

Klasifikasi Tingkat Kematangan Buah Strawberry Menggunakan Algoritma Logistic Regression

Zulkifli^{1)*}, Riyadhul Fajri²⁾

¹⁾ Program Studi Sistem Informasi, Universitas Almuslim Bireuen, ²⁾ Program Studi Informatika, Universitas Almuslim Bireuen

¹⁾ zulladasicupak@gmail.com, