

Penerapan Data Mining Dalam Prediksi Hasil Produksi Kelapa Sawit PT Borneo Ketapang Indah Menggunakan Metode *Linier Regression*

Nuning Rusmilawati¹, Putri Taqwa Prasetyaningrum²

^{1,2}Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta
Yogyakarta, Indonesia

e-mail: ¹nuningrusmila@gmail.com, ²putri@mercubuana-yogya.ac.id

Diajukan: 26 Oktober 2020; Direvisi: 08 Februari 2021; Diterima: 02 Maret 2021

Abstrak

Berdasarkan data yang ada seperti Luas Lahan, Curah Hujan, Pemupukan dan Jumlah Produksi sebagai variable yang akan diprediksi. Oleh karena itu, peneliti melakukan analisis Data Mining Hasil Produksi kelapa sawit agar menjadi informasi yang dapat di gunakan oleh PT Borneo Ketapang Indah. Dalam melakukan analisis peneliti menggunakan tools Rapidminer, metode yang digunakan dalam penelitian ini adalah Linier Regression. Setelah dilakukan analisis prediksi Linier Regression pada tools rapidminer maka di dapatlah 2 variable factor yang mempengaruhi Hasil Produksi kelapa sawit yaitu variable X1 (Luas Lahan) dan variable X2 (Curah Hujan) dengan nilai coefficient masing-masing variable sebesar 0,646 untuk variable Luas Lahan, variable Curah Hujan sebesar 0,530.

Kata Kunci : Data Mining, Linier Regression, Rapidminer

Abstract

Based on existing data such as land area, rainfall, fertilization and total production. as a variable to be predicted. Therefore, the researchers conducted an analysis of Oil Palm Production Data Mining to provide information that could be used by PT Borneo Ketapang Indah. In conducting the analysis, the researcher uses Rapidminer tools, the method used in this research is Linear Regression. After analyzing the Linear Regression prediction on the rapid mining tool, there are 2 variable factors that affect oil palm production, namely variable X1 (land area) and variable X2 (rainfall) with the coefficient value of each variable of 0.646 for the variable land area, variable Rainfall of 0.530.

Keywords: Data Mining, Linier Regression, Rapidminer

1. Pendahuluan

Tanaman Kelapa Sawit saat ini merupakan salah satu jenis tanaman perkebunan yang menduduki posisi penting disektor pertanian umumnya, dan sektor perkebunan khususnya, hal ini disebabkan karena dari sekian banyak tanaman yang menghasilkan minyak atau lemak, kelapa sawit yang menghasilkan nilai ekonomi terbesar per hektarnya di dunia. Peramalan atau forecasting merupakan seni dan ilmu yang dilakukan untuk mengetahui atau memperkirakan apa yang terjadi pada masa yang akan datang. Peramalan menjadi sangat penting karena penyusunan suatu rencana diantaranya didasarkan pada suatu proyeksi atau ramalan [1][2].

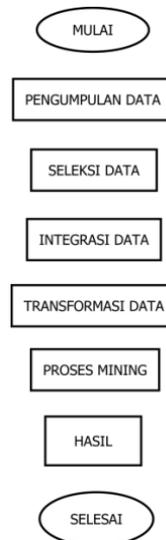
PT Borneo Ketapang Indah merupakan suatu perusahaan yang bergerak dalam bidang perkebunan produksi kelapa sawit yang terletak di pulau Kalimantan. Dalam memprediksi jumlah produksi kelapa sawit dapat membantu perusahaan untuk melihat factor apa saja yang mempengaruhi dalam proses produksi kelapa sawit pada masa yang akan datang dengan menggunakan data yang ada seperti Luas Lahan, Curah Hujan, Pemupukan dan Jumlah Produksi sebagai variable yang akan diprediksi.

