

## IMPLEMENTASI ALGORITMA MULTIPLE LINEAR REGRESSION DALAM MENGESTIMASI HASIL PANEN TANAMAN TEMBAKAU

Diah Nawang Wulan<sup>1</sup>, Mula Agung Barata<sup>2</sup>, Ita Aristia Sa'ida<sup>3</sup>

Universitas Nahdlatul Ulama Sunan Giri, Bojonegoro

email: <sup>1</sup>nawangwulan2512@gmail.com, <sup>2</sup>mula.ab26@gmail.com,

<sup>3</sup>itaaristia@unugiri.ac.id

**Abstract:** *The purpose of this study is to use the multiple linear regression algorithm to estimate tobacco crop yields in Balongrejo Village. Four independent variables are included in the data: rainfall, irrigation volume, fertilization amount, and seed count. Both statistical software and manual analysis were used. The findings indicate that 95.5% of the yield variation can be explained by the regression model. The t-test finds three significant predictors, while the F-test validates the model's overall significance.*

**Keyword:** *multiple linear regression, tobacco yield, predictive model, t-test, f-test*

**Abstrak:** Penelitian ini bertujuan untuk menaksir panen tembakau petani di Desa Balongrejo menggunakan algoritma regresi linier khusus. Data yang digunakan terdiri dari empat variabel dasar: jumlah bit, jumlah pembelian, jumlah transaksi, dan jumlah jam. Analisis dilakukan secara manual dan dengan bantuan alat statistik. Hasil analisis data menunjukkan bahwa model regresi dapat menjelaskan 95,5% variansi dalam data. Selain itu, uji F menunjukkan semua variabel memiliki pengaruh yang signifikan secara bersamaan, sedangkan uji t mengidentifikasi tiga variabel yang memiliki pengaruh signifikan secara terpisah.

**Kata kunci:** regresi linear berganda, panen tembakau, prediksi hasil, uji t, uji f.

### PENDAHULUAN

Sektor pertanian memegang peranan krusial dalam menopang perekonomian Indonesiadengan komoditas tembakau menjadi salah satu andalan di beberapa wilayah. Desa Balongrejo yang terletak di kabupaten Bojonegoro merupakan contoh nyata sebuah komunitas agraris yang sangat bergantung pada budidaya tembakau sebagai tulang punggung mata pencarian petani lokal.

Namun, karakteristik tanaman tembakau yang sangat sensitif seperti karena ketidakpastian cuaca, keragaman teknik budaya, , hingga fluktuasi harga input pertanian yang tidak stabil, semuanya berkontribusi pada kesulitan dalam memprediksi hasil panen tembakau secara akurat. [Click or tap here to enter text.](#)

Fenomena ketidakpastian ini secara langsung berdampak pada perencanaan petani, mulai dari alokasi sumber daya, manajemen resiko, hingga proyeksi pendapatan. Tanpa perkiraan hasil panen yang memadai, petani seringkali dihadapkan pada dilema dalam menentukan kapan waktu terbaik untuk menanam, berapa banyak pupuk yang harus digunakan, atau strategi pemasaran seperti apa yang paling efektif. Oleh karena itu, kebutuhan akan metode prediktif yang andal dan akurat untuk mengestimasi hasil panen tembakau menjadi sangat mendesak. Adanya alat prediksi semacam ini tidak hanya akan membantu petani membuat keputusan yang lebih terinformasi sebelum masa panen tiba, tetapi juga dapat meningkatkan efisiensi operasional dan stabilitas ekonomi mereka.

Dalam konteks inilah, penelitian ini hadir dengan memanfaatkan algoritma Multiple linear regression, sebuah pendekatan statistik yang terbukti efektif dalam memodelkan hubungan antara beberapa variabel prediktor dengan satu variabel respons. Regresi linear berganda memungkinkan identifikasi dan kuantifikasi sejauh mana setiap faktor, seperti jumlah bibit yang ditanam, frekuensi dan volume pemupukan, intensitas penyiraman, serta curah hujan, secara individual dan kolektif mempengaruhi total hasil panen tembakau. Dengan menganalisis data historis dan menerapkan model ini, diharapkan dapat dibangun sebuah sistem estimasi yang mampu memberikan proyeksi hasil panen yang lebih presisi, sehingga memberdayakan petani Desa Balongrejo untuk menghadapi tantangan pertanian tembakau dengan lebih percaya diri dan strategis.

