 ISSN:2302-8149
e-ISSN:2540-9719

Optimasi Algoritma *Linear Regression* menggunakan *GridSearchCV* pada Prediksi Produksi Tanaman Padi

Optimization of the Linear Regression Algorithm using GridSearchCV for Rice Crop Production Prediction

[1]Imel, [2]Norhikmah*, [3]Irma Rofni Wulandari, [4]Ali Mustofa, [5]Niken Larasati, [6]Subektiningsih
[1,2,3,4,5]Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta, Indonesia
[6]Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta, Indonesia
[1,2,3,4,5,6]Jl. Ring Road Utara, Condong Catur, Sleman, Yogyakarta
*e-mail: hikmah@amikom.ac.id

(***received*:** 24 November 2025, ***revised*:** 12 December 2025, ***accepted*:** 15 December 2025)

Abstrak

Produksi padi di Provinsi Jawa Tengah mengalami fluktuasi setiap tahun, memengaruhi ketahanan pangan dan distribusi hasil pertanian. Oleh karena itu, diperlukan metode prediksi yang akurat untuk membantu pemangku kepentingan dalam perencanaan sektor pertanian dan pengambilan keputusan strategis. Penelitian ini menerapkan algoritma Linear Regression untuk memprediksi produksi padi berdasarkan data historis 2014–2023 dari website resmi Dinas Pertanian dan Perkebunan Provinsi Jawa Tengah. Model dikembangkan menggunakan regresi linear berganda dengan variabel luas tanam, luas panen, dan produktivitas. Kebaruan penelitian ini terletak pada penerapan hyperparameter tuning menggunakan *GridSearchCV* secara terstruktur untuk mengoptimalkan performa *Linear Regression*, serta integrasi pipeline preprocessing berbasis stabilisasi distribusi data guna meningkatkan akurasi dan kemampuan generalisasi model. Proses penelitian mencakup pengumpulan data, preprocessing, pemodelan, optimasi, evaluasi model, dan deployment dalam aplikasi berbasis website menggunakan Streamlit Cloud. Hasil optimasi dengan *GridSearchCV* menunjukkan akurasi *cross-validation* sebesar 98,26%, menegaskan kemampuan prediksi yang sangat baik. Evaluasi model menghasilkan nilai R^2 sebesar 0,9754 dengan MAE 0,0957, MSE 0,0307, dan RMSE 0,1753, yang menunjukkan kesalahan prediksi rendah dan stabilitas model yang baik. Model yang telah dioptimasi diimplementasikan sebagai aplikasi web menggunakan *Streamlit Cloud* sehingga dapat digunakan secara langsung oleh pengguna. Penelitian selanjutnya disarankan untuk menambahkan variabel seperti curah hujan, suhu, dan jenis varietas padi atau membandingkan performa dengan algoritma lain seperti *Random Forest*, *Support Vector Regression*, atau *Long Short-Term Memory* guna meningkatkan akurasi prediksi.

Kata kunci: Prediksi, *linear regression, machine learning, evaluasi, gridsearchcv, streamlit cloud.*

Abstract

Rice production in Central Java Province fluctuates annually, affecting food security and agricultural output distribution. Therefore, accurate prediction methods are essential to assist stakeholders in agricultural planning and strategic decision-making. This study applies the Linear Regression algorithm to predict rice production based on historical data from 2014 to 2023 obtained from the official website of the Central Java Provincial Agriculture and Plantation Office. The model is developed using multiple linear regression with variables including planted area, harvested area, and productivity. The novelty of this study lies in the structured application of hyperparameter tuning using GridSearchCV to optimize Linear Regression performance, as well as the integration of a preprocessing pipeline based on data distribution stabilization to improve accuracy and model generalization. The research process includes data collection, preprocessing, modeling, optimization, model evaluation, and deployment as a web-based application using Streamlit Cloud. GridSearchCV optimization results indicate a cross-validation accuracy of 98.26%, confirming the model's strong

 ISSN:2302-8149
e-ISSN:2540-9719

predictive capability. Model evaluation shows an R^2 value of 0.9754, with MAE of 0.0957, MSE of 0.0307, and RMSE of 0.1753, indicating low prediction errors and stable model performance. The optimized model is implemented as a web application via Streamlit Cloud, enabling direct use by end-users. For future research, it is recommended to incorporate additional variables such as rainfall, temperature, and rice variety, or to compare performance with other algorithms such as Random Forest, Support Vector Regression, or Long Short-Term Memory (LSTM) to further enhance prediction accuracy.

Keywords: *hyperparameter tuning, linear regression, machine learning, model evaluation, rice production prediction, streamlit cloud*

1 Pendahuluan

Provinsi Jawa Tengah merupakan salah satu wilayah dengan kontribusi produksi padi terbesar di Indonesia dan memegang peranan penting dalam menjaga ketahanan pangan nasional. Pada tahun 2023, Jawa Tengah menempati posisi kedua sebagai penghasil padi terbesar setelah Jawa Timur dengan kontribusi sebesar 17,5% terhadap kebutuhan beras nasional [1]. Peran strategis ini menjadikan kestabilan produksi padi di Jawa Tengah sebagai faktor krusial dalam mendukung sistem pangan nasional.

Namun demikian, produksi padi di Jawa Tengah dalam beberapa tahun terakhir menunjukkan fluktuasi yang signifikan. Pada tahun 2023, produksi padi tercatat sebesar 9,08 juta ton Gabah Kering Giling (GKG), mengalami penurunan sebesar 0,27 juta ton atau 2,91% dibandingkan tahun 2022 yang mencapai 9,36 juta ton GKG. Penurunan ini sejalan dengan berkurangnya luas panen dari 1,69 juta hektare pada tahun 2022 menjadi 1,64 juta hektare pada tahun 2023. Produksi beras untuk konsumsi juga menurun dari 5,38 juta ton menjadi 5,22 juta ton[2]. Tren ini menunjukkan adanya dinamika yang perlu dianalisis untuk mendukung upaya stabilisasi produksi.

Badan Pusat Statistik mengemukakan bahwa perubahan produksi padi dipengaruhi oleh berbagai faktor seperti luas tanam, luas panen, produktivitas, kondisi cuaca, kualitas benih, serta ketersediaan air [2], Variabel luas tanam, luas panen, dan produktivitas merupakan komponen utama dalam memahami dinamika produksi padi di tingkat kabupaten/kota[3].

Dalam penelitian ini, tiga variabel utama yang menjadi fokus adalah luas tanam, luas panen, dan produktivitas [3]. Luas tanam merupakan indikator awal potensi produksi, namun hasil panen tidak hanya bergantung pada jumlah lahan yang ditanami, tetapi juga dipengaruhi faktor eksternal seperti cuaca dan kualitas benih[4]. Perbedaan antara luas tanam dan luas panen dapat mengindikasikan adanya risiko gagal panen akibat hama, penyakit tanaman, atau cuaca ekstrem [5]. sementara produktivitas menjadi indikator efisiensi lahan dalam menghasilkan padi[6].

Dengan kompleksitas faktor-faktor tersebut, kemampuan melakukan prediksi produksi padi yang akurat menjadi penting dalam mendukung pengambilan keputusan berbasis data. Teknik prediksi pada umumnya digunakan untuk memperkirakan nilai kontinu berdasarkan hubungan antara variabel independen dan dependen[7]. sebagaimana diterapkan pada berbagai penelitian sebelumnya seperti prediksi kegagalan siswa dengan *Naïve Bayes* [8].Prediksi hasil pertanian menggunakan *Linear Regression* [9],[10]. *Linear Regression* memiliki keunggulan dalam memodelkan hubungan linear antarvariabel serta memberikan hasil yang mudah diinterpretasikan. Namun, algoritma ini juga memiliki kelemahan, seperti sensitivitas terhadap outlier dan keterbatasan dalam menangani pola non-linear [9]. Penelitian-penelitian sebelumnya yang menggunakan *Linear Regression* untuk memprediksi hasil pertanian menunjukkan akurasi yang cukup tinggi, seperti tingkat keandalan 94,51% dalam memprediksi hasil panen berdasarkan variabel pertanian tertentu [10], atau nilai R^2 sebesar 0,99 pada studi prediksi produksi GKP [5], Produksi padi di Kabupaten Grobogan mengalami tren penurunan sehingga diperlukan model prediksi berbasis Linear Regression menggunakan data BPS dan BMKG, yang setelah dibangun menghasilkan nilai MSE 6550,24 dan MAE 121.247.657,97 sebagai indikator performa prediksi yang cukup baik [11].

Penelitian ini melakukan pendekatan berupa optimasi terstruktur *Linear Regression* melalui *GridSearchCV* yang dikombinasikan dengan *preprocessing* berbasis stabilisasi distribusi data dan evaluasi *multimetric*. Dengan tujuan untuk meningkatkan performa model secara signifikan pada fitur

 ISSN:2302-8149
e-ISSN:2540-9719

produksi padi, dan model prediksi ini akan diimplementasikan menggunakan platform Streamlit Cloud agar dapat diakses secara online,

2 Tinjauan Literatur

Metode regresi linear untuk membantu UMKM Keripik Assri dalam manajemen stok dan penjualan. Menggunakan dataset historis dari 2019 hingga 2024, penelitian ini menggunakan pendekatan Knowledge Discovery in Database (KDD) dan membagi data menjadi latih dan uji. Hasil evaluasi menunjukkan nilai RMSE 0.000, menandakan akurasi prediksi yang sangat baik [12]. Dengan dataset 1000 pasien dari tahun 2022, hasil menunjukkan bahwa Linear Regression lebih akurat dibandingkan Decision Tree dengan MAPE berkisar antara 12,73% hingga 12,99% [7].

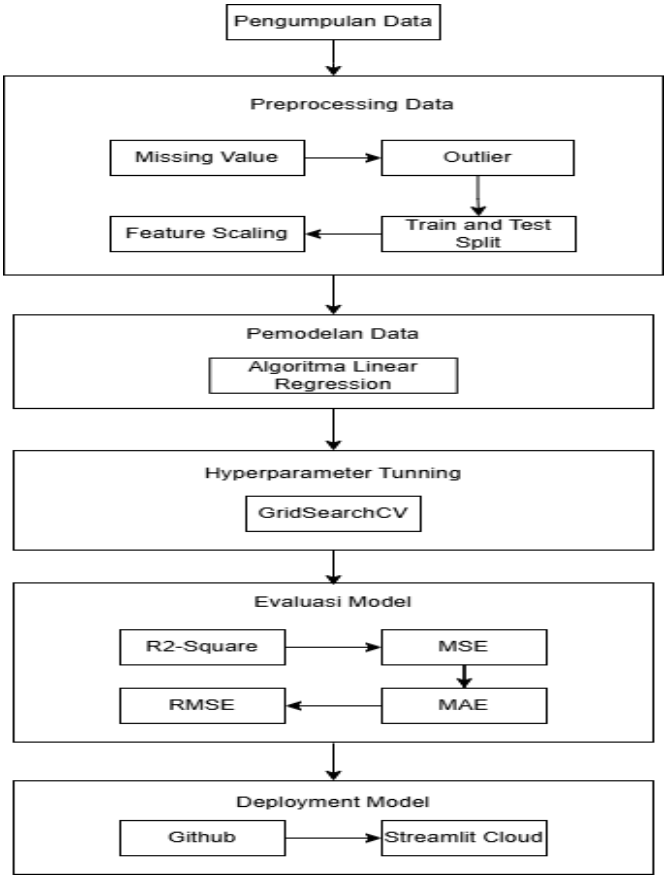
Linear Regression menunjukkan performa terbaik dengan R^2 98,33% dan MAE 27.764,86, menjadikannya metode paling akurat untuk prediksi produksi padi [6]. Menggunakan dataset tanaman padi 2011-2022 (190 data dari 19 kecamatan). Model diuji dengan MSE 6.542,46 dan MAE 117.823.406,92, menunjukkan akurasi yang cukup baik dalam prediksi produksi padi [11].Regresi linear memiliki akurasi tinggi dalam prediksi hasil panen [13].

Regresi linear memiliki performa lebih baik dengan RMSE 1,6 dan MAPE 1,1%, dibandingkan Random Forest yang memiliki RMSE 921,19 [14]. Linear Regression dengan hyperparameter tuning menghasilkan R^2 86,90%, lebih tinggi dibandingkan Decision Tree (81,42%), menjadikannya metode lebih efektif dalam prediksi produksi padi [15]. Kebaruan penelitian menerapan GridSearchCV sebagai mekanisme optimasi *Linear Regression* dalam konteks produksi padi Jawa Tengah, dan validasi model dengan skema evaluasi komprehensif yang tujuan untuk meningkatkan performa signifikan atas model baseline.serta membangun aplikasi prediksi berbasis web atau *Streamlit Cloud*,

3 Metode Penelitian

Penelitian ini menggunakan pendekatan regresi linear berganda untuk memodelkan hubungan antara luas tanam, luas panen, dan produktivitas terhadap produksi padi. Optimasi parameter menggunakan GridSearchCV untuk meningkatkan performa model dan implementasi *Streamlit Cloud*. berikut tahapan alur penelitian yang ditunjukkan pada Gambar 1.

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 1 Alur penelitian

3.1. Pengumpulan Data

Penelitian ini menggunakan dataset produksi padi sebagai objek utama, dengan variabel dependen berupa produksi padi dan variabel independen mencakup luas tanam, luas panen, serta produktivitas. Data diperoleh dari website resmi Dinas Pertanian dan Perkebunan, mencakup 350 data historis dari tahun 2014 hingga 2023 di 35 Kabupaten/Kota di Provinsi Jawa Tengah[1].

3.2. Preprocessing Data

Setelah data dikumpulkan, tahap preprocessing dilakukan untuk membersihkan dan mempersiapkan data agar siap digunakan dalam pembangunan model yang lebih akurat [16]. Proses ini mencakup identifikasi dan penanganan missing value dengan metode interpolasi atau penghapusan data yang tidak valid, serta pengecekan outlier menggunakan metode statistik seperti z-score atau interquartile range (IQR) untuk menghindari bias dalam pemodelan [17], [18]. Selain itu, dataset dibagi menjadi data latih dan data uji (train and test split) guna memastikan model dapat diuji secara independen sebelum diterapkan pada data baru [19]. Untuk meningkatkan efektivitas algoritma, dilakukan feature scaling dengan metode standardization agar skala data lebih seragam, sehingga model dapat memahami pola dengan lebih baik dan menghasilkan prediksi yang lebih akurat [20].

3.3. Pemodelan Data

Pemodelan data dalam penelitian ini menggunakan algoritma Linear Regression, khususnya Linear Regression, untuk memprediksi nilai variabel dependen berdasarkan hubungan dengan variabel independen [6]. Linear Regression digunakan karena terdapat lebih dari satu variabel independen yang memengaruhi variabel dependen. Dalam penelitian ini, variabel dependen yang diprediksi adalah produksi padi, sedangkan variabel independen meliputi luas tanam, luas panen, dan produktivitas [3], [21]. Adapun formula dari kedua variabel tersebut dapat dilihat dengan rumus (1) sebagai berikut :

$$Y = b + b_1X_1 + b_2X_2 + \ldots + b_2X_2 \qquad (1)$$

Dimana :
Y = Variabel dependen

 ISSN:2302-8149
e-ISSN:2540-9719

b = Konstanta
$b_1 b_2 b_3$ = Nilai koefisien regresi
$X_1 X_2 b_3$ =Variabel independent

3.4. Hyperparameter Tuning

Hyperparameter tuning adalah proses mencari kombinasi parameter optimal untuk meningkatkan performa model, dan dalam penelitian ini digunakan teknik GridSearchCV [15](Jamiluddin et al., 2024). Metode ini memungkinkan pencarian sistematis terhadap berbagai kombinasi parameter, seperti fit_intercept, copy_X, n_jobs, dan positive, yang diuji melalui validasi silang dengan cv=5. Model dilatih menggunakan setiap kombinasi parameter, kemudian dievaluasi berdasarkan score R^2. Hasil tuning menunjukkan bahwa model menjadi lebih stabil, dengan perbedaan minimal antara score pelatihan dan pengujian, serta prediksi yang lebih akurat. Dengan demikian, penerapan GridSearchCV membantu meningkatkan kestabilan dan kemampuan generalisasi model terhadap data baru [22].

3.5. Evaluasi Model

Beberapa metrik yang digunakan dalam evaluasi ini antara lain R-squared (R^2), yang mengukur proporsi variabilitas data yang dapat dijelaskan oleh model, di mana nilai mendekati 1 menunjukkan model yang baik [20]. Selain itu, digunakan Mean Absolute Error (MAE) untuk menghitung rata-rata perbedaan absolut antara nilai aktual dan prediksi [23]. Mean Squared Error (MSE) yang menghitung rata-rata kuadrat selisih antara keduanya, memberikan penalti lebih besar terhadap kesalahan yang lebih besar [24]. Root Mean Squared Error (RMSE), sebagai akar dari MSE, memberikan ukuran kesalahan dalam satuan yang sama dengan data asli sehingga lebih mudah diinterpretasikan. Evaluasi ini memastikan model memiliki performa yang baik, akurat, dan dapat diandalkan dalam pengambilan keputusan berbasis data [18].

3.6. Deployment Model

Tahap deployment model merupakan langkah akhir dalam penelitian ini, di mana model yang telah dikembangkan diimplementasikan dalam sistem berbasis web agar dapat diakses oleh pengguna. Model yang telah dilatih dan diuji disimpan dalam repositori GitHub untuk mempermudah pengelolaan dan kolaborasi. Selanjutnya, model diterapkan menggunakan *Streamlit Cloud* guna membangun antarmuka interaktif yang memungkinkan pengguna mengakses hasil prediksi secara daring [25]. Model regresi yang telah dibuat diintegrasikan ke dalam website dengan tampilan front-end yang menampilkan variabel-variabel dari dataset secara real-time. Dengan adanya deployment ini, pengguna dapat memasukkan data baru dan memperoleh hasil prediksi secara langsung [26].

4 Hasil dan Pembahasan

4.1 Pengumpulan Data

Pengumpulan data dilakukan melalui rekapan data dari sumber website Dinas Pertanian dan Perkebunan Provinsi Jawa Tengah untuk variabel luas tanam(Ha), luas panen(Ha), produktivitas(Ku/Ha), dan produksi(Ton). Dari proses pengumpulan data diperoleh jumlah 350 data yang dapat dilihat pada Gambar 2 berikut ini.