Berdasarkan urai-uraian diatas, penulis tertarik untuk menulis judul “Penerapan Data Mining dalam Prediksi Hasil Produksi Kelapa Sawit PT Borneo Ketapang Indah Menggunakan Metode Linear Regression”. Dimana penerapan Data Mining diharapkan dapat mengatasi permasalahan. Penelitian pertama [3] oleh yang berjudul “Penerapan Metode Regresi Linier untuk Memprediksi Kebiasaan Pelanggan Studi Kasus: PT. Mensa Binasukses”. PT. Mensa Binasukses merupakan distributor usaha yang bergerak di bidang pelayanan kesehatan, khususnya dalam hal penjualan obat-obatan, baik itu penjualan obat melalui resep dokter maupun apotek.

Permasalahan yang muncul adalah bagaimana mengolah data transaksi pelanggan yang sudah ada menjadi sebuah prediksi kebiasaan pelanggan. Hasil prediksi ini akan digunakan sebagai strategi perusahaan untuk memberi rekomendasi pembelian produk dalam bentuk paket (untuk pelanggan) dan strategi pemasaran serta pembuatan katalog (untuk perusahaan). Proses prediksi dilakukan dengan cara menggali informasi hubungan asosiasi antar item produk yang terjadi pada setiap transaksi. Salah satu Metode yang dapat digunakan untuk menggali informasi potensial ini adalah dengan pendekatan data mining menggunakan metode regresi linier. Penelitian ini bertujuan untuk mengimplementasikan sebuah perangkat lunak untuk memprediksi kebiasaan pelanggan menggunakan metode regresi linier dengan memanfaatkan data transaksi pelanggan yang sudah ada. Penelitian kedua oleh [4] yang berjudul “Penerapan Algoritme Linear Regression untuk Prediksi Hasil Panen Tanaman Padi”. Hasil panen padi sangat berpengaruh dalam pemenuhan kebutuhan pangan pokok beras. Dikarenakan kebutuhan beras tiap tahunnya selalu mengalami peningkatan, diperlukan prediksi hasil panen untuk memperkirakan penanaman yang akan dilakukan, sehingga dapat mencukupi pemenuhan kebutuhan beras. Untuk memperoleh hasil panen, diperlukan informasi mengenai luas lahan, jumlah bibit yang akan di tanam, dan pupuk yang digunakan sebagai penunjang kebutuhan tanaman padi. Variabel yang paling berpengaruh terhadap panen adalah faktor cuaca atau curah hujan, namun untuk faktor harga pada suatu waktu atau pada kondisi dimana faktor produksi yang besar tidak menjamin harga akan lebih murah atau stabil [5]. Model regresi yang digunakan memberikan hasil 57,78% yang berarti curah hujan dipengaruhi oleh 57,78% variable bebas yang diukur pada penelitian ini [6]. Algoritma Regresi Linear Berganda mempunyai Kelebihan diantaranya melakukan generalisasi dan ekstraksi dari pola data tertentu, mampu mengakuisisi pengetahuan walau tidak ada kepastian, dan mampu melakukan perhitungan secara paralel sehingga proses lebih singkat [7]. Oleh karena itu, dalam penelitian ini membahas tentang pengelompokan jumlah data bencana dan prediksi data bencana yang akan terjadi 5 tahun kedepan menggunakan teknik data mining. Algoritmadata mining yang digunakan adalah K-Means untuk clustering dan Linear Regression untuk prediksi data bencana [8]. Teknik data mining dan machine learning dapat digunakan untuk memprediksi berdasarkan data-data masa lalu[9]. Ilmu Pengetahuan dari proses data mining adalah pihak pengambil keputusan dapat mengetahui kebutuhan lahan yang harus disediakan [10].

2. Metode Penelitian

Pada penelitian Penerapan Data Mining Dalam Prediksi Jumlah Hasil Produksi Kelapa Sawit PT Borneo Ketapang Indah menggunakan metode Linear Regression akan melalui beberapa proses, berikut diagram alur proses jalan penelitian pada gambar.



Gambar 1. Flowchart alur penelitian

3. Hasil dan Pembahasan

Data yang digunakan dalam penelitian ini yaitu data jumlah hasil produksi kelapa sawit berdasarkan data historis 3 tahun terakhir yaitu 2017, 2018 dan 2019 data tersebut berasal dari PT Borneo Ketapang Indah. Berdasarkan data yang diperoleh langsung dari PT Borneo ketapang Indah yang dimana data tersebut

telah melalui proses selection yang terdiri dari beberapa atribut yaitu Bulan, Luas Lahan, Curah Hujan, Pupuk dan Jumlah produksi yang dapat dilihat pada table dibawah.

