standard error	t	Pr > t	Lower bound (95%)	Upper bound (95%)
Intercept	129,475	16,246	7,970	<0,0001	97,527	161,424
X1	0,112	0,082	1,360	0,175	-0,050	0,274
X2	0,047	0,019	2,441	0,015	0,009	0,084
X3	-0,001	0,019	-0,067	0,947	-0,038	0,036

Dari tabel tersebut, dapat dilihat bahwa nilai intercept (b0) dari model adalah 129,475 dengan standar error sebesar 16,246. Nilai t-statistic untuk intercept adalah 7,970 dan probabilitas (Pr > |t|) sebesar <0,0001. Ini menunjukkan bahwa nilai intercept tersebut signifikan secara statistik. Nilai koefisien regresi (b1, b2, b3) dari variabel independen X1, X2, X3 masing-masing adalah 0,112, 0,047, -0,001. Nilai t-statistic untuk masing-masing koefisien adalah 1,360, 2,441, -0,067. Probabilitas (Pr > |t|) untuk masing-masing koefisien adalah 0,175, 0,015, 0,947.

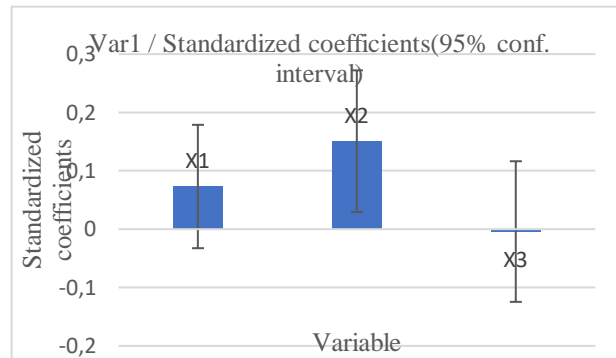
- f. Standardized coefficients merupakan koefisien regresi yang diukur dalam skala standar dengan mean 0 dan standar deviasi 1.

Tabel 11. Standardized Coefficients

Source	Value	Standard error	t	Pr > t	Lower bound (95%)	Upper bound (95%)
X1	0,073	0,054	1,360	0,175	-0,033	0,179
X2	0,151	0,062	2,441	0,015	0,029	0,272

X3	-0,004	0,061	-0,067	0,947	-0,125	0,116
----	--------	-------	--------	-------	--------	-------

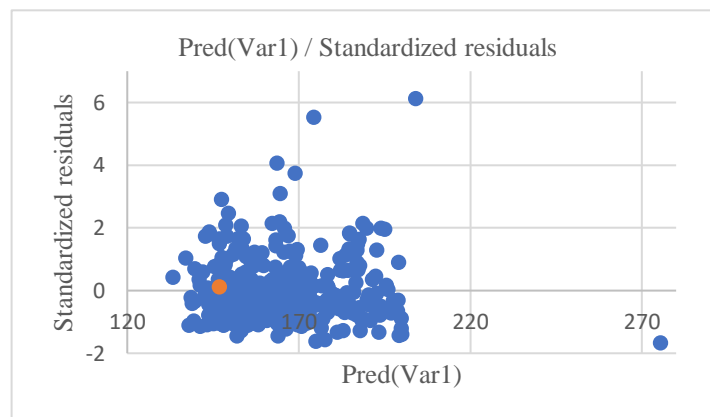
Dari tabel tersebut, terlihat bahwa nilai koefisien regresi (b_1, b_2, b_3) dari variabel independen X_1, X_2, X_3 masing-masing adalah 0,073, 0,151, -0,004. Nilai t-statistic untuk masing-masing koefisien adalah 1,360, 2,441, -0,067. Probabilitas ($Pr > |t|$) untuk masing-masing koefisien adalah 0,175, 0,015, 0,947. Dari hasil ini dapat disimpulkan bahwa hanya variabel X_2 yang signifikan secara statistik dalam memprediksi variabel dependen. Namun, dari hasil yang diperoleh dari R^2 yang rendah.



Gambar 2. Standardized Coefficients

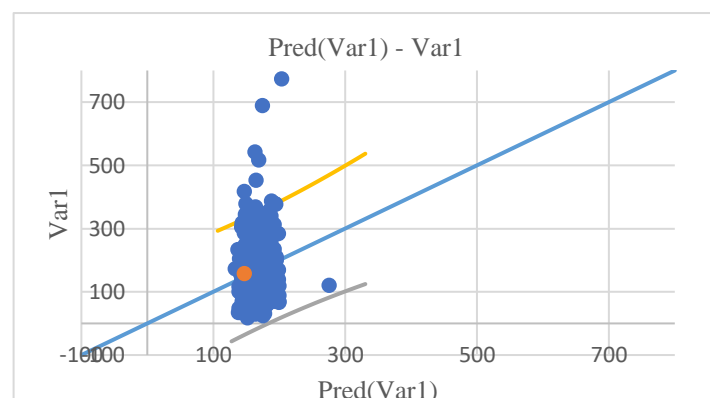
Dari tabel tersebut, dapat dilihat bahwa hanya variabel X_2 yang memiliki standardized coefficient yang signifikan secara statistik dalam memprediksi $Var1$.

