JRIIN: Jurnal Riset Informatika dan Inovasi
Volume 2, No. 10 Maret 2025
ISSN 3025-0919 (media online)
Hal 1927-1934

Implementasi Model *Machine Learning* Untuk Klasifikasi Dan Prediksi Pada Dataset Bunga Iris

Abdurahman[1], Alda Ailsa Alvita[2], Indika Saputra[3], Kamil Haedar[4], Muhammad Ario Ardhi[5], Muhammad Edria Zaiz Faidullah[6], Muhammad Fahmi Alfarizi[7]

[1]Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia.
Email: [1]rahmanabdu702@gmail.com , [2]aldaailsaalvita@gmail.com , [3]indikasaputra29@gmail.com , [4]kamilhaedar01@gmail.com , [5]ario.ardhi2eeeee@gmail.com , [6]zoyedria.231@gmail.com , [7]fahmimal99@gmail.com

Abstrak− Penelitian ini bertujuan untuk mengimplementasikan berbagai model machine learning dalam menganalisis *dataset* Iris, termasuk *Logistic Regression*, *Random Forest*, dan *Linear Regression*. *Dataset* Iris digunakan karena sifatnya yang sederhana namun mencakup masalah klasifikasi dan regresi. Proses melibatkan *preprocessing* data, seperti *scaling* fitur, serta pembagian dataset menjadi data latih dan data uji. Evaluasi dilakukan menggunakan metrik akurasi, *confusion matrix*, *mean squared error* (MSE), dan *R-squared* (R^2). Hasil penelitian menunjukkan bahwa semua model mencapai performa sempurna pada tugas klasifikasi, dengan akurasi 100% dan *confusion matrix* yang hanya menunjukkan prediksi benar. Model *Linear Regression* juga menunjukkan kinerja yang sangat baik dengan nilai $R^2 = 1.00$, mengindikasikan kemampuan untuk menjelaskan seluruh variasi data. Korelasi antar variabel mengungkapkan bahwa fitur petal memiliki pengaruh lebih signifikan dibandingkan fitur sepal dalam membedakan spesies. Visualisasi hasil prediksi menunjukkan pola distribusi nilai aktual dan prediksi yang hampir identik.

Kata Kunci: *Machine Learning; Classification; Linear Regression; Random Forest; Prediction*

Abstract− *This study aims to implement various machine learning models in analyzing Iris datasets, including Logistic Regression, Random Forest, and Linear Regression. The Iris dataset is used because of its simple nature but covers classification and regression issues. The process involves preprocessing data, such as scaling features, as well as dividing the dataset into training data and test data. The evaluation was carried out using accuracy metrics, confusion matrix, mean squared error (MSE), and R-squared (R^2). The results showed that all models achieved perfect performance on classification tasks, with 100% accuracy and a confusion matrix that only showed correct predictions. The Linear Regression model also showed excellent performance with a value of $R^2 = 1.00$, indicating the ability to explain the entire variation in data. The correlation between variables revealed that the petal feature had a more significant influence than the sepal feature in distinguishing species. The visualization of prediction results shows the distribution patterns of actual and predicted values that are almost identical.*

Keywords: *Machine Learning; Classification; Linear Regression; Random Forest; Prediction*

1. PENDAHULUAN

Bunga Iris termasuk salah satu dataset populer yang sering digunakan sebagai pembelajaran dasar dari konsep *machine learning* karena dataset ini lengkap dan mudah digunakan dalam pemrograman python. Bunga Iris terdiri dari tiga spesies utama, yaitu Iris setosa, Iris versicolor, dan Iris virginica. Dari ketiga spesies tersebut terdapat 150 bunga Iris yang diidentifikasi berdasarkan panjang mahkota, lebar mahkota, panjang kelopak dan lebar kelopak (Rahman dkk., 2023). Dataset ini sangat cocok digunakan sebagai pembelajaran klasifikasi dan prediksi dalam *machine learning* (Setia Budi dkk., 2024).

Machine learning berkaitan dengan kemampuan mesin untuk belajar agar dapat menyerupai perilaku manusia dalam pengambilan keputusan dan menjalankannya secara otomatis (Fahmi, 2023). Selain itu teknik klasifikasi digunakan untuk membagi objek dan mengelompokkan objek tersebut ke dalam suatu kelas berdasarkan pola yang dipelajari, sedangkan prediksi digunakan untuk memperkirakan hasil atau nilai berdasarkan data yang telah di analisis (Arifin & Rofianto, 2023).