Tabel 1. Atribut

Bulan	Luas Lahan	Curah Hujan	Pupuk (kg)	Jumlah Produksi
Jan-17	6,28	31140	7500	338535
Feb-17	11,01	22170	7500	395119
Mar-17	10,13	41070	7550	450037
Apr-17	17,71	22470	8250	480269
Mei-17	18,89	24090	8250	53638
Juni-17	15,35	18730	7810	507942

PERHITUNGAN MANUAL

Adapun rumus Linear Regression adalah sebagai berikut :

$$Y' = a + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

Keterangan :

- a. Y' = Variable Dependen (Nilai yang diprediksikan)
- b. X₁, X₂, X_n = Variable Independen
- c. Konstanta (nilai Y' apabila X₁, X₂, X_n=0)
- d. b₁, b₂, b_n = Koefisien Regresi

Rumus Linear Regression kedua, untuk mencari a dan b dapat di peroleh dengan menggunakan rumus dibawah ini :

- a. $\sum Y_i = na + \sum b_1X_1 + \sum b_2X_2 + \sum b_3X_3$
- b. $\sum X_1Y_i = a\sum X_1 + b_1\sum (X_1)^2 + b_2\sum X_1X_2 + b_3\sum X_1X_3$
- c. $\sum X_2Y_i = a\sum X_2 + b_1\sum X_1X_2 + b_2\sum (X_2)^2 + b_3\sum X_2X_3$
- d. $\sum X_3Y_i = a\sum X_3 + b_1\sum X_1X_3 + b_2\sum X_2X_3 + b_3\sum (X_3)^2$

Tahapan perhitungan manual Liniear Regression.

- a. Siapkan data *prediksi jumlah produksi
- b. Hitung berapa jumlah data set (n). *n=19
- c. Tentukan mana yang X dan mana Y. Y disini adalah nilai yang akan diprediksi. *X₁=Luas Tanah, X₂=

Curah Hujan, X₃=Pupuk, Y= Jumlah Produksi

- d. Hitung (X₁)², (X₂)², (X₃)³
- e. Hitung X₁*X₂, X₁*X₃, X₂*X₃, X₁*Y, X₂*Y, X₃*Y
- f. Hitung $\sum X_1$, $\sum X_2$, $\sum X_3$, Y
- g. Hitung $\sum (X_1)^2$, $\sum (X_2)^2$, $\sum (X_3)^2$
- h. Hitung $\sum X_1X_2$, $\sum X_1X_3$, $\sum X_2X_3$, $\sum X_1Y$, $\sum X_2Y$, $\sum X_3Y$
- i. Masukkan perhitungan kedalam rumus persamaan mencari a, b₁, b₂, b₃
- j. Hitung nilai a, b₁, b₂, b₃ menggunakan perhitungan matrix
- k. Buat persamaan regresi linier

Perhitungan Manual Liniear Regression dengan membagi data training dan testing dapat dilihat pada table dibawah ini yang dimana atribut Luas Lahan menjadi variable X₁, atribut Curah Hujan menjadi variable X₂, atribut Pupuk menjadi variable X₃ dan atribut Jumlah Produksi menjadi variable Y yang nantinya akan diprediksi.

	X1	X2	X3	Y
Bulan	Luas Lahan (Ha)	Curah Hujan (mm)	Pupuk (kg)	Jumlah Produksi
Jan-17	6,28	31140	7500	338535
Feb-17	11,01	22170	7500	395119
Mar-17	10,13	41070	7550	450037
Apr-17	17,71	22470	8250	480269
May-17	18,89	24090	8250	53638
June-17	15,35	18730	7810	507942
July-17	11,86	13310	7900	494224
Aug-17	17,40	20960	7950	581595
Sept-17	19,73	8480	7500	487000
Oct-17	15,51	12830	8000	484627
Nov-17	13,13	29440	8000	472433
Dec-17	9,77	33780	7870	425819
Jan-18	13,68	31140	7870	354726
Feb-18	19,23	399	7870	296990
Mar-18	10,54	30250	7500	402782
Apr-18	20,62	16820	7500	410989
Mei-18	20,39	12390	7870	410989
June-18	15,24	7260	8250	380655
July-18	9,48	11160	8000	268894