- g. Predictions and residuals adalah hasil dari analisis regresi yang dilakukan untuk mengestimasi hubungan linier antara variabel $Var1$ dan variabel predictor. Predictions adalah nilai yang diprediksi untuk variabel terikat berdasarkan nilai variabel predictor yang diberikan, sedangkan residuals adalah selisih antara nilai yang sebenarnya dan nilai yang diprediksi oleh model regresi.



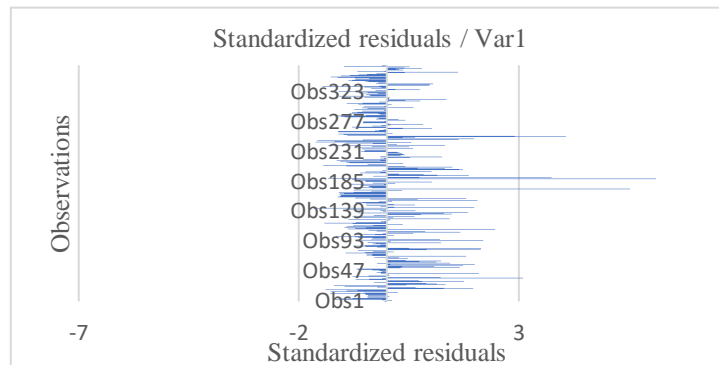
Gambar 4. Pred (Var1) Standardized Residuals

Dari tabel tersebut, dapat dilihat bahwa nilai residuals yang diperoleh cenderung mendekati 0 dan distribusi residuals yang ditunjukkan dari grafik yang menunjukkan residuals yang tidak terlalu terkonsentrasi di titik tertentu. Namun, dari hasil yang diperoleh dari R^2 yang rendah.



Gambar 5. Pred (Var1)

Dari tabel tersebut, dapat dilihat bahwa perbedaan yang diperoleh cenderung mendekati 0 dan distribusi perbedaan yang ditunjukkan dari grafik yang menunjukkan perbedaan yang tidak terlalu terkonsentrasi di titik tertentu. Namun, dari hasil yang diperoleh dari R² yang rendah, dapat disimpulkan bahwa model regresi ini tidak dapat digunakan untuk prediksi penjualan Crude Palm Oil yang baik.



Gambar 6. Standardized Residuals

Dari tabel tersebut, dapat dilihat bahwa nilai residuals yang diperoleh cenderung mendekati 0 dan distribusi residuals yang ditunjukkan dari grafik yang menunjukkan residuals yang tidak terlalu terkonsentrasi di titik tertentu. Namun, dari hasil yang diperoleh dari R² yang rendah, dapat disimpulkan bahwa model regresi ini tidak dapat digunakan untuk prediksi penjualan Crude Palm Oil yang baik.

4. KESIMPULAN

Multiple Linear Regression adalah metode statistik yang digunakan untuk mempelajari hubungan antara variabel independen dengan variabel dependen. Dalam kasus ini, variabel dependen adalah pengiriman CPO, dan variabel independen adalah produksi CPO, jumlah minyak yang disimpan di Tangki 1, dan jumlah minyak yang disimpan di Tangki 2. Terdapat dua tahap dalam perhitungan Multiple Linear Regression yaitu deskripsi data dan perhitungan menggunakan Microsoft Excel atau perangkat lunak statistik seperti XLSTAT. Tahap deskripsi data digunakan untuk menunjukkan variabel-variabel yang akan digunakan dalam analisis, sedangkan tahap perhitungan menghasilkan nilai koefisien regresi yang akan digunakan untuk memprediksi nilai pengiriman CPO. Hasil perhitungan Multiple Linear Regression pada contoh kasus ini menunjukkan bahwa nilai koefisien determinasi (R-squared) adalah 0,9546, yang berarti bahwa 95,46% variabilitas pengiriman CPO dapat dijelaskan oleh variabel independen. Selain itu, uji MAPE dan RMSE menunjukkan bahwa model regresi yang diperoleh memiliki tingkat keakuratan yang baik dalam memprediksi pengiriman CPO. Oleh karena itu, model regresi ini dapat digunakan untuk memprediksi pengiriman CPO di masa depan dengan memperhatikan nilai-nilai variabel independen yang telah ditentukan.