	Kabupaten	Tahun	Luas Tanam	Luas Panen	Produktivitas	Produksi
0	Cilacap	2014	135533	132074	52.84	697918
1	Cilacap	2015	140723	138864	63.99	888642
2	Cilacap	2016	165280	144212	62.15	888979
3	Cilacap	2017	147719	152967	58.53	895352
4	Cilacap	2018	145528	124011	63.97	793265
...	...	...	...	...	...	...
345	Kota Tegal	2019	344	619	58.06	3594
346	Kota Tegal	2020	485	585	35.63	2085
347	Kota Tegal	2021	263	505	48.40	2443
348	Kota Tegal	2022	339	626	51.36	3215
349	Kota Tegal	2023	172	545	58.37	3181

350 rows × 6 columns

 ISSN:2302-8149
e-ISSN:2540-9719

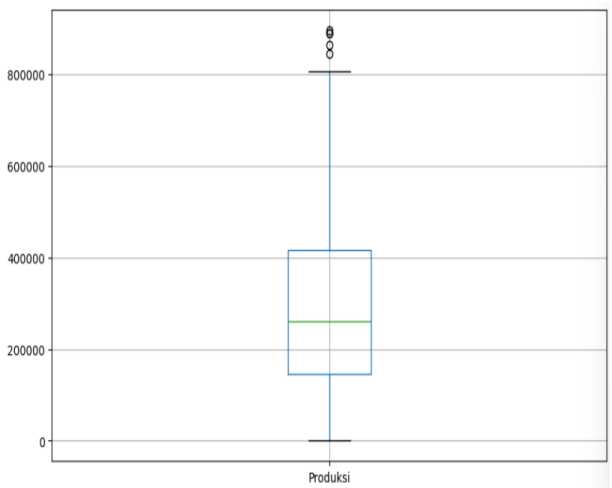
Gambar 2 Dataset yang digunakan

4.2 Preprocessing Data

Tahap berikutnya yaitu menyeleksi ulang data untuk melakukan pembersihan data. Memeriksa apakah pada dataset yang digunakan terdapat *missing value* atau tidak. Berikut merupakan pengecekan *missing value*, dapat dilihat pada Gambar 3 berikut ini.

Gambar 3 Cek *missing value*

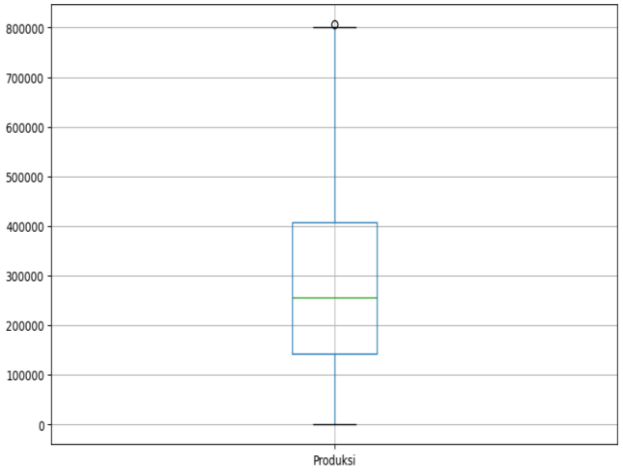
Kemudian mengecek o*utlier* (pencilan) langkah pertama yaitu perlu memastikan apakah terdapat outlier pada dataset yang akan digunakan. Diagram kotak *(box plot)* terdapat outlier pada data produksi padi, ditunjukkan oleh titik-titik di luar batas normal diagram. Berikut ini pengecekan *outlier* yang dapat dilihat pada Gambar 4 di bawah ini.



Gambar 4 Pengecekan *outlier*

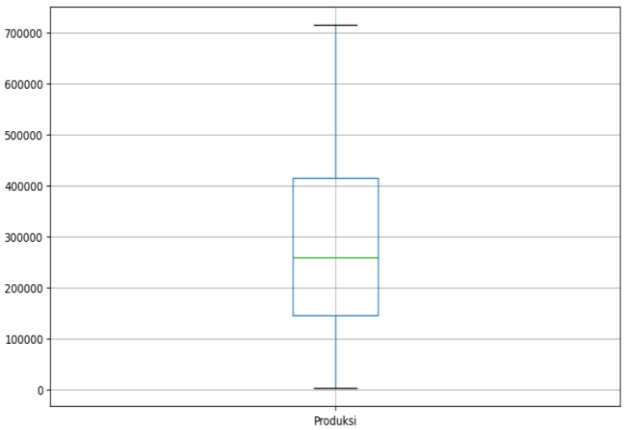
Kemudian mencari dan menampilkan data-data yang dianggap sebagai outlier menggunakan metode *Interquartile Range* (IQR). Setelah melakukan pencarian data yang outlier, bahwasanya terdapat 5 baris data outlier yaitu pada kabupaten Cilacap terdapat 3 data *outlier* dan Grobogan terdapat 2 data *outlier*. Langlah selanjutnya yaitu menampilkan *boxplot* 'Produksi' yang baru dapat dilihat pada Gambar 5 berikut ini.

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 5 *Boxplot outlier*

Untuk menangani outlier pada kolom 'Produksi' dalam data dengan teknik *Winsorization*. Dengan demikian, dampak *outlier* pada analisis statistik dapat diminimalkan tanpa menghilangkan data tersebut sepenuhnya. Berikut hasil *outlier* yang yang dapat dilihat pada Gambar 6 di bawah ini.



Gambar 6 *Boxplot tanpa outlier*

Tahapan berikutnya *train and test split*, Dalam pembagian ini, 80% data digunakan untuk *data training*, sedangkan 20% digunakan *data testing* dengan pembagian secara acak, yang dapat dilihat pada Gambar 7 berikut ini.

```
x_train : (276, 3)
x_test : (69, 3)
y_train : (276, 1)
y_test : (69, 1)
```

Gambar 7 *Train and test split*

Tahapan berikutnya *feature scaling*, Menyeragamkan skala data numerik menggunakan StandardScaler agar tidak ada fitur yang mendominasi proses pembelajaran model. Adapun proses *scaling data train* dan *scaling data test* tersebut dapat dilihat pada Gambar 8 dan 9.

 ISSN:2302-8149
e-ISSN:2540-9719

```
[[ 0.54766591 -0.01908766 -0.17385659]
 [ 1.54627101  1.69117177  0.52144924]
 [-0.34556167 -0.2918815   0.14551834]
 [ 0.591996    0.92397671 -0.76104071]
 [-1.43243879 -1.40198299 -0.08403239]]
[[-0.0549728 ]
 [ 1.76943968]
 [-0.27241016]
 [ 0.68473907]
 [-1.34862432]]
```

Gambar 8 *Scaling data train*

```
[[-1.46792504 -1.44448068 -0.70282132]
 [-0.16657792 -0.54853082 -0.77601141]
 [-0.76488155 -0.58494922  0.17213291]
 [-1.37139136 -1.36134924 -0.96730368]
 [-0.31476072 -0.14557706 -1.41808809]]
[[-1.38903533]
 [-0.60604937]
 [-0.55386793]
 [-1.31811907]
 [-0.33964602]]
```

Gambar 9 *Scaling data test*

4.3 Pemodelan Data

Langkah pertama adalah mengimpor modul LinearRegression dari library sklearn. Selanjutnya, buat sebuah variabel untuk menyimpan model yang telah dibuat agar lebih mudah dipanggil pada tahap berikutnya. Setelah model selesai dibuat, proses pelatihan dilakukan menggunakan data latih (train) yang telah dipersiapkan pada tahap pra-pemrosesan, di mana x_train berisi fitur dan y_train berisi label pada Gambar 10 berikut ini.

