

---

## PREDIKSI LUAS PANEN DI KECAMATAN PURWODADI MENGGUNAKAN ALGORITMA REGRESI LINEAR BERGANDA

Muhammad Akbar Mustofa<sup>1</sup>; Andre Tiyono<sup>2</sup>; Agus Susilo Nugroho<sup>3</sup>

---

### ABSTRACT

*Agriculture, particularly rice cultivation, is highly vulnerable to climate change because it depends on water cycles and weather conditions to maintain productivity. Climate change affects crop growth, development, and yields, as agricultural activities are heavily dependent on weather and climate. This study utilizes data mining to introduce a new breakthrough in addressing rice farming issues in Grobogan Regency, Purwodadi District. The method used is multiple linear regression, with the dependent variable being harvested area and the independent variables including planted area and rainfall. The objective of this research is to test and develop data mining methods to predict yield levels, thereby assisting local governments in decision-making during crop failures, based on agricultural data from 2019-2023. The research process involves data collection, preprocessing, algorithm implementation, and result evaluation. The analysis shows that the multiple linear regression model provides reasonably accurate predictions, with a Root Mean Square Error (RMSE) value of 209.042 and a Relative Root Squared Error (RRSE) of 0.111. Furthermore, the analysis reveals that planted area significantly influence the harvested area. These findings offer insights for local governments as policymakers in providing aid during crop failures.*

**Keywords:** *Data mining*1; *Yield prediction*2; *Multiple linear regression*3

---

#### Correspondence:

Muhammad Akbar Mustofa

Universitas An Nuur, Email; [101akbarm101@gmail.com](mailto:101akbarm101@gmail.com)

---

## PENDAHULUAN

Indonesia merupakan negara yang mempunyai kekayaan sumber daya alam yang melimpah. Sumber daya alam ini meliputi sumber daya air, lahan, hutan, laut, serta keanekaragaman binatang dan tanaman yang ada dan terdistribusi dari sabang sampai merauke. Terutama sumberdaya pertanian dan perkebunan yang menjadi produksi bahan pangan pokok di Indonesia. Pertanian sebagai kegiatan bercocok tanam yang menjadi

dasar utama dalam menyediakan makanan yang diperlukan untuk kebutuhan gizi dan energi manusia. Terutama padi yang menjadi kebutuhan utama bagi penduduk Indonesia yang menjadi sumber tenaga dan karbohidrat. Selain itu, padi juga merupakan tanaman yang sangat vital bagi jutaan petani kecil yang tersebar di berbagai daerah di Indonesia.

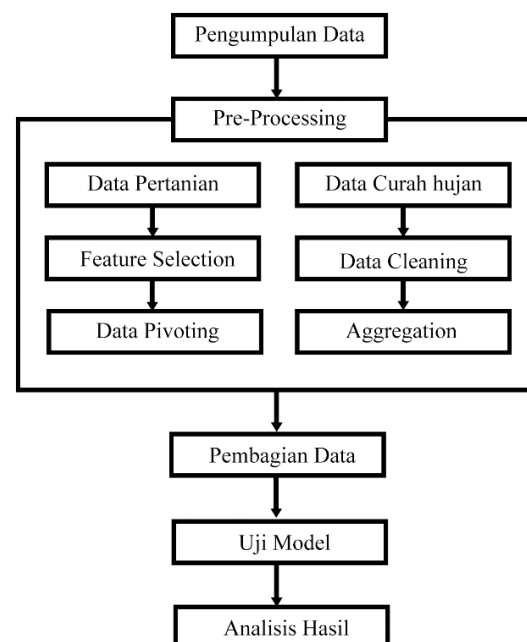
Pertanian sangat rentan terhadap dampak perubahan iklim karena bergantung pada siklus air dan kondisi

cuaca untuk mempertahankan produktivitas. Perubahan iklim memengaruhi pertanian dengan cara mempengaruhi pertumbuhan, perkembangan, dan hasil tanaman, karena aktivitas pertanian sangat tergantung pada kondisi cuaca dan iklim. Perubahan iklim yang tidak wajar sangat berdampak pada sektor pertanian yang dapat dibuktikan dengan hasil panen yang berada di kabupaten grobogan pada kecamatan purwodadi, yang mengalami gagal panen dikarenakan terlambat nya datangnya musim hujan atau kekeringan, kegagalan panen artinya mengurangi persediaan beras lokal, sehingga masyarakat terpaksa harus membeli beras inport yang cenderung mempunyai harga yang lebih mahal dari beras lokal, harga beras adalah faktor krusial yang mempengaruhi kehidupan masyarakat, terutama bagi mereka yang memiliki pendapatan terbatas, Untuk memantau harga beras dibutuhkan informasi yang berguna dalam memprediksi hasil panen padi sebagai dasar dari pengambilan keputusan pemerintah dan memberi informasi kepada masyarakat dikala akan adanya kenaikan harga beras, untuk mengatasi masalah tersbut maka di perlukan prediksi yang akurat untuk memprediksi luas panen berdasarkan luas tanam dan curha hujan.

Dalam penelitian ini, digunakan metode regresi linier berganda untuk memperkirakan hasil panen padi di Kabupaten Grobogan dengan attribut luas tanam dan curah hujan. Tujuan dari penelitian ini adalah untuk mengetahui seberapa akurat algoritma regresi linear untuk memprediksi luas panen padi sehingga dapat di gunakan untuk membantu pengambilan keputusan pemerintah dikala diprediksi kegagalan panen.

## METODE PENELITIAN

### Desain Penelitian



Gambar 1. Gambar Desain Penelitian

Gambar di atas adalah bagan pemodelan yang di usulkan oleh peneliti, dalam gambar di atas terdapat langkah langkah penelitian, langkah tersebut diantara lain adalah:

1. Pengumpulan data
2. Pre-processing
3. Pembagian data
4. Pelatihan Model
5. Analisis Hasil

### Pengumpulan data

Pengumpulan data pertanian dilakukan dengan datang secara langsung ke kantor badan pusat statistik dan meminta data yang di butuhkan serta mewawancarai mengenai atribut atribut yang terdapat pada data tersebut, pengumpulan data curah hujan dilakukan dengan mengunjungi situs data online bmkg <https://dataonline.bmkg.go.id>, membuat akun, dan mendownload data yang di butuhkan.

### Pre-processing

*Pre-processing* adalah tahap untuk menghilangkan berbagai masalah yang dapat mengganggu proses data. Ini diperlukan karena banyak data memiliki format yang tidak konsisten. Data preprocessing merupakan teknik awal sebelum melakukan data mining. Proses dalam data *pre-processing* meliputi pembersihan, integrasi, transformasi, dan reduksi data, preprocessing dilakukan kepada data pertanian adalah *feature selection* dan *data pivoting* sedangkan data curah hujan dilakukan *data cleaning* dan proses agregasi.

### Pembagian Data

Pembagian data adalah pemecahan data menjadi dua bagian, yaitu data pelatihan (*Trainng Set*) dan data pegujian (*Testing Set*), pada peneitian ini data akan di belah menjadi 80/20 atau 80% untuk data training dan 20% untuk testing, data training di olah menggunakan metode regresi linear dan data testing untuk pembuktian akurasi prediksi dari data testing

### Pelatihan Model

Untuk menghitung prediksi luas panen dengan menggunakan algoritma regresi linier harus dilakukan beberapa tahap pada data *training* yaitu mencari varians dan kovarians dimana lakukan perhitungan terhadap variabel  $X_n^2$  dan  $Y^2$ , dimana X Adalah variabel bebas (independen) dan Y Adalah variabel terkait (dependen)  $X_n Y$  merupakan perkalian antar kedua vaiabel , $X_n X_n$  merupakan perkalian antar variabel bebas yang berbeda, selanjutnya menghitung koefisien regresi a, b1 dan b2 dengan persamaan sebagai berikut:

$$b1 = \frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_2 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2} \quad (1)$$

$$b2 = \frac{(\sum x_1^2)(\sum x_2 y) - (\sum x_1 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2} \quad (2)$$

$$a = \left(\frac{\sum Y}{n}\right) - b1 \left(\frac{\sum X_1}{n}\right) - b2 \left(\frac{\sum X_2}{n}\right) \quad (3)$$

### Analisis hasil

Data *training* yang sudah di latih dengan regresi linear berganda akan menghasilkan nilai hasil prediksi, hasil perediksi tersebut akan di bandingkan dengan data testing dan di analisis menggunakan *Root Mean Squared Error* (RMSE) dan *Root Relative Squared Error* (RRSE), dari nilai RMSE dan RRSE tersebut akan di ketahui ke akurasion dari prediksi yang di lakuakn, semua kalkulasi akan di lakukan pada aplikasi rapid miner