Gambar 2. Data Training

Bulan	Luas Lahan (Ha)	Curah Hujan (mm)	Pupuk (kg)	Jumlah Produksi
Jan-19	17,96	175	8000	322242
Feb-19	10,92	149	7500	259420
Mar-19	14,78	172	7550	412263
Apr-19	14,78	50	8000	?
May-19	10,92	100	7500	?
June-19	17,96	100	8000	?

Gambar 3. Data Testing

Hitung $\sum X_1, \sum X_2, \sum X_3, \sum Y$

	X1	X2	X3	Y
Bulan	Luas Lahan (Ha)	Curah Hujan (mm)	Pupuk (kg)	Jumlah Produksi
Jan-17	6,28	31140	7500	338535
Feb-17	11,01	22170	7500	395119
Mar-17	10,13	41070	7550	450037
Apr-17	17,71	22470	8250	480269
May-17	18,89	24090	8250	53638
June-17	15,35	18730	7810	507942
July-17	11,86	13310	7900	494224
Aug-17	17,40	20960	7950	581595
Sept-17	19,73	8480	7500	487000
Oct-17	15,51	12830	8000	484627
Nov-17	13,13	29440	8000	472433
Dec-17	9,77	33780	7870	425819
Jan-18	13,68	31140	7870	354726
Feb-18	19,23	399	7870	296990
Mar-18	10,54	30250	7500	402782
Apr-18	20,62	16820	7500	410989
Mei-18	20,39	12390	7870	410989
June-18	15,24	7260	8250	380655
July-18	9,48	11160	8000	268894
Jumlah	276,36	387.889,00	148.890,00	8.180.008,00

Gambar 4. Jumlah Variable

Hitung $(X_1)^2, (X_2)^2, (X_3)^2$, dan $\sum(X_1)^2, \sum(X_2)^2, \sum(X_3)^2$

	X1 ²	X2 ²	X3 ²
39,43	969699600	56250000	
121,32	491508900	56250000	
102,66	1686744900	57002500	
313,50	504900900	68062500	
356,82	580328100	68062500	
235,49	350812900	60996100	
140,67	177156100	62410000	
302,91	439321600	63202500	
389,19	71910400	56250000	
253,18	1646088400	64000000	
172,46	866713600	64000000	
95,49	1141088400	61936900	
187,25	969699600	61936900	
369,71	159201	56250000	
111,18	915062500	56250000	
425,09	282912400	63202500	
415,74	153512100	61936900	
232,34	5270600	61936900	
89,83	124545600	68062500	
Jumlah	4.354,27	9.943.393.301,00	1.167.998.700,00

Gambar 5. Jumlah pengkuadratan

Hitung $X_1 * X_2, X_1 * X_3, X_1 * X_3, X_2 * X_3, X_1 * Y, X_2 * Y, X_3 * Y$, dan $\sum X_1 X_2, \sum X_1 X_3, \sum X_1 X_3, \sum X_2 X_3, \sum X_1 Y, \sum X_2 Y, \sum X_3 Y$

Hitung nilai a, b1, b2, b3 menggunakan perhitungan matriks. Cara dari perhitungan matriks secara manual dapat dilihat pada pembahasan nomor 4, maka setelah memasukan rumus matriks yang di peroleh hasil matriks sebagai berikut.

$$\sum y^1 = na + \sum b_1 X_1 + \sum b_2 X_2 + \sum b_3 X_3$$

	19		8.180.008
A	276,36	B	121.152.068
	387.889,00		169.533.345.850
	148.890		64.198.775.700

Tabel 2. Perhitungan Matriks

A	-177.630,76
b1	11.310,59
b2	3,96
b3	46,28

Tabel 3. Hasil Perhitungan Matriks

Bulan	Luas Lahan (ha)	Curah Hujan (mm)	Pupuk (kg)	Jumlah Produksi	Output Perhitungan Manual
Jan-19	17,96	175	8000	322242	396.393,86
Feb-19	10,92	149	7500	259420	293.560,94
Mar-19	14,78	172	7500	412263	337.314,23
Apr-19	14,78	50	8000	?	359.977,76
May-19	10,92	100	7500	?	293.376,88
Jun-19	17,96	100	7500	?	373.003,44

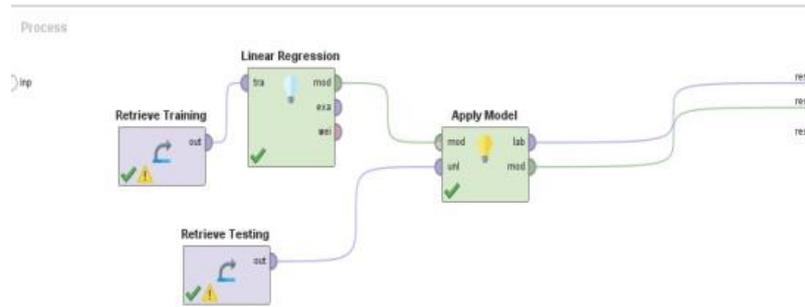
Gambar 6. Output perhitungan manual

Pengujian Menggunakan Rapidminer



Gambar 7. Antar muka halaman utama

Menghubungkan panel susunan operator dapat dilihat pada gambar dibawah dan akan dijelaskan fungsinya, panel operators retrieve yang mana telah dijelaskan sebelumnya berfungsi untuk memproses objek data, linier regression yaitu model algoritma yang berfungsi untuk menghasilkan model berdasarkan algoritma yang telah dipilih yaitu linier regression, apply model yang dimana panel operator ini berfungsi untuk menerapkan model yang digunakan.



Gambar 8. Panel Operators

Setelah semua panel operator terhubung dengan mengklik icon play maka akan muncul tab result hasil prediksi dengan jumlah produksi yang diprediksikan dalam 3 bulan kedepan yakni meliputi bulan April diprediksikan menghasilkan jumlah produksi sebesar : 350862.531, bulan Mei sebesar : 304206.794, Jun sebesar : 389668.603.

Row No.	Jumlah Prod...	prediction,Jumla...	Bulan	Luas Lahan	Curah Hujan...	Pupuk (kg)
1	322242	389922.494	Jan 1, 2019	17.956	175	8000
2	259420	304384.915	Feb 1, 2019	10.919	149	7500
3	412263	351349.696	Mar 1, 2019	14.779	172	7500
4	?	350862.531	Apr 1, 2019	14.780	50	8000
5	?	304206.794	May 1, 2019	10.920	100	7500
6	?	389668.603	Jun 1, 2019	17.960	100	7500

Gambar 9. Hasil Prediksi

Dapat dilihat pada gambar dibawah ini tampilan nilai error, yang dimana factor yang mempengaruhi produksi kelapa sawit di PT Borneo Ketapang Indah adalah Luas Lahan dan Curah Hujan dengan nilai Coefficient masing-masing variable 0,646 variable X1 (Luas Lahan), 0,530 variable X2 (Curah Hujan).

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
Luas Lahan	12139.492	4738.922	0.646	0.679	2.582	0.021	**
Curah Hujan (m...	4.052	1.829	0.530	0.679	2.103	0.052	*
(Intercept)	171238.704	98175.113	?	?	1.744	0.100	

Gambar 10. Tampilan Nilai error

4. Kesimpulan

Berdasarkan hasil dan pembahasan penelitian mengenai Penerapan Data Mining Dalam Prediksi Hasil Produksi Kelapa Sawit Pada PT Borneo Ketapang Indah (BKI) dengan menggunakan proses data mining dengan metode Linear Regression menghasilkan kesimpulan sebagai berikut :

- a. Metode Linear Regression dapat diterapkan untuk memprediksi faktor apa saja yang berpengaruh pada hasil produksi kelapa sawit PT. BORNEO Ketapang Indah (BKI) dengan menggunakan hasil produksi tahun 2017-2019.
- b. Model diuji dengan menggunakan tools Rapidminer dengan membagi data Testing dan Training. Data training di ambil dari tahun 2017-2018, sedangkan data testing diambil dari data tahun 2019. Dari data yang telah diuji menggunakan tools rapidminer maka dapat disimpulkan faktor-faktor apa saja yang mempengaruhi hasil produksi kelapa sawit menggunakan Algoritma Linier Regression.