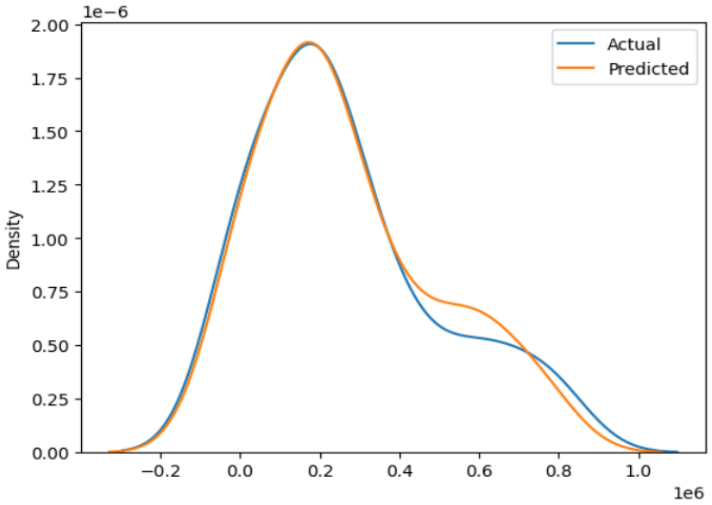
```
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
lr.fit(x_train, y_train)
```

```
▾ LinearRegression
LinearRegression()
```

Gambar 10 Pembuatan algoritma *linear regression*

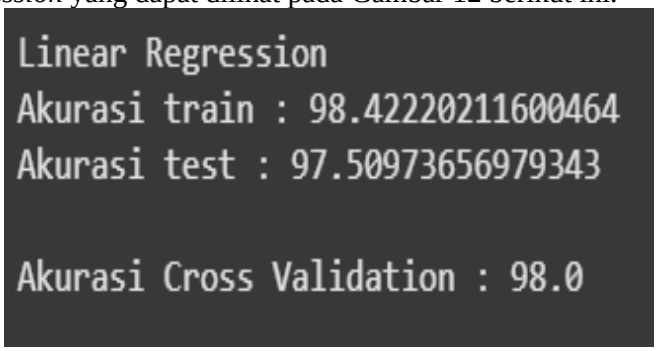
Setelah model selesai dilatih, langkah selanjutnya adalah menguji model dengan melihat nilai akurasi dan prediksi melalui grafik plot. Tujuannya adalah untuk memberikan gambaran mengenai seberapa akurat model yang telah dibangun dalam melakukan prediksi. Berikut dapat dilihat nilai aktual dan prediksi pada Gambar 11 di bawah ini

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 11 Grafik *linear regression*

Pada Gambar 11, menunjukkan bahwa algoritma *Linear Regression* cukup baik dalam memprediksi data prediksi, meskipun terdapat sedikit perbedaan terutama setelah puncak utama. Ini bisa jadi diindikasi bahwa model tersebut cukup akurat, tetapi mungkin perlu beberapa penyesuaian untuk meningkatkan prediksi. Selanjutnya adalah mengecek score *train*, *test* dan *cross validation* pada algoritma *Linear Regression* yang dapat dilihat pada Gambar 12 berikut ini.



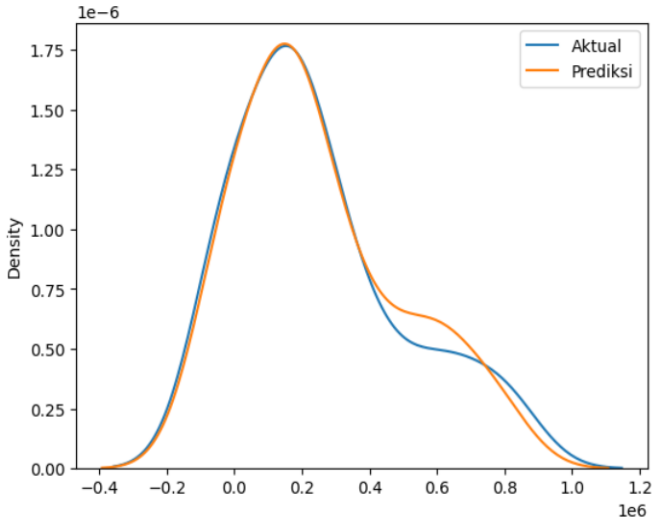
Gambar 12 Akurasi *cross validation*

Hasil akurasi pada Gambar 12 diatas, data pelatihan mencapai 98,42%, akurasi pada data pengujian sebesar 97.50%, dan akurasi rata-rata berdasarkan *cross-validation* adalah 98,0%. Kesimpulannya adalah algoritma *Linear Regression* memiliki performa yang sangat baik dan dapat diandalkan untuk memprediksi produksi padi di Provinsi Jawa Tengah.

4.4 *Hyperparameter tuning*

Teknik yang akan digunakan yaitu *GridSearchCV*. Sebelum menggunakan kedua teknik tersebut untuk menemukan *hyperparameter* terbaik, perlu mengetahui nilai default atau nilai terkini dari *hyperparameter* untuk menentukan ruang pencarian dengan *method lr.get_params()* untuk mengambil nilai terkini dari semua *hyperparameter* yang telah diatur untuk estimator yang dapat dilihat pada Gambar 13 berikut ini.