JRIIN: Jurnal Riset Informatika dan Inovasi
Volume 2, No. 10 Maret 2025
ISSN 3025-0919 (media online)
Hal 1927-1934

Beberapa algoritma yang sering digunakan untuk klasifikasi dan prediksi adalah *Logistic Regression*, *Linear Regression*, dan *Random Forest* (Fahmi, 2023). Hubungan antara satu atau lebih variabel dikatakan *Regresi Logistic* bila variabel independen dan dependen bersifat biner atau memiliki kategori nominal (Simanjuntak dkk., 2023). *Regresi Linear* digunakan untuk memahami dan memprediksi hubungan antar variabel (Susetyoko dkk., 2022). Sementara itu, metode *Random Forest* membantu untuk melihat seberapa pengaruhnya karakteristik bunga iris seperti panjang dan lebar mahkota, serta panjang dan lebar kelopak bunga dalam mengidentifikasi jenis bunga iris (Rahman dkk., 2023). Ketiga algoritma ini sering diterapkan untuk menyelesaikan berbagai masalah machine learning karena keandalannya dalam menangani data dengan pola yang kompleks dan beragam.

Salah satu hal yang sering terjadi ketika mengimplementasikan sebuah dataset kedalam model *machine learning* adalah dataset yang tidak lengkap atau seimbang seperti isi sel data yang kosong, oleh karena itu diperlukan sebuah *library scikit-learn* yaitu *SimpleImputer* untuk menangani data yang hilang (Simanjuntak dkk., 2023). *SimpleImputer* mengisi nilai yang hilang dengan memprediksi nilai imputasi berdasarkan nilai yang sering muncul ataupun rata-rata dan median dalam dataset pelatihan (Fadlil dkk., 2023).

Berbagai analisis model *machine learning* terutama dalam klasifikasi dan prediksi pernah dilakukan. Beberapa penelitian ini menggunakan metode *Random Forest* dan *Logistic Regression* untuk klasifikasi emosi pada twitter. Algoritma *Random Forest* menghasilkan akurasi sebesar 73.72% dan *Logistic Regression* mampu menghasilkan akurasi sebesar 79.06% (Simanjuntak dkk., 2023). Berdasarkan penelitian tersebut perlu dikembangkan lebih lanjut dengan menambahkan algoritma *Linear Regression* untuk mengeksplorasi kemampuan prediksi kepada dataset Iris. Selain itu, kombinasi penggunaan algoritma *Logistic Regression*, *Random Forest, dan Linear Regression* dalam penelitian ini dapat memberikan pemahaman yang lebih mendalam mengenai performa masing-masing algoritma.

2. METODE

Penelitian ini menggunakan dataset Iris, yang merupakan dataset klasik dalam machine learning. Dataset ini terdiri dari tiga kelas spesies bunga, yaitu Iris-setosa, Iris-versicolor, dan Iris-virginica, dengan total 150 sampel (Rahman dkk., 2023). Setiap sampel memiliki empat fitur utama, yaitu panjang dan lebar kelopak (sepal) serta panjang dan lebar mahkota bunga (petal) (Rahman dkk., 2023). Sebelum diimplementasikan ke dalam model, data ini diproses melalui tahap preprocessing, termasuk scaling menggunakan metode standardisasi. Proses *scaling* ini dilakukan untuk memastikan bahwa semua fitur berada dalam rentang nilai yang seragam, sehingga dapat meningkatkan kinerja algoritma tertentu yang sensitif terhadap perbedaan skala antar variabel.

Tahapan implementasi model melibatkan tiga algoritma utama. Pertama, *Logistic Regression* digunakan untuk tugas klasifikasi tiga kelas spesies bunga. *Logistic Regression* dipilih karena kemampuannya dalam menghasilkan probabilitas prediksi yang baik pada masalah klasifikasi (Fadlil dkk., 2023). Kedua, *Random Forest* digunakan untuk mengevaluasi distribusi prediksi melalui *confusion matrix*. Algoritma ini dipilih karena sifatnya yang kuat dalam menangani data yang kompleks serta kemampuan memberikan interpretasi melalui pentingnya fitur (Susetyoko dkk., 2022). Ketiga, *Linear Regression* diterapkan untuk memprediksi nilai kontinu, seperti panjang mahkota bunga (petal length) berdasarkan fitur lainnya. Algoritma ini dipilih karena kesederhanaannya dan kemampuannya untuk menangkap hubungan linear antar variabel.

Untuk mengevaluasi performa model, berbagai metrik digunakan. Akurasi menjadi indikator utama dalam menilai keberhasilan model klasifikasi, sementara *confusion matrix* memberikan gambaran distribusi kesalahan dan prediksi yang benar. Untuk model regresi, metrik *mean squared error* (MSE) digunakan untuk mengukur rata-rata kesalahan prediksi, sedangkan *R-squared* (R^2) digunakan untuk mengukur proporsi variabilitas data yang dapat dijelaskan oleh model (Fadlil dkk., 2023). Selain itu, visualisasi hasil evaluasi, seperti plot distribusi dan *heatmap* korelasi,

JRIIN: Jurnal Riset Informatika dan Inovasi
Volume 2, No. 10 Maret 2025
ISSN 3025-0919 (media online)
Hal 1927-1934

dilakukan untuk memberikan pemahaman yang lebih mendalam mengenai hubungan antar variabel dan performa model.

3. HASIL DAN PEMBAHASAN

Berdasarkan program yang sudah dibuat menggunakan bahasa python mengenai klasifikasi dan prediksi sesuai dengan model algoritma *Logistic Regression*, *Linear Regression*, dan *Random Forest*, maka diperoleh ouput yang menunjukkan data tersebut.

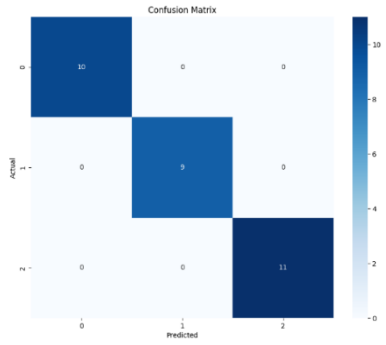
3.1 *Confusion Matrix (Logistic Regression)*

```
Akurasi model: 1.00

Confusion Matrix:
[[10  0  0]
 [ 0  9  0]
 [ 0  0 11]]
```

Gambar 1. Akurasi Model dan *Confusion Matrix*

Confusion matrix menunjukkan hasil prediksi model untuk masing-masing spesies bunga. Dari total 30 sampel data uji yang terdiri dari 10 bunga Iris Setosa, 9 bunga Iris Versicolor, dan 11 bunga Iris Virginica, model berhasil membedakan ketiga spesies dengan sempurna. Pada baris pertama, seluruh 10 bunga Iris Setosa diprediksi dengan benar sebagai Setosa, tanpa kesalahan prediksi ke Versicolor maupun Virginica. Pada baris kedua, seluruh 9 bunga Iris Versicolor juga diprediksi dengan benar sebagai Versicolor, tanpa adanya prediksi salah ke Setosa maupun Virginica. Begitu pula pada baris ketiga, semua 11 bunga Iris Virginica diprediksi dengan benar sebagai Virginica, tanpa kesalahan ke Setosa maupun Versicolor. Hasil ini menunjukkan bahwa model mampu memanfaatkan fitur-fitur seperti **SepalLengthCm**, **SepalWidthCm**, **PetalLengthCm**, dan **PetalWidthCm** secara optimal untuk mengklasifikasikan spesies bunga dengan akurasi sempurna.



Gambar 2. Visualisasi *Confusion Matrix* Logistic Regression

3.2 *Confusion Matrix Random Forest*

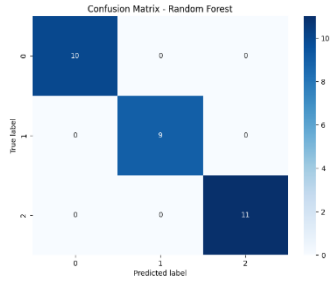
```
Random Forest Classifier Accuracy: 1.0000
```

Gambar 3. Akurasi *Random Forest*

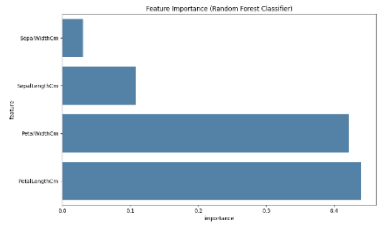
JRIIN: Jurnal Riset Informatika dan Inovasi
Volume 2, No. 10 Maret 2025
ISSN 3025-0919 (media online)
Hal 1927-1934

3.2.1 Akurasi Model

Nilai akurasi 1.0000 atau 100% menunjukkan bahwa model *Random Forest* berhasil mengklasifikasikan semua sampel pada data uji dengan benar. Tingkat akurasi ini setara dengan yang ditunjukkan oleh *Logistic Regression*.