## HASIL DAN PEMBAHASAN

### Tahap Pengumpulan Data

Pengumpulan data luas tanam dan luas panen di lakukan dengan observasi langsung ke kantor BPS (Badan Pusat Statistik) Purwodadi di sektor permintaan data pada tanggal 6 juni 2024, dengan mengajukan permohonan permintaan data luas tanam dan luas panen pada kecamatan purwodadi dari tahun 2019 – 2023, Pengumpulan data curah hujan di lakukan dengan mengunjungi website *Repository online* BMKG (Badan Meteorologi, Klimatologi, dan Geofisika) yaitu “<https://dataonline.bmkg.go.id/home>”, data dapat di unduh setelah membuat akun dan mengisi survei kepuasan pelayanan yang muncul pada proses pengunduhan data.

### Tahap *Pre Processing*

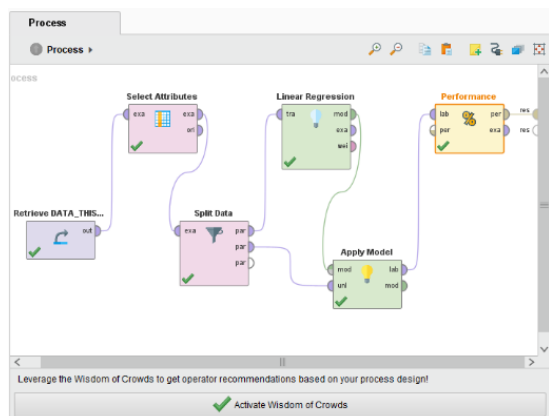
Pada tahap *Pre-processing* data pertanian akan melalui tahap *feature selection* pada tahap ini akan di pilih variabel yang akan digunakan, setelah itu akan dilakukan *data pivoting* yaitu pengaturan ulang kolom atau baris sehingga data menjadi lebih terstruktur. Sedangkan data curah hujan akan melalui tahap *data celaning* dimana data yang yang mempunyai nilai *0/null value* akan di hilangkan, lalu akan melalui tahap *aggreggasi* yaitu penggabungan data yang mempunyai level rendah menjadi data yang mempunyai level yang lebih tinggi, hal ini dilakukan sehingga data curah hujan dan data pertanian mempunyai format waktu yang sama yaitu per bulan.

Tabel 1. Tabel hasil tahap pre-processing

Waktu tanam	Luas Tanam (Hk)	curah hujan (MM)	waktu panen	Luas Paenen (Hk)
Jan-19	176,1	7,24	Apr-19	208,0
Feb-19	566,9	8,3592	May-19	535,0
Mar-19	3.458,8	5,8483	Jun-19	3.507,2
Apr-19	557,3	7,6548	Jul-19	508,9
Oct-19	286,4	0,2645	Jan-20	286,4
Nov-19	1.811,2	2,3566	Feb-20	1.811,1
Dec-19	1.609,9	7,7033	Mar-20	1.600,2

Tabel diatas sampel data setelah melalui tahap pre-processing

## Tahap Pemrosesan Data



Gambar 2. Gambar alur proses Rapid Miner

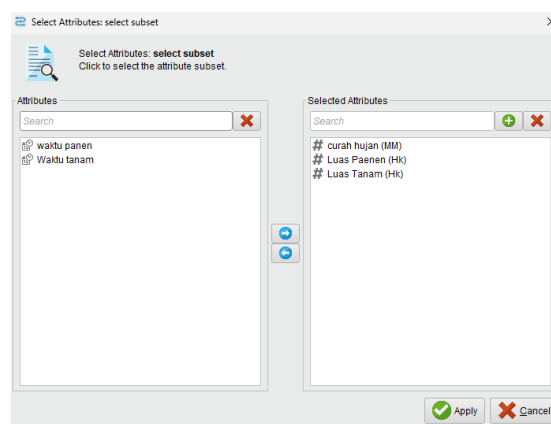
Gambar di atas adalah alur proses dilakukan menggunakan aplikasi rapid miner, proses tersebut meliputi: *Retrieve*, *Select Attributes*, *Split data*, *Linear Regression*, *Apply Model*, dan *Performance*, proses-proses tersebut akan di bahas sebagai berikut:

### 1. Retrieve data

*Retrieve data* adalah proses memuat data pada penyimpanan komputer ke aplikasi rapid miner, sebelum data dapat di olah, harus di lakukan pengaturan *role* data pada variabel dependen (luas panen), hal ini dilakukan agar rapid miner mengetahui jika variabel tersebut adalah variabel target yang nantinya akan di lakukan prediksi.