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 13 Grafik dengan *gridsearchcv*

Pada gambar 13, hasil prediksi setelah optimalisasi dengan *GridSearchCV* lebih mendekati data aktual, menunjukkan peningkatan performa model dalam menangkap pola data. Setelah melakukan *hyperparameter tuning*, langkah selanjutnya adalah mengevaluasi kembali akurasi model pada data pelatihan dan pengujian.

```
Train Score setelah tuning: 98.42159105558065
Test Score setelah tuning: 97.542515940772
Rata-rata skor cross-validation: 98.26589357298114
```

Gambar 14 *Cross validation* dengan *gridsearchcv*

Pada Gambar 14, setelah optimalisasi dengan *GridSearchCV*, model Linear Regression menunjukkan peningkatan performa, dengan akurasi prediksi 98.42, cross-validation score 97.54 yang menandakan kestabilan, serta score pelatihan 98.26 tanpa overfitting signifikan. Untuk perbedaan hasil algoritma sebelum dan sesudah dilakukan *hyperparameter tuning* yang dapat dilihat pada Tabel 1 berikut ini:

Tabel 1. Perbandingan Hasil *Model*

Metode	Score Train	Score Test	Rata-rata Cross Validation Score
Linear Regression	98.42220211600464	97.50973656979343	98.0
Linear Regression GridSearch CV	98.42159105558065	97.542515940772	98.26

Dari Tabel 1, Hasil perbandingan antara model *Linear Regression* sebelum dan sesudah dilakukan *hyperparameter tuning* menggunakan *GridSearchCV* menunjukkan adanya peningkatan performa yang meskipun tidak besar, namun signifikan dari sisi stabilitas dan kemampuan generalisasi model. Nilai *score train* pada kedua model hampir identik, yaitu 98,42% untuk model baseline dan 98,42% untuk model hasil *tuning*, yang mengindikasikan bahwa proses optimasi tidak menimbulkan gejala *overfitting* maupun perubahan drastis pada perilaku model terhadap data latih. Perbedaan mulai terlihat pada *score test*, di mana model tanpa *tuning* memperoleh nilai 97,50%, sedangkan model yang telah dioptimasi meningkat menjadi 97,54%. Kenaikan ini menunjukkan bahwa *GridSearchCV* membantu model bekerja lebih baik pada data yang belum pernah dilihat sebelumnya. Peningkatan yang paling mencolok terdapat pada nilai *cross-validation score*, yaitu dari 98,00% pada model *baseline* menjadi 98,26% setelah *tuning*. Hal ini menegaskan bahwa model hasil optimasi memiliki konsistensi performa yang lebih baik di berbagai subset data, sehingga lebih *robust* dan memiliki kemampuan generalisasi yang lebih kuat.

 ISSN:2302-8149
e-ISSN:2540-9719

4.5 Evaluasi Model

Evaluasi yang digunakan yaitu R2-Square, *Mean Absolute Error (MAE*), *Mean Squared Error (MSE)*, dan *Root Mean Squared Error (RMSE)* seperti pada Tabel 2 di bawah ini:

Tabel 2 Hasil evaluasi model

Teknik Evaluasi	Hasil
R2-Square	0.9754251594077201
MAE	0.09569555945881004
MSE	0.03073218293780972
RMSE	0.1753059694870934
R2-Square	0.9754251594077201

Dari tabel 2 di atas, tahap evaluasi model menggunakan R^2, *MAE*, *MSE*, dan *RMSE* menunjukkan bahwa Linear Regression memiliki performa yang sangat baik. Dengan R^2 sebesar 0.9754, model mampu menjelaskan 97,54% variasi data. MAE sebesar 0.0957 menunjukkan rata-rata kesalahan prediksi yang kecil, sedangkan MSE sebesar 0.0307 dan RMSE sebesar 0.1753 mengindikasikan bahwa model memiliki tingkat akurasi tinggi dan kesalahan prediksi rendah. Oleh karena itu, model ini dapat diandalkan untuk prediksi pada dataset yang digunakan. Selanjutnya adalah memprediksi atau meramalkan hasil produksi padi menggunakan kode program dengan menggunakan data baru seperti gambar 15 berikut ini.

```
# Create DataFrame for new data with only the 3 features
data_baru = pd.DataFrame([[76235, 3572, 51.06]], columns=['Luas Tanam', 'Luas Panen', 'Produktivitas'])

# Make prediction
prediction = lr_model.predict(data_baru)
print(prediction)
```

Gambar 15 Kode program proses prediksi baru

Pada Gambar 15, prediksi dilakukan dengan model Linear Regression yang telah dilatih. Data baru disiapkan dalam DataFrame dengan tiga fitur utama: luas tanam (87.235), luas panen (65.572), dan produktivitas (5.106). Model lr_model kemudian memprediksi nilai target menggunakan perintah *lr_model.predict(data_baru)*, dan pada Gambar 16 menunjukkan bahwa hasil prediksi poduksi padi berdasarkan analisis data baru yang telah dilakukan menghasilkan nilai sebesar 381970.59 Ton.

```
[45211.23789415]
```

Gambar 16 Prediksi menggunakan data baru

4.6 *Deployment Model*

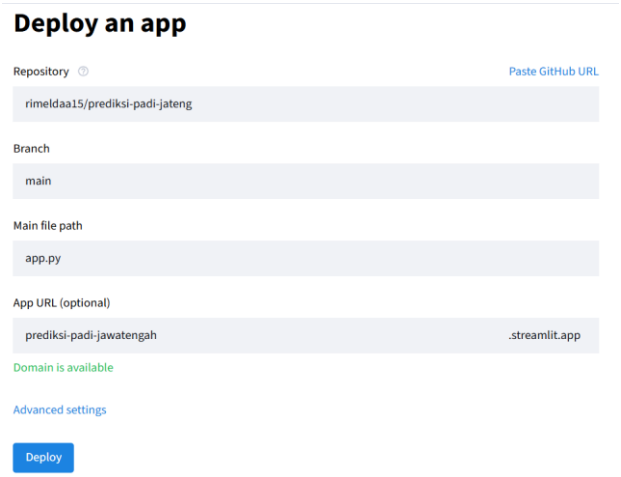
Tahap deployment model dilakukan dengan membangun website interaktif berbasis Streamlit Cloud agar dapat diakses secara daring. Proses ini meliputi penyimpanan model yang telah dilatih dalam file models prediksi.pkl, pembuatan repository GitHub untuk menyimpan file, serta penulisan kode tampilan antarmuka di Streamlit pada Gambar 17 berikut ini.

 ISSN:2302-8149
e-ISSN:2540-9719

```python
import streamlit as st
import pickle
import numpy as np

# Load the model
with open('model_prediksi.pkl', 'rb') as file:
    model = pickle.load(file)

st.title("Prediksi Hasil Produksi Padi Provinsi Jawa Tengah")  # Mengganti judul

# Informasi Akurasi Model
st.markdown("**Terdiri dari 3 Variabel yaitu Luas Tanam (Hektare), Luas Panen (Hektare), dan Produktivitas (Kuin

# Luas Tanam dengan batasan minimal dan maksimal, dan satuan hektar
Luas_Tanam = st.number_input("Luas Tanam (Ha) (min: 0, maks: 1000000) ", min_value=0.0, max_value=1000000.0, for
# Luas Panen dengan satuan hekta
Luas_Panen = st.number_input("Luas Panen (Ha) (min: 0, maks: 1000000) ", min_value=0.0, max_value=1000000.0, for
# Produktivitas dengan satuan Ku/Ha
Produktivitas = st.number_input("Produktivitas (Ku/Ha) (min: 0, maks: 10000) ", min_value=0.0, max_value=10000.0
```

Gambar 17 Kode program tampilan antarmuka di *streamlit*

Selanjutnya, semua file app.py, model_prediksi.pkl, dan *requirements*, diunggah ke GitHub dan dikoneksikan ke Streamlit Cloud untuk menjalankan aplikasi secara online. Kemudian, repository dikoneksikan ke Streamlit Cloud untuk menjalankan aplikasi sebagai website interaktif. Proses ini mencakup autentikasi dengan akun GitHub, memilih repository, menentukan file utama (app.py), dan memulai deployment agar aplikasi dapat diakses secara daring yang dapat dilihat pada Gambar 18 berikut ini.



Gambar 18 Pengkoneksian *Streamlit*

Tahap akhir dalam proses deployment adalah menampilkan hasilnya dalam bentuk website yang berisi fitur prediksi berdasarkan variabel masukan, seperti luas tanam (*hectare area planted*), luas panen (*hectare area harvested*), dan produktivitas (*productivity rate*). Melalui website ini, pengguna dapat memasukkan nilai dari masing-masing variabel tersebut, kemudian model akan memproses data yang diberikan dan menghasilkan prediksi berdasarkan pola yang telah dipelajari sebelumnya. Berikut tampilan akhir dari website hasil deployment dapat dilihat pada Gambar 19 berikut ini.

Gambar 19 Tampilan *deployment*

 ISSN:2302-8149
e-ISSN:2540-9719

Berdasarkan Gambar 20, dengan memasukkan data yang sama dapat disimpulkan bahwa hasil prediksi produksi padi berdasarkan analisis data yang telah dilakukan adalah 45211.24 Ton. Hasil prediksi yang diperoleh melalui website sama dengan hasil yang diperoleh menggunakan kode program pada Gambar sebelumnya. Sistem prediksi yang peneliti kembangkan adalah untuk mengetahui prediksi produksi padi setiap kabupaten di Provinsi Jawa Tengah.

Gambar 20 Implementasi prediksi

5 Kesimpulan

Penelitian ini berhasil mengembangkan model prediksi produksi padi di Provinsi Jawa Tengah melalui pendekatan metodologis yang mengintegrasikan *Linear Regression* dengan optimasi *hyperparameter GridSearchCV*. Pipeline *preprocessing* yang terstruktur, mencakup penanganan *missing value, stabilisasi outlier*, standarisasi fitur, serta pembagian data latih–uji, terbukti meningkatkan stabilitas model dan mendukung proses optimasi secara efektif. Hasil penelitian menunjukkan bahwa model optimal memiliki performa yang sangat baik dengan nilai R^2 sebesar 0.9754, MAE 0.0957, MSE 0.0307, dan RMSE 0.1753, yang merepresentasikan peningkatan akurasi dibandingkan model baseline sebelum tuning. Penelitian selanjutnya dapat memperluas variabel prediktor dengan memasukkan faktor iklim seperti curah hujan, suhu, dan jenis varietas padi serta melakukan perbandingan performa dengan model non-linear seperti Random Forest, Support Vector Regression, atau LSTM untuk meningkatkan ketepatan prediksi pada pola data yang lebih kompleks.

Referensi

[1] Subagprogram, “Data Series Tanaman Pangan Jawa Tengah 1998-2022,” Data Jateng Prov. Accessed: Jan. 03, 2025. [Online]. Available: https://data.jatengprov.go.id/dataset/data-series-tanaman-pangan-jawa-tengah

[2] BPS, “Luas Panen dan Produksi Padi di Provinsi Jawa Tengah Tahun 2023 (Angka Tetap), ,” Badan Pusat Stattistik Provinsi Jawa Tengah. Accessed: Jan. 03, 2025. [Online]. Available: https://jateng.bps.go.id/id/pressrelease/2024/03/01/1528/luas-panen-dan-produksi-padi-di-provinsi-jawa-tengah-tahun-2023--angka-tetap-.html

[3] BPS, “Luas Panen, Produksi, dan Produktivitas Padi Menurut Kabupaten/Kota di Provinsi Jawa Tengah (Kuintal/Hektar), 2023,” Badan Pusat Stattistik Provinsi Jawa Tengah. Accessed: Jan. 03, 2025. [Online]. Available: https://jateng.bps.go.id/id/statistics-table/2/NDYzIzI=/luas-panen--produksi--dan-produktivitas-padi-menurut-kabupaten-kota-di-provinsi-jawa-tengah.html

[4] E. Triyanto, H. Sismoro, and A. D. Laksito, “Implementasi Algoritma Regresi Linear Berganda untuk memprediksi Produksi Padi di Kabupaten Bantul,” *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, Vol. 4, No. 2, pp. 66–75, 2019, DOI: 10.36341/rabit.v4i2.666.

[5] R. Andia, K. Kaslani, S. Eka Permana, and T. Handayani, “Peramalan Hasil Panen Padi Kabupaten Cirebon menggunakan Algoritma Regresi Linear Berganda,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, Vol. 8, No. 1, pp. 738–747, 2024, DOI: 10.36040/jati.v8i1.8446.

 ISSN:2302-8149
e-ISSN:2540-9719

[6] J. Hutahaean, D. Yusup, and Purwantoro, "Perbandingan Metode *Linear Regression, Random Forest & K-Nearest Neighbor* untuk Prediksi Produksi Hasil Panen Padi di Provinsi Jawa Barat," *Jurnal Mahasiswa Teknik Informatika (JATI)*, Vol. 8, No. 3, Jun. 2024, DOI: https://doi.org/10.36040/jati.v8i3.9821.

[7] A. T. Nurani, A. Setiawan, and B. Susanto, "Perbandingan Kinerja Regresi *Decision Tree* dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma," *Jurnal Sains dan Edukasi Sains*, vol. 6, no. 1, pp. 34–43, 2023, doi: 10.24246/juses.v6i1p34-43.