## METODE

Tujuan dari penelitian ini adalah untuk mengestimasi hasil tembakau di Desa Balongrejo menggunakan pemodelan regresi. Desain penelitian ini dimaksudkan untuk diterapkan lebih lanjut oleh peneliti lain atau mahasiswa, dengan metodologi yang ditetapkan dengan jelas.

Algoritma Multiple linear regression merupakan pemodelan hubungan antara satu variabel dependen dan dua atau lebih variabel independen. Dalam konteks ini, multiple linear regression digunakan untuk memperkirakan beberapa efek signifikan dari berbagai faktor input pada hasil tembakau.

Bentuk rumus umum dalam *Multiple Linear Regression* dapat dinyatakan secara statistik sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$$

Keterangan:

Y = Variabel dependen  
 $X_1, X_2, \dots, X_n$  = Variabel independen  
 $\beta_0$  = Intercept (nilai konstanta)  
 $\beta_1, \beta_2, \dots, \beta_n$  = koefisien regresi untuk masing-masing variabel *Independent*  
 $\epsilon$  = *error* atau gangguan *residual*

Metode *Multiple Linear Regression* di perlukan Intercape atau koefisien-koefisien dalam perhitungannya, koefisien-koefisien  $\beta_0, \beta_1, \beta_2, \dots, \beta_n$  dihitung menggunakan metode kuadrat terkecil (*least squares*), yang bertujuan meminimalkan jumlah kuadrat dari selisih antara nilai aktual dan prediksi.

Rumus untuk menghitung koefisien-koefisien *Multiple Linear Regression* ( $\beta$ ) adalah sebagai berikut:

$$\beta = (X^T X)^{-1} X^T Y$$

Keterangan:

X : matriks dari variabel independent.  
 Y : Vektor hasil Variabel independent.  
 $X^T$  : Transpos dari matriks X.  
 $(X^T X)^{-1}$  : Invers dari matriks  $X^T X$

Metode *Multiple Linear Regression* diperlukan bebrapa tahapan perhitungan untuk mengukur akurasi prediksi menggunakan *RMSE (Root Mean Squared Error)* dan *MAPE (Mean Absolute Error)*. *RMSE (Root Mean Squared Error)* digunakan untuk mengukur rata-rata kesalahan kuadrat antara nilai aktual dan nilai prediksi, yang memberikan gambaran seberapa jauh prediksi model dari data yang sebenarnya.

Rumus *RMSE (Root Mean Squared Error)* sebagai berikut:

$$RMSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Keterangan:

$Y_i$  : nilai hasil panen aktual pada data ke-i.  
 $\hat{Y}_i$  : nilai hasil panen prediksi pada data ke-i.  
 n : jumlah total data.

Koefisien determinasi ( $R^2$ ) merupakan indikator statistik yang menunjukkan seberapa besar proporsi variansi variabel dependen (Y) yang bisa dijelaskan oleh variabel-variabel independen (X) dalam model. Rumus Koefisien determinasi ( $R^2$ ) sebagai berikut:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Keterangan:

$y_i$  : nilai aktual

$\hat{y}_i$  : nilai prediksi

$\bar{y}$  : Rata-rata keseluruhan Y aktual.

### HASIL DAN PEMBAHASAN

Penelitian ini didasarkan pada data awal yang dikumpulkan secara komprehensif dari petani tembakau di Desa Balongrejo, Kabupaten Bojonegoro. Dataset yang digunakan menunjukkan hubungan antarasatu variable terikat dan tiga variable bebas. Variabel terikat, yang berfungsi sebagai estimasi, adalah berat dalam kilogram. Sebaliknya, variabel yang diidentifikasi dan diukur meliputi jumlah bibit yang ditanam (dalam batang), jumlah total pemupukan (dalam kilogram), volume udara yang digunakan untuk penyiraman (dalam liter), dan total curah hujan selama periode tanam (dalam milimeter). Variabel-variabel yang terwakili dalam setiap data sampel memberikan spektrum data yang representatif untuk memodelkan hubungan antara faktor input dan output. Berikut adalah dataset hasil panen tembakau petani desa Balongrejo.

**Table 1 Dataset**

Jumlah bibit	pemupukan	penyiraman	Curah hujan	Hasil panen
1000	50	500	150	700
1200	60	550	152	750

800	45	400	153	650
1500	70	600	151	800
1100	55	520	150	720
900	40	450	149	640
1300	65	570	151	760
950	50	480	150	670
1400	75	620	151	820
1000	50	500	150	700
...	...	...	...	...
1050	53	510	151	725
1220	68	590	150	765
1120	61	570	151	740
1270	74	610	150	790
1180	69	590	151	755

### Algoritma Multiple Linear Regression

Berdasarkan perhitungan matriks yang telah dilakukan, diperoleh hasil sebagai berikut:

#### Matriks X'X

Hasil perhitungan matriks X'X menunjukkan hubungan antara variabel independen:

$$X'X = \begin{bmatrix} 100 & 115270 & 6355 & 56590 & 1508 \\ 11527 & 13454 & 7443 & 6588 & 1738 \\ 6355 & 7443 & 4133 & 3646 & 9587 \\ 5659 & 6588 & 3646 & 3231 & 8536 \\ 1508 & 1738 & 9587 & 8536 & 2275 \end{bmatrix}$$

#### Matrik (X'X)<sup>-1</sup>

$$= \begin{bmatrix} 254,5 & 0,001 & 0,155 & -0,029 & -1,64 \\ 0,001 & 5,570 & -2,9 & -7,46 & -9,81 \\ 0,155 & -2,92 & 0,001 & -0,0002 & -0,0005 \\ -0,02 & -7,46 & -0,00028 & 7,009 & 0,0001 \\ -1,649 & -9,81 & -0,00058 & 0,0001 & 0,0108 \end{bmatrix}$$

#### Matriks X'Y

$$X^T Y = \begin{bmatrix} 7517 \\ 8722 \\ 4820 \\ 4277 \\ 1133 \end{bmatrix}$$

#### Persamaan Regresi

Dari perhitungan matriks di atas, diperoleh persamaan regresi linear berganda sebagai berikut:

$$Y = 95,24 + 0,14X_1 + 1,09X_2 + 0,31X_3 + 1,64X_4$$

Berikut adalah tabel hasil persamaan:

#### Table 2 Hasil perhitungan

Jumlah bibit	pupukan	penyiraman	Curah hujan	Hasil panen	Hasil Prediksi
1000	50	500	150	700	692,94
1200	60	550	152	750	751,316
800	45	400	153	650	632,854
1500	70	600	151	800	819,278
1100	55	520	150	720	718,96
900	40	450	149	640	650,702
1300	65	570	151	760	775,698
950	50	480	150	670	679,57
1400	75	620	151	820	816,298
..	..	..	..	..	..
1120	51	570	151	740	745,278
1270	74	510	150	790	791,67
1180	59	590	151	755	768,738

**Uji signifikan model**

**Uji F (Uji kelayakan model)**

Uji F dilakukan untuk menguji signifikansi model secara keseluruhan:

Niai F-statistik : 505,45

Signifikan F(p-value): 4,45 x 10<sup>-63</sup>

Interpretasi:

Karena p-value < 0.05, model regresi ini signifikan secara statistik dalam memprediksi hasil panen tembakau.

**Uji t (Uji signifikan parsial)**

Uji t dilakukan untuk menguji signifikan masing masing variabel independen:

**Table 3 uji signifikan parsial**

variabel	Koefisien	t-hit	p-value	signifikan
Konstanta	95.24	0,595	0,553	X
Jumlah bibit (X1)	0,144	6,118	2,10	√
Jumlah pemupukan (X2)	1,088	2,430	0,017	√
Jumlah penyiraman (X3)	0,36	3,644	0,0004	√
Curah hujan (X4)	1.638	1,569	0,120	X

**Evaluasi model**

**R-Square (R<sup>2</sup>)**

Rumus :

$$R^2 = 1 - \frac{SSE}{SST}$$

$$= 1 - \frac{9561,98}{213061,00}$$

$$= 1 - 0,0449$$

$$= 0,9551$$

**RMSE (Root mean squared error)**

Rumus:

$$RMSE = \sqrt{\frac{SSE}{n}} = \sqrt{\frac{9561,98}{100}} = \sqrt{95.6198} = 9,78$$

**Hasil akhir**

R2 = 0,955 (95,5%)

RMSE = 9,78 Kg

**SIMPULAN**

Penelitian ini berhasil mengembangkan model prediksi hasil tembakau dengan akurasi tinggi (R2 95,5%) menggunakan regresi linier berganda. Salah satu temuan yang paling mencolok adalah tidak signifikannya curah hujan sebagai faktor penentu, yang menunjukkan bahwa dalam sistem pertanian modern dengan irigasi terkendali, ketergantungan pada faktor alam dapat diminimalkan

Model ini menggambarkan bagaimana tiga faktor budidaya utama bibit, pupuk, dan penyiraman dapat memberikan kontrol yang lebih signifikan terhadap produktivitas tanaman. Secara metodologis, penelitian ini berhasil menghubungkan teori statistik klasik dengan kebutuhan praktis di lapangan melalui analisis komputer yang menyeluruh.

Hasilnya tidak hanya memberikan dasar untuk mengembangkan komoditas pertanian yang lebih persuasif, tetapi juga menciptakan peluang untuk mengembangkan komoditas pertanian lainnya. Hal ini membuktikan bahwa transformasi digital dalam industri farmasi Semua ini menunjukkan bahwa transformasi digital bukanlah hal yang tak terelakkan.

## DAFTAR PUSTAKA

- Ali, M., & Hariyadi, B. W. (2018). Teknik Budidaya Tembakau. *Universitas Merdeka Surabaya*, 1–8.
- badan pusat statistik. (2023). *Sensus Pertanian tahun 2023*. <https://sensus.bps.go.id/main/index/st2023>
- Eliza, A. L., Manalu, D. R., & Yohanna, M. (2024). PREDIKSI HARGA KELAPA SAWIT MENGGUNAKAN METODE REGRESI LINEAR BERGANDA Studi Kasus PT. Bakrie Sumatera Plantations, Tbk. *METHOMIKA Jurnal Manajemen Informatika Dan Komputerisasi Akuntansi*, 8(1), 89–95. <https://doi.org/10.46880/jmika.vol8n01.pp89-95>
- Himawan, I., Nurdiawan, O., & Dwilestari, G. (2022). Algoritma Regresi Linier Berganda. *JURSIMA Jurnal Sistem Informasi Dan Manajemen*, 10(3), 239–247.
- Maulana, A., Martanto, M., & Ali, I. (2024). Prediksi Hasil Produksi Panen Bawang Merah Menggunakan Metode Regresi Linier Sederhana. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(4), 2884–2888. <https://doi.org/10.36040/jati.v7i4.7281>
- Maulana, G., & Danar Dana, R. (2024). Prediksi Hasil Produksi Jagung Di Jawa Barat Dengan Metode Algoritma Regresi Linear Menggunakan Google Collab. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 827–837. <https://doi.org/10.36040/jati.v8i1.8816>
- Media, P., Momentum, P., Impuls, D. A. N., Sistem, B., Android, O., Meningkatkan, U., Berpikir, K., Siswa, K., & Aliyah, M. (2017). *Digital Digital Repository Repository Universitas Universitas Jember Jember Digital Digital Repository Repository Universitas Universitas Jember Jember*.
- Nursela, I., Agus, R. T. A., & Rohminatin, R. (2024). Penerapan Data Mining Untuk Memprediksi Jumlah Produksi Kelapa Sawit Dengan Metode Regresi Linear Berganda Di Pt. Bsp Tbk. *J-Com (Journal of Computer)*, 4(2), 145–153. <https://doi.org/10.33330/j-com.v4i2.3213>
- Rachman, R., Kusdinar, A. B., Indrayana, D., Informatika, T., Sukabumi, U. M., Sukabumi, K., & Mungil, T. (2024). *PENERAPAN REGRESI LINEAR BERGANDA DALAM PREDIKSI DAN*. 8(5), 10499–10506.
- Radar Jatim. (2023). *Harga tembakau di Bojonegoro*. <https://radarbojonegoro.jawapos.com/ekonomi/714945445/harga-tembakau-bojonegoro-tembus-rp-45-ribu-per-kilogram>
- Raehan, M. F. I., Kusdinar, A. B., & Indrayana, D. (2024). *MEMPREDIKSI HASIL PANEN KACANG KEDELAI ( STUDI KASUS : KECAMATAN SURADE )*. 8(5), 10572–10579.
- Ronaldi, A. A., & Nanang Hunafi. (2020). Implementasi Data Mining Untuk Prediksi. *Seminar Nasional Hasil Penelitian & ...*, 1(1), 250–257. <http://jurnal.poliupg.ac.id/index.php/snp2m/article/download/2382/2094>