### 2. Select attributes

*Select Attributes* adalah tahap pemilihan atribut atau variabel yang nantinya akan di olah dengan algoritma regresi linear

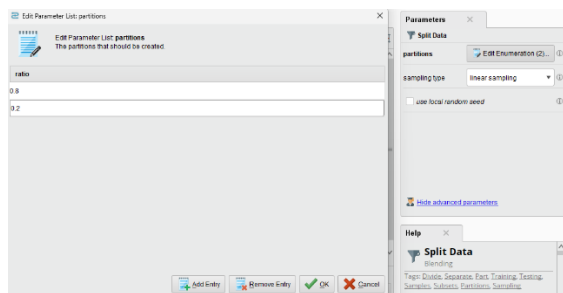


Gambar 3. Gambar Proses pemilihan Atribut

Gambar diatas adalah proses pemilihan variabel, kolom di sebelah kiri merupakan data yang tidak di gunaka dan kolom sebelah kanan adalah variabel yang akan di gunakan yaitu luas panen, curah hujan, dan luas tanam.

### 3. Split Data

*Split data* adalah proses pemisahan data menjadi dua bagian, peneliti memecah data yang berjumlah 22 baris menjadi dua set data, data tersebut akan di pisah dengan rasio 80% untuk *training* dan 20% untuk *testing*, *data training* digunakan untuk mempelajari pola, hubungan, dan karakteristik data sehingga model dapat memberikan prediksi yang akurat, Sedangkan data testing di gunakan untuk untuk menguji dan mngevaluasi dari prediksi yang telah di dapatkan



Gambar 4. Pengaturan rasio dan tipe sampling

Gambar di atas adalah proses *split data* pada rapid miner, dimana rasio entri pertama berisi 0,8 yang artinya mengatur data *training* 80% dari seluruh data dan entri kedua berisi 0,2 yang artinya mengatur data *testing* 20% dari seluruh data, tipe *sampling* dari partisi di atur sebagai *linear sampling* atau secara runtut karena data berbentuk *timeseries* atau jangka waktu sehingga data testing di ambil dari data yang paling baru.

#### 4. Penerapan Regresi Linear

pada tahap ini akan kakukan perhitungan untuk mendapatkan konstanta, koefisien, dan persamaan regresi pada data training yang di awali dengan perhitungan varians dan kovarians yang akan di jelaskan sebagai berikut:

$$\sum X_1 = 22.862$$

$$\sum X_2 = 124,8$$

$$\sum Y = 22.813,5$$

$$n = 18$$

Menghitung Varians

$$\Sigma x_1^2 = \Sigma x_1^2 - \frac{(\Sigma X_1)^2}{n} = 24958667,42 \quad (4)$$

$$\Sigma x_2^2 = \Sigma x_2^2 - \frac{(\Sigma X_2)^2}{n} = 211,78$$

$$\Sigma y^2 = \Sigma y^2 - \frac{(\Sigma Y)^2}{n} = 24569590,61$$

Menghitung Kovarians

$$\Sigma x_1 x_2 = \Sigma x_1 x_2 - \frac{(\Sigma X_1)(\Sigma X_2)}{n} = -20746,49$$

$$\Sigma x_1 y = \Sigma x_1 y - \frac{(\Sigma X_1)(\Sigma Y)}{n} = 2 4740712,87$$

$$\Sigma x_2 y = \Sigma x_2 y - \frac{(\Sigma X_2)(\Sigma Y)}{n} = -19879,74$$

Setelah Varians dan Kovarians ditemukan kemudian dilakukan perhitungan koefisiensi dan intercept.

$$b_1 = \frac{(\Sigma x_2^2)(\Sigma x_1 y) - (\Sigma x_2 y)(\Sigma x_1 x_2)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2} = 0,994 \quad (5)$$

$$b_2 = \frac{(\Sigma x_1^2)(\Sigma x_2 y) - (\Sigma x_1 y)(\Sigma x_1 x_2)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2} = 3,524 \quad (6)$$

$$a = \left(\frac{\Sigma Y}{n}\right) - b_1 \left(\frac{\Sigma X_1}{n}\right) - b_2 \left(\frac{\Sigma X_2}{n}\right) = -19,757 \quad (7)$$