- c. Hasil prediksi jumlah produksi kelapa sawit pada PT Borneo Ketapang Indah dalam 3 bulan kedepan meliputi bulan April adalah sebesar : 350862.531, Mei : 304206.794, dan Juni sebesar : 389668.603.
- d. Dari tiga variable yang digunakan untuk melakukan prediksi jumlah produksi kelapa sawit, meliputi Luas Lahan(X1), Curah Hujan(X2) dan Pupuk(X3). Maka variable Curah Hujan dan Luas lahan yang paling berpengaruh pada jumlah produksi kelapa sawit dengan masing-masing nilai coefficient Luas Lahan (X1) sebesar 0,646 dan Curah Hujan (X2) sebesar 0,530. Dengan perhitungan metode Linier Regression dapat menjadi solusi bagi perusahaan dalam mengatasi masalah ketidakstabilan dalam hasil produksi kelapa sawit.

Daftar Pustaka

- [1] D. Setyamidjaja, *Kelapa sawit : teknik budi daya; panen dan pengolahan*. Jakarta: Kanisius, 2006.
- [2] D. Setyamodjaja, *Kelapa sawit : teknik budi daya, panen dan pengolahan / Djoehana Setyamidjaja*, Ed. rev.. Yogyakarta: Jakarta : Kanisius, 2006, 2006.
- [3] A. Kurniadi, Jasmir, and Y. Novianto, "Penerapan Metode Regresi Linier untuk Memprediksi Kebiasaan Pelanggan Studi Kasus: PT. Mensa Binasukses," vol. 2, No.2, J, 2020.
- [4] H. W. Herwanto, T. Widiyaningtyas, and P. Indriana, "Penerapan Algoritme Linear Regression untuk Prediksi Hasil Panen Tanaman Padi," vol. 8, No. 4, 2019.
- [5] D. Virdaus and P. T. Prasetyaningrum, "Penerapan Data Mining Untuk Memprediksi Harga Bawang Merah Di Yogyakarta Menggunakan Metode K-Nearest Neighbor," *J. Inf. Syst. Artif. Intell.*, vol. 1, no. 1, pp. 1–8, 2020.
- [6] I. Budiman and N. A. Artesya, "Aplikasi Data Mining Menggunakan Multiple Linear Regression Untuk Pengenalan Pola Curah Hujan," *J. Ilmu Komput.*, vol. 02, no. 01, pp. 34–44, 2015.
- [7] I. L. L. Gaol, S. Sinurat, and E. R. Siagian, "Implementasi Data Mining Dengan Metode Regresi Linear Berganda Untuk Memprediksi Data Persediaan Buku Pada Pt. Yudhistira Ghalia Indonesia Area Sumatera Utara," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 130–133, 2019, doi: 10.30865/komik.v3i1.1579.
- [8] M. Iqbal Ramadhan, "Penerapan Data Mining Untuk Analisis Data Bencana Milik Bnpb Menggunakan Algoritma K-Means Dan Linear Regression," *J. Inform. dan Komput.*, vol. 22, no. 1, pp. 57–65, 2017.
- [9] A. Saifudin, "Metode Data Mining Untuk Seleksi Calon Mahasiswa," *J. Teknol.*, vol. 10, no. 1, pp. 25–36, 2018.
- [10] R. Yanto, "Implementasi Data Mining Estimasi Ketersediaan Lahan Pembuangan Sampah menggunakan Algoritma Simple Linear Regression," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 2, no. 1, pp. 361–366, 2018, doi: 10.29207/resti.v2i1.282.