Gambar 4. Visualisasi *Confusion Matrix Random Forest*



Gambar 5. *Feature Importance Random Forest*

3.2.2 Pentingnya Fitur pada *Random Forest*

Model *Random Forest* menunjukkan bahwa fitur-fitur yang digunakan memiliki kontribusi berbeda dalam membedakan jenis Iris:

1. *PetalLength*Cm (Panjang Mahkota Bunga): Memberikan pengaruh terbesar, yaitu 44%.
2. *PetalWidth*Cm (Lebar Mahkota Bunga): Memberikan pengaruh besar, yaitu 42%.
3. *SepalLength*Cm (Panjang Kelopak Bunga): Memberikan pengaruh lebih kecil, yaitu 11%.
4. *SepalWidth*Cm (Lebar Kelopak Bunga): Memberikan pengaruh paling kecil, yaitu 3%.

Untuk membedakan jenis *Iris*, fitur yang terkait dengan mahkota bunga (*PetalLengthCm* dan *PetalWidthCm*) lebih penting dibandingkan dengan fitur pada kelopak bunga (*SepalLengthCm* dan *SepalWidthCm*).

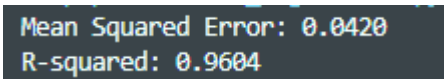
3.2.3 Perbandingan Performa Model

1. *Random Forest* adalah model yang mengandalkan analisis mendalam terhadap fitur-fitur bunga untuk menghasilkan prediksi. Dengan memanfaatkan algoritma ensemble learning, model ini sangat akurat dan berhasil mengklasifikasikan jenis bunga dengan akurasi sempurna sebesar 100% (Schonlau & Zou, 2020).
2. *Logistic Regression* menggunakan pendekatan berbasis "rumus sederhana" yang lebih ringan dibandingkan *Random Forest* (Joshi & Dhakal, 2021). Meskipun demikian, model ini juga menunjukkan akurasi yang sangat baik dengan hasil sempurna (100%). Hal ini menunjukkan bahwa algoritma sederhana pun dapat menghasilkan performa tinggi pada dataset dengan pola yang jelas.
3. *Linear Regression* dirancang untuk memprediksi nilai kontinu (*continuous value*) dan lebih cocok digunakan dalam analisis hubungan linear antara variabel input dan *output* (Fadlil dkk., 2023). Contoh penerapan model ini adalah visualisasi

JRIIN: Jurnal Riset Informatika dan Inovasi
Volume 2, No. 10 Maret 2025
ISSN 3025-0919 (media online)
Hal 1927-1934

grafik prediksi vs nilai aktual, yang menunjukkan hubungan linear yang kuat. Model ini efektif untuk menggambarkan pola prediksi kontinu dalam data.

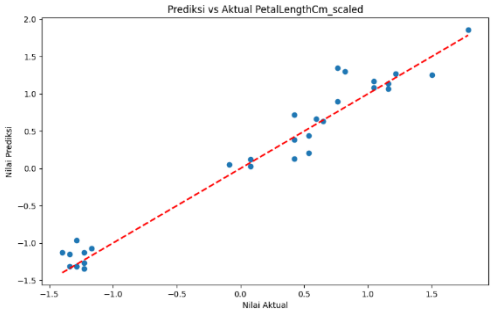
3.3 *Iris Preprocessed Regression Prediction (Linear Regression)*



Gambar 6. *MSE* dan *R-squared Linear Regression*

Hasil evaluasi model menggunakan *Mean Squared Error* (MSE) dan *R-squared* (R^2) menunjukkan performa yang sangat baik. Nilai MSE sebesar 0.0420 mengindikasikan rata-rata kesalahan kuadrat antara nilai prediksi dan nilai aktual yang sangat kecil, yang berarti model memberikan prediksi yang akurat. Semakin mendekati 0 nilai MSE, semakin baik kualitas prediksi model.