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
Luas Tanam (H)	0.994	0.011	1.002	0.919	89.418	0	----
curah hujan (MM)	3.524	3.817	0.010	0.919	0.923	0.370	
(Intercept)	-19.757	35.643	?	?	-0.554	0.588	

Gambar 5. Gambar Hasil perhitungan rapid miner

Gambar di atas adalah gambar hasil pengolahan pada aplikasi rapid miner, attribut di atas dapat di interpretasikan sebagai berikut:

1. *Intercept* (a) : Adalah nilai konstanta (Y) ketika semua variable independen (X1,X2) atau Luas tanam dan Curah hujan bernilai 0, dalam konteks ini *intercept* bernilai -19,757 yang artinya itu adalah nilai prediksi tanpa di pengaruhi variabel independen

2. *Coefficient* Luas Tanam (b1) : perubahan rata-rata pada konstanta (Y) yang mengasumsikan jika variabel *intercept* (a) dan *coefficient* curah hujan (b2) adalah konstan, dalam konteks ini *coefficient* luas tanam (b1) bernilai 0,994 yang artinya setiap kenaikan luas tanam sebesar 1 , maka nilai luas panen akan naik sebesar 0,994 dengan asumsi variabel curah hujan di anggap konstan
3. *Coefficient* Cruah hujan (b2) : perubahan rata-rata pada konstanta (Y) yang mengasumsikan jika variabel *intercept* (a) dan *coefficient* luas tanam (b1) adalah konstan, dalam konteks ini *coefficient* curah hujan (b2) bernilai 3,524 yang artinya setiap kenaikan curah hujann sebesar 1, maka nilai luas panen akan naik sebesar 3,524 degnan asusmsi variabel luas tanam di anggap konstan
4. Hasil pengolahan menggunakan rapid miner menunjukkan bahwa pada kolom “code”, variabel luas tanam memiliki bintang, bintang tersebut dapat dianggap sebagai skor untuk menilai apakah variabel mempunyai dampak pada hasil prediksi, dalam konteks ini variabel

luas tanam mempunyai empat bintang yang menandakan variabel tersebut sangat berdampak pada hasil prediksi, sedangkan variabel curah hujan tidak mempunyai bintang yang berarti berdampak kecil pada hasil prediksi.

### 5. Apply model

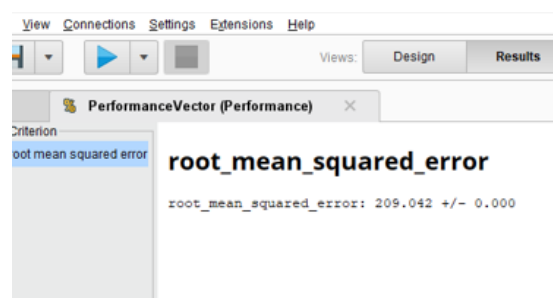
Row No.	Luas Paene...	prediction(L...	Luas Tanam...	curah hujan ...
1	1003.300	1041.276	1033.300	9.571
2	3130.900	3546.020	3550.700	10.125
3	90.900	105.587	90.900	9.923
4	4713.600	4685.119	4713.700	5.258

Gambar 6. Gambar hasil perbandingan apply model

Gambar di atas adalah hasil dari *Apply model*, *apply model* adalah perbandingan antara data testing dan data prediksi dimana data prediksi di hitung menggunakan rumus  $y = a + b1X_1 + b2X_2$ . (8)

### 6. Performance

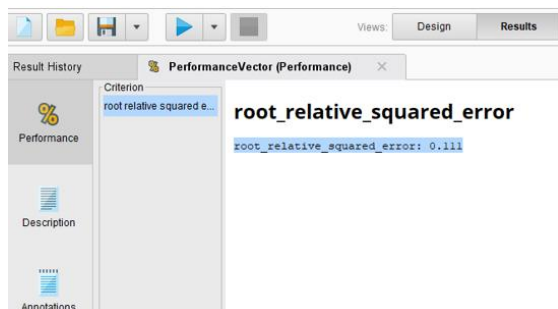
Pada tahap ini akan dilakukan perhitungan performa dari hasil prediksi yang sudah di ketahui, perhitungan prediksi akan menggunakan RMSE (*Root Mean Square Error*) dan RRSE (*Root Relative Square Error*).



Gambar 7. Gambar hasil RMSE



Gambar diatas menunjukkan bahwa nilai RMSE (*Root Mean Squared Error*) berjumlah 209.042 angka tersebut menunjukkan seberapa besar perbedaan antara prediksi dan nilai sebenarnya, semakin kecil nilai RMSE (*Root Mean Squared Error*) mendekati nol semakin bagus performa dari prediksi.



Gambar 8. Gambar hasil RRSE

Gambar di atas menunjukkan bahwa nilai *Root Relative Squared Error* (RRSE) adalah 0.111. Nilai *Root Relative Squared Error* (RRSE) merupakan metrik yang menghitung antara rata-rata kesalahan prediksi model terhadap rata-rata deviasi kesalahan dari model awal yang dihitung sebagai akar perbandingan antara jumlah kuadrat kesalahan model dengan jumlah kuadrat kesalahan dari model dasar (baseline), Nilai *Root Relative Squared Error* (RRSE) sebesar 0,111 menunjukkan bahwa kesalahan prediksi model hanya 11,1% dari kesalahan yang akan terjadi jika menggunakan model dasar. Ini berarti model prediksi memiliki kesalahan relatif kecil dibandingkan dengan model dasar,

dapat di simpulkan bahwa prediksi memiliki akurasi sebesar 88,9%.

## SIMPULAN DAN SARAN

### SIMPULAN

Curah hujan tidak mempunyai dampak yang besar yang di buktikan pada perhitungan variabel curah hujan dan koefisien curah hujan, dan juga di buktikan oleh skor yang di berikan oleh rapid miner, hal ini dapat di karenakan data curah hujan di ambil di ukur pada stasiun pengukuran bmg yang berlokasi di wilayah semarang.

Akurasi dari proses perhitungan RMSE adalah 209.042 dan nilai akurasi *Root Relative Squared Error* (RRSE) menunjukkan angka 0,111 artinya model estimasi memiliki error sebesar 11,1% yang dapat di sumpulkan prediksi mempunyai akurasi sebesar 88,9%

### SARAN

Saran yang diberikan penulis kepada peneliti selanjutnya adalah penambahan variabel yang dapat berpengaruh lebih besar terhadap penanaman padi dan pengambilan data curah hujan dapat di lakukan pada tempat penanaman padi sehingga prediksi lebih akurat

### DAFTAR PUSTAKA

Al-Tashi, Q., Abdulkadir, S. J., Rais, H. M., Mirjalili, S., & Alhussian, H.