[8] R. Norhikmah, "Prediksi Kegagalan Siswa dalam Data Mining dengan menggunakan Metode *Naïve Bayes*," *Terakreditasi DIKTI*, Vol. 3, No. 1, pp. 42–46, 2019, DOI: 10.13140/RG.2.2.22726.42560.

[9] F. H. Hamdanah and D. Fitrianah, "Analisis Performansi *Algoritma Linear Regression* dengan *Generalized Linear Model* untuk Pradiksi Penjualan pada Usaha Mikro, Kecil dan Menengah," *Jurnal Nasional Pendidikan Teknik Informatika : JANAPATI*, Vol. 10, No. 1, pp. 23–32, 2021.

[10] H. W. Herwanto, T. Widiyaningtyas, and P. Indriana, "Penerapan *Algoritme Linear Regression* untuk Prediksi Hasil Panen Tanaman Padi," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, Vol. 8, No. 4, p. 364, 2019, DOI: 10.22146/jnteti.v8i4.537.

[11] M. Adin Musababa, "Implementasi *Algoritma Linear Regression* untuk Prediksi Produksi Tanaman Padi di Kabupaten Grobogan," *Data Sciences Indonesia (DSI)*, Vol. 3, No. 2, pp. 68–78, 2024, DOI: 10.47709/dsi.v3i2.3118.

[12] A. H. Pradhana, M. Irfa, A. Ali, A. Ristyawan, and E. Daniati, "Penerapan Regresi Linear menggunakan *RapidMiner* untuk memprediksi Penjualan dan Persediaan," *Semnas Inotek*, Vol. 8, pp. 291–297, 2024.

[13] Diyanti, Martanto, and A. Bahtiar, "Jurnal Informatika Terpadu Prediksi Hasil Panen Padi Tahun 2023 menggunakan Metode Regresi Linear di Kabupaten Indramayu," *Jurnal Informatika Terpadu*, Vol. 9, No. 1, pp. 18–23, 2023, [Online]. Available: https://journal.nurulfikri.ac.id/index.php/JIT

[14] D. Pramesti and Wiga Maulana Baihaqi, "Perbandingan Prediksi Jumlah Transaksi Ojek Online menggunakan Regresi Linier dan *Random Forest*," *Generation Journal*, Vol. 7, No. 3, pp. 21–30, 2023, DOI: 10.29407/gj.v7i3.20676.

[15] F. Jamiluddin, S. Faisal, S. A. P. Lestari, and A. Fauzi, "Implementasi *Hyperparameter Tuning GridSearchCV* pada Prediksi Produksi Padi menggunakan Algoritma Linear Regresi," *Journal of Information System Research (JOSH)*, Vol. 6, No. 1, pp. 480–488, 2024, DOI: 10.47065/josh.v6i1.5930.

[16] W. Nugraha and A. Sasongko, "*Hyperparameter Tuning on Classification Algorithm with Grid Search*," *Sistemasi*, Vol. 11, No. 2, p. 391, 2022, DOI: 10.32520/stmsi.v11i2.1750.

[17] I. Hakim and T. Afriliansyah, "Implementasi Algoritma Komputasi Linear Regression untuk Optimasi Prediksi Hasil Pertanian," *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, Vol. 5, No. 3, pp. 1423–1434, 2024.

[18] E. P. Yudha, Arif Rohmadi, and Agung Teguh Setyadi, "Sistem Prediksi Produksi Padi di Sumatera menggunakan Regresi Linear," *Jurnal Manajemen Informatika dan Sistem Informasi*, Vol. 8, No. 1, pp. 81–89, 2025, DOI: 10.36595/misi.v8i1.1411.

[19] E. M. Pusung and I. N. Dewi, "Optimasi RoBERTa dengan *Hyperparameter Tuning* untuk Deteksi Emosi berbasis Teks," *Jurnal Nasional Teknologi dan Sistem Informasi*, Vol. 10, No. 3, pp. 240–248, 2025, DOI: 10.25077/teknosi.v10i3.2024.240-248.

[20] K. Sofi, A. S. Sunge, S. R. Riady, and A. Z. Kamalia, "Perbandingan Algoritma Linear *Regression*, LSTM, dan Gru dalam memprediksi Harga Saham dengan Model *Time Series*," *Seminastika*, Vol. 3, No. 1, pp. 39–46, 2021, DOI: 10.47002/seminastika.v3i1.275.

[21] M. K. B. Seran, F. Tedy, Ign. P. A. N. Samane, P. Batarius, P. A. Nani, and A. A. J. Sinlae, "Analisis Data Pertanian Tanaman Pangan untuk memprediksi Hasil Panen di Kabupaten Malaka menggunakan Metode *Multiple Linear Regression*," *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, Vol. 4, No. 1, pp. 209–221, 2024, DOI: 10.24002/konstelasi.v4i1.8970.

[22] I. Muhamad Malik Matin, "*Hyperparameter Tuning* menggunakan *GridsearchCV* pada *Random Forest* untuk Deteksi Malware," *Multinetics*, Vol. 9, No. 1, pp. 43–50, 2023, DOI: 10.32722/multinetics.v9i1.5578.

 ISSN:2302-8149
e-ISSN:2540-9719

[23] U. Lathifah and R. Danar Dana, "Implementasi Metode *Linear Regression* untuk Prediksi Harga Properti *Real Estate* menggunakan *Rapidminer*," *JATI (Jurnal Mahasiswa Teknik Informatika)*, Vol. 8, No. 1, pp. 1129–1137, 2024, DOI: 10.36040/jati.v8i1.8919.

[24] H. H. Nuha, "*Mean Squared Error (MSE)* dan Penggunaannya," *Social Science Research Network*, Vol. 52, pp. 2021–2022, 2023.

[25] M. A. Dewi, P. T. L., Dewanta, F., Nugroho, "Implementasi *Machine Learning Model Deployment* pada *Website* Pemantauan Kondisi Sungai Citarum menggunakan *Platform-As-A-Service," e-Proceeding of Engineering*, Vol. 8, No. 6, pp. 3064–3074, 2022.

[26] G. A. Syafarina and Zaenuddin, "Implementasi *Framework Streamlit* sebagai Prediksi Harga Jual Rumah dengan Linear Regresi," *Metik Jurnal*, Vol. 7, pp. 121–125, 2023, DOI: 10.47002/metik.v7i2.680.