Sementara itu, nilai *R-squared* (R^2) sebesar 0.9604 menunjukkan bahwa model mampu menjelaskan 96.04% variasi dalam panjang petal (*PetalLength*) berdasarkan tiga fitur input yang digunakan. Nilai R^2 yang mendekati 1 menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam menjelaskan hubungan antara variabel *input* dan *output*, sehingga memberikan hasil yang sangat andal untuk prediksi kontinu.



Gambar 7. Hubungan Antara Nilai Prediksi dan Nilai Aktual (Setelah *Scaling*)

Visualisasi ini menggambarkan hubungan antara nilai prediksi dari model regresi dan nilai aktual dari data yang telah diskalakan. Pada grafik, sumbu X mewakili nilai aktual fitur *PetalLength*Cm setelah melalui proses *scaling*, sedangkan sumbu Y menunjukkan nilai prediksi dari model regresi untuk fitur yang sama. Garis merah dalam visualisasi merepresentasikan hubungan ideal di mana prediksi model sama persis dengan nilai aktual.

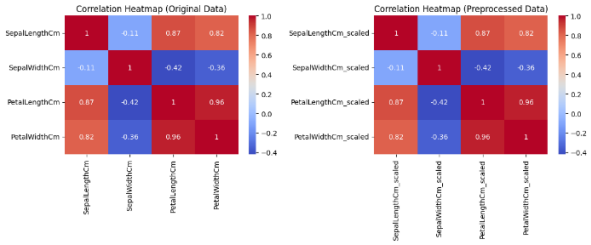
Titik-titik yang berada dekat dengan garis merah menunjukkan bahwa prediksi model sangat akurat, sementara titik-titik yang lebih jauh mengindikasikan adanya perbedaan kecil antara nilai prediksi dan nilai aktual, meskipun tetap berada dalam batas toleransi yang wajar.

Data terbagi menjadi tiga kelompok berdasarkan panjang petal. Kelompok bawah (sekitar -1.5) mewakili Iris Setosa, yang memiliki petal terpendek. Kelompok tengah (sekitar 0.5) adalah Iris Versicolor, dengan panjang petal sedang. Sementara itu, kelompok atas (sekitar 1.5) menggambarkan Iris Virginica, yang memiliki petal paling panjang. Visualisasi ini tidak hanya menunjukkan akurasi prediksi model tetapi juga memperkuat pola pemisahan spesies berdasarkan panjang petal.

3.4 *Correlation Heatmap*

JRIIN: Jurnal Riset Informatika dan Inovasi
Volume 2, No. 10 Maret 2025
ISSN 3025-0919 (media online)
Hal 1927-1934

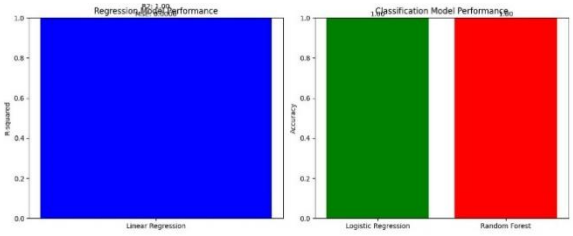
Program ini menampilkan dua ***heatmap*** untuk memperlihatkan korelasi antar variabel dalam dataset Iris sebelum dan sesudah proses *preprocessing* (*scaling*).



Gambar 8. *Correlation Heatmap Origin* dan *After Preprocessing*

A. *Correlation Heatmap* (Data Asli) *heatmap* ini menggambarkan hubungan korelasi antara variabel panjang dan lebar pada sepal serta petal. Warna merah menunjukkan korelasi positif yang kuat, sedangkan warna biru menunjukkan korelasi negatif. Sebagai contoh, panjang petal (PetalLengthCm) memiliki korelasi positif yang sangat kuat dengan lebar petal (*PetalWidth*Cm) dengan nilai korelasi sebesar 0.96. Hal ini menunjukkan bahwa semakin panjang petal, semakin lebar pula ukurannya.
B. *Correlation Heatmap* (Data Setelah *Preprocessing*) Setelah data melalui proses *scaling*, korelasi antar variabel tetap tidak berubah. Namun, data telah berada dalam skala yang lebih seragam, yang membantu meningkatkan kinerja model tertentu dengan memastikan bahwa fitur-fitur memiliki pengaruh yang lebih seimbang dalam analisis. Sebagai contoh, korelasi positif yang kuat antara panjang dan lebar petal tetap sama dengan nilai sebesar 0.96, meskipun data telah diskalakan. *Scaling* memastikan hasil analisis dan pemodelan menjadi lebih andal tanpa mengubah pola hubungan antar fitur.

3.5 *Performance Model*

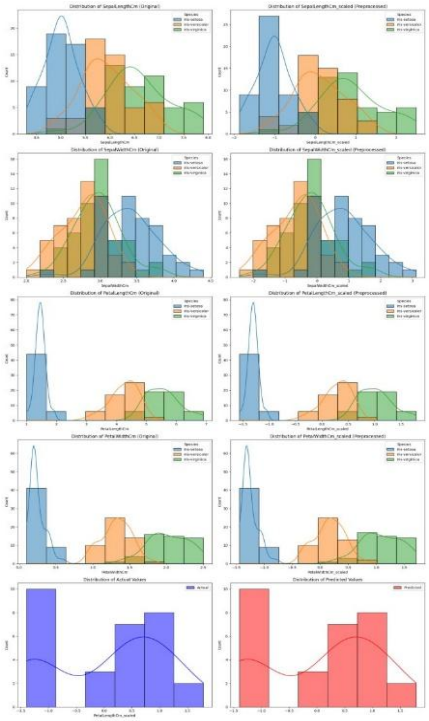


Gambar 9. Visualisasi Diagram Performa Model

Grafik di sisi kiri menunjukkan performa model *Linear Regression* dengan nilai R^2 sebesar 1.00, yang diilustrasikan dalam bentuk batang biru. Nilai R^2 = 1.00 ini mengindikasikan bahwa model mampu menjelaskan 100% variasi dalam data, sehingga menunjukkan performa sempurna dalam regresi. Sementara itu, grafik di sisi kanan membandingkan akurasi dua model klasifikasi, yaitu *Logistic Regression* (ditampilkan sebagai batang hijau) dan *Random Forest* (ditampilkan sebagai batang merah). Kedua model berhasil mencapai akurasi 100% dalam mengklasifikasikan spesies Iris, mencerminkan performa yang sama-sama sangat tinggi. Pada grafik ini, terdapat gap putih di tengah yang digunakan untuk memisahkan kedua model, sehingga memudahkan perbandingan visual.

JRIIN: Jurnal Riset Informatika dan Inovasi
Volume 2, No. 10 Maret 2025
ISSN 3025-0919 (media online)
Hal 1927-1934

3.6 Distribusi



Gambar 10. Visualisasi Distribusi Diagram

1. *SepalLength*Cm (2 Grafik Atas) Pada data asli, distribusi menunjukkan rentang nilai yang berbeda untuk setiap spesies: Iris-setosa berada di rentang 4.5–5.5 cm, *Iris-versicolor* pada 5.5 -6.5 cm, dan Iris-virginica pada 6.5–7.5 cm. Setelah *preprocessing*, data telah diskalakan ke rentang -2 hingga 2, namun pola distribusi tetap konsisten dengan data asli.
2. *SepalWidth*Cm (2 Grafik Tengah Atas) Pada data asli, terdapat overlap signifikan di rentang 2.0–4.5 cm untuk ketiga spesies, sehingga sulit untuk membedakan spesies berdasarkan fitur ini. Setelah preprocessing, data diskalakan ke rentang -2 hingga 3, dengan pola distribusi tetap serupa namun lebih seragam.
3. *PetalLength*Cm (2 Grafik Tengah Bawah) Data asli menunjukkan bahwa *Iris-setosa* terpisah jelas di rentang 1–2 cm, *Iris-versicolor* berada pada 4–5 cm, dan *Iris-virginica* pada 5–6 cm. Setelah diskalakan ke rentang -1.5 hingga 1.5, pemisahan antar spesies tetap terlihat jelas.
4. *PetalWidth*Cm (2Grafik Bawah) Pada data asli, *Iris-setosa* berada di rentang 0.0–0.5 cm, *Iris versicolor* di 1.0–1.5 cm, dan *Iris-virginica* pada 1.5–2.5 cm. Setelah preprocessing ke rentang -1.5 hingga 1.5, pola distribusi tetap mencerminkan pemisahan antar spesies yang jelas.
5. Distribusi Actual vs Predicted (2 Grafik Paling Bawah) Grafik ini menunjukkan distribusi nilai actual (biru) dibandingkan dengan nilai prediksi (merah). Kesamaan pola distribusi antara keduanya menunjukkan bahwa model menghasilkan prediksi yang sangat akurat.

4. KESIMPULAN

Dengan menggunakan algoritma *Regresi Logistik*, *Random Forest*, dan *Regresi Linier* pada dataset Iris berhasil memberikan hasil yang sangat baik. Sementara *Random Forest* dan *Regresi Logistic* mencapai akurasi 100% dalam tugas klasifikasi, *Regresi Linier* menghasilkan R2 = 1.00, yang menunjukkan kemampuan prediksi yang sangat akurat.

Dalam hal mengidentifikasi karakteristik, fitur petal (panjang dan lebar mahkota bunga) memiliki keunggulan yang signifikan dibandingkan dengan fitur sepal (panjang dan lebar kelopak bunga). Visualisasi hasil prediksi menunjukkan distribusi nilai aktual dan prediksi yang hampir sama. Penelitian ini menunjukkan keefektifan algoritma yang digunakan untuk memecahkan masalah klasifikasi dan prediksi dalam dataset yang tidak terkait tetapi relevan, seperti dataset Iris.

Dengan demikian hasil ini dapat menjadi dasar untuk pengembangan penelitian lebih lanjut dengan memperluas dataset atau menggunakan algoritma tambahan untuk meningkatkan kinerja model dan generalisasi.

REFERENCES

Arifin, O., & Rofianto, D. (2023). Perbandingan metode klasifikasi SOM dan LVQ pada data bunga iris dengan parameter dimodifikasi. *Decode: Jurnal Pendidikan Teknologi Informasi*, *3*(1), 130–138. https://doi.org/10.51454/decode.v3i1.135

Fadlil, A., Herman, & Dikky Praseptian, M. (2023). Single imputation using statistics-based and K Nearest Neighbor methods for numerical datasets. *Ingenierie des Systemes d'Information*, *28*(2), 451–459. https://doi.org/10.18280/isi.280221

Fahmi, M. N. (2023). Implementasi machine learning menggunakan Python library: Scikit-learn (supervised dan unsupervised learning). *Sains Data Jurnal Studi Matematika dan Teknologi*, *1*(2), 87–96. https://doi.org/10.52620/sainsdata.v1i2.31

Joshi, R. D., & Dhakal, C. K. (2021). Predicting type 2 diabetes using logistic regression and machine learning approaches. *International Journal of Environmental Research and Public Health*, *18*(14). https://doi.org/10.3390/ijerph18147346

Rahman, B., Fauzi, F., & Amri, S. (2023). Perbandingan Hasil Klasifikasi Data Iris menggunakan Algoritma K-Nearest Neighbor dan Random Forest. *Journal Of Data Insights*, *1*(1), 19–26. https://doi.org/10.26714/jodi.v1i1.135

Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *Stata Journal*, *20*(1), 3–29. https://doi.org/10.1177/1536867X20909688

Setia Budi, E., Nofriyaldi Chan, A., Priscillia Alda, P., & Arif Fauzi Idris, M. (2024). Optimasi model machine learning untuk klasifikasi dan prediksi citra menggunakan algoritma convolutional neural network. *Media Online*, *4*(5), 502–509. https://doi.org/10.30865/resolusi.v4i5.1892

Simanjuntak, W. O., Bijaksana, A., Negara, P., & Septriana, R. (2023). Perbandingan algoritma logistic regression dan random forest (Studi kasus: Klasifikasi emosi tweet). *Jurnal Aplikasi dan Riset Informatika*, *2*, 160–164. https://doi.org/10.26418/juara.v2i1.69682

Susetyoko, R., Yuwono, W., Purwantini, E., & Ramadijanti, N. (2022). Perbandingan metode random forest, regresi logistik, naïve bayes, dan multilayer perceptron pada klasifikasi uang kuliah tunggal (UKT). *Jurnal Infomedia: Teknik Informatika, Multimedia & Jaringan*, *7*(1), 8–16.