- (2020). Approaches to Multi-Objective Feature Selection: A Systematic Literature Review. *IEEE Access*, 8, 125076–125096. <https://doi.org/10.1109/ACCESS.2020.3007291>
- Arinal, V., & Azhari, M. (2023). Penerapan Regresi Linear Untuk Prediksi Harga Beras Di Indonesia. *Jurnal Sains Dan Teknologi*, 5(1), |pp. <https://doi.org/10.55338/saintek.v5i1.1417>
- Diyanti, Martanto, & Bahtiar, A. (2023). Jurnal Informatika Terpadu PREDIKSI HASIL PANEN PADI TAHUN 2023 MENGGUNAKAN METODE REGRESI LINIER DI KABUPATEN INDRAMAYU. *Jurnal Informatika Terpadu*, 9(1), 18–23. <https://journal.nurulfikri.ac.id/index.php/JIT>
- Fitri Santika, D., & Yakub, S. (n.d.). Penerapan Data Mining Untuk Memprediksi Keuntungan Penjualan Melalui Aplikasi Gofood Pada Coffeescape Medan Dengan Menggunakan Metode Regresi Linear Berganda. In *Jurnal CyberTech: Vol. x. No.x*. <https://ojs.trigunadharma.ac.id/>
- Hutahaean, J., & Yusup, D. (2024). PERBANDINGAN METODE LINEAR REGRESSION, RANDOM FOREST & K-NEAREST NEIGHBOR UNTUK PREDIKSI PRODUKSI HASIL PANEN PADI DI PROVINSI JAWA BARAT. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 3).
- Iffah Luthfiyah Arham, & Soeryo Adiwibowo. (2022). Pengaruh Kemarau Panjang 2019 Sebagai Indikasi Perubahan Iklim Terhadap Kesejahteraan Rumah Tangga Petani Padi Desa Tenajar Kidul, Indramayu. *Jurnal Sains Komunikasi Dan Pengembangan Masyarakat*, 6(1), 86–100.
- Kaliappan, J., Srinivasan, K., Mian Qaisar, S., Sundararajan, K., Chang, C. Y., & Suganthan, C. (2021). Performance Evaluation of Regression Models for the Prediction of the COVID-19 Reproduction Rate. *Frontiers in Public Health*, 9. <https://doi.org/10.3389/fpubh.2021.729795>
- Kurniawan, R., Halim, A., & Melisa, H. (2023). KLIK: Kajian Ilmiah Informatika dan Komputer Prediksi Hasil Panen Pertanian Salak di Daerah Tapanuli Selatan Menggunakan Algoritma SVM (Support Vector Machine). *Media Online*, 4(2), 903–912. <https://doi.org/10.30865/klik.v4i2.1246>
- Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Edik Informatika*, 2(2), 213–219. <https://doi.org/10.22202/ei.2016.v2i2.1465>
- Mergono Adi Ningrat, Carolina Diana Mual, & Yohanis Yan Makabori. (2021). Pertumbuhan dan Hasil Tanaman Padi (*Oryza sativa* L.) pada Berbagai Sistem Tanam di Kampung Desay, Distrik Prafi, Kabupaten Manokwari. *Prosiding Seminar Nasional Pembangunan Dan Pendidikan Vokasi Pertanian*, 2(1), 325–332. <https://doi.org/10.47687/snppvp.v2i1.191>
- Moh. Rodiaminollah, & Lu'lu' Nurul Qomariyah. (2023). PEMBANGUNAN SEKTOR PERTANIAN DALAM PERENCANAAN PEMBANGUNAN EKONOMI. *Istisodina Jurnal Ekonomi Syariah & Hukum Islam*, 6(1), 49–55.
- Nur Latifah, A., Sulistiyono, M., Sidauruk, A., Satria, B., & Tofa Nurcholis, M.

- (2023). Prediksi Curah Hujan Menggunakan Algoritma Regresi Linear Berganda. *Jurnal ICT : Information Communication & Technology*, 23(1), 39–44.  
<https://ejournal.ikmi.ac.id/index.php/jict-ikmi>
- Rahman, M. A., Kang, S. C., Nagabhatla, N., & Macnee, R. (2017). Impacts of temperature and rainfall variation on rice productivity in major ecosystems of Bangladesh. *Agriculture and Food Security*, 6(1).  
<https://doi.org/10.1186/s40066-017-0089-5>
- Santoso, A. B., Supriana, T., & Girsang, M. A. (2022). Pengaruh Curah Hujan pada Produksi Padi Gogo di Indonesia. *Jurnal Ilmu Pertanian Indonesia*, 27(4), 606–613.  
<https://doi.org/10.18343/jipi.27.4.606>
- Suad, A., & Wesam, B. (2017). Review\_of\_Data\_Preprocessing\_Techniques. *Journal of Engineering and Applied Science*, 12(16), 4102–4107.
- Surmaini, E., & Faqih, D. A. (n.d.). *Kejadian Iklim Ekstrem dan Dampaknya Terhadap Pertanian Tanaman Pangan di Indonesia Extreme Climate Events and their Impacts on Food Crop in Indonesia*.
- Triyanto, E., Sismoro, H., & Laksito, A. D. (2019). IMPLEMENTASI ALGORITMA REGRESI LINEAR BERGANDA UNTUK MEMPREDIKSI PRODUKSI PADI DI KABUPATEN BANTUL. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, 4(2), 66–75.  
<https://doi.org/10.36341/rabit.v4i2.666>
- Widaningsih, S. (2019). PERBANDINGAN METODE DATA MINING UNTUK PREDIKSI NILAI DAN WAKTU KELULUSAN MAHASISWA PRODI TEKNIK INFORMATIKA DENGAN ALGORITMA C4,5, NAÏVE BAYES, KNN DAN SVM. *Jurnal Tekno Insentif*, 13(1), 16–25.  
<https://doi.org/10.36787/jti.v13i1.78>
- Yuli Mardi. (2019). Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, 2(2), 213–219.
- Zai, C. (2023). IMPLEMENTASI DATA MINING SEBAGAI PENGOLAHAN DATA. In *Portaldata.org* (Vol. 2, Issue 3).