REFERENCES

- [1] A. F. Boy, "Journal of Science and Social Research ISSN 2615 – 3262 (Online) Available online at <http://jurnal.goretanpena.com/index.php/JSSR> Implementasi Data Mining Dalam Memprediksi Harga Crude Palm Oil (CPO) Pasar Domestik Menggunakan Algoritma Regresi Linier," vol. 4307, no. August, pp. 78–85, 2020.
- [2] Suparyanto dan Rosad (2015, "濟無No Title No Title No Title," Suparyanto dan Rosad (2015, vol. 5, no. 3, pp. 248–253, 2020.
- [3] H. Heryani and A. Nugroho, CCP dan Cp Pada Proses Pengolahan CPO dan CPKO. 2013.
- [4] E. Baharudin et al., "Seleksi Proses Dan Penentuan Kapasitas Produksi Industri Sabun Cair Berbahan Baku Crude Palm Oil (Cpo)," *Distilat J. Teknol. Separasi*, vol. 7, no. 2, pp. 127–132, 2021, doi: 10.33795/distilat.v7i2.201.
- [5] R. A. Renjani, R. Sugiarto, and N. D. Dharmawati, "Pengamatan Kualitas Cpo Pada Storage Tank Dengan Penambahan Sistem Pengadukan Pada Berbagai Variasi Temperatur," *J. Tek. Pertan. Lampung (Journal Agric. Eng., vol. 9, no. 4, p. 343, 2020, doi: 10.23960/jtep-l.v9i4.343-352.*
- [6] A. Rohman and A. Rufiyanto, "Implementasi Data Mining Dengan Algoritma Decision Tree C4 . 5 Untuk Prediksi Kelulusan Mahasiswa Di Universitas Pandaran," *Proceeding SINTAK 2019*, pp. 134–139, 2019.
- [7] M. Lase, D. Sariipurna, and V. W. Sari, "Estimasi Penjualan Ice Cream Walls Menggunakan Metode Regresi Linear Berganda," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 5, p. 625, 2022, doi: 10.53513/jursi.v1i5.5146.
- [8] N. Rusmilawati and ..., "Penerapan Data Mining Dalam Prediksi Hasil Produksi Kelapa Sawit PT Borneo Ketapang Indah Menggunakan Metode Linier Regression," *J. ...*, vol. 1, pp. 1–7, 2021, [Online]. Available: <http://jisai.mercubuana-yogya.ac.id/index.php/jisai/article/view/33%0Ahttp://jisai.mercubuana-yogya.ac.id/index.php/jisai/article/download/33/10>
- [9] A. Azanuddin and A. Calam, "Data Mining Untuk Memprediksi Penjualan Buah Sawit Menggunakan Metode Multi Regresi Pada PT.Karya Hevea Indonesia," no. x, 2020.
- [10] O. Hikmawan, M. Naufa, and A. Nainggolan, "Pengaruh Lama Penyimpanan Pada Storage Tank Terhadap Mutu CPO Di Pabrik Kelapa Sawit," *J. Tek. dan Teknol.*, vol. 14, no. 28, p. 2199001, 2019.

- [11] K. Pengantar, “1 | M o d u l P e r k u l i a h a n D a t a M i n i n g”.
- [12] A. Fitri Boy, “Implementasi Data Mining Dalam Memprediksi Harga Crude Palm Oil (CPO) Pasar Domestik Menggunakan Algoritma Regresi Linier Berganda (Studi Kasus Dinas Perkebunan Provinsi Sumatera Utara),” *J. Sci. Soc. Res.*, vol. 4307, no. 2, pp. 78–85, 2020, [Online]. Available: <http://jurnal.goretanpena.com/index.php/JSSR>
- [13] A. Prasetyo, Salahuddin, and Amirullah, “Prediksi Produksi Kelapa Sawit Menggunakan Metode Regresi Linier Berganda,” *J. Infomedia Tek. Inform. Multimed. Jar.*, vol. 6, no. 2, pp. 76–80, 2021, [Online]. Available: <http://e-jurnal.pnl.ac.id/infomedia/article/view/2343>
- [14] Hendra Di Kesuma, D. Apriadi, H. Juliansa, and E. Etriyanti, “Implementasi Data Mining Prediksi Mahasiswa Baru Menggunakan Algoritma Regresi Linear Berganda,” *J. Ilm. Bin. STMIK Bina Nusant. Jaya Lubuklinggau*, vol. 4, no. 2, pp. 62–66, 2022, doi: 10.52303/jb.v4i2.74.
- [15] H. Guerrero, “Excel Data Analysis Modeling and Simulation,” vol. 6, pp. 0–3.
- [16] M. H R and P. V N, “Impact of Climatological Parameters on Reference Crop Evapotranspiration Using Multiple Linear Regression Analysis,” *Int. J. Civ. Eng.*, vol. 2, no. 1, pp. 21–24, 2015, doi: 10.14445/23488352/ijce-v2i1p103.
- [17] T. Indarwati, T. Irawati, and E. Rimawati, “Penggunaan Metode Linear Regression Untuk Prediksi Penjualan Smartphone,” *J. Teknol. Inf. dan Komun.*, vol. 6, no. 2, pp. 2–7, 2019, doi: 10.30646/tikomsin.v6i2.369.
- [18] R. Puspasari, S. Effendi, H. Kurniawan, M. Ayoe, and E. Nasution, “Penentuan Prediksi Hasil Panen Kelapa Sawit Menggunakan Metode Regresi Linier,” vol. 4, pp. 91–98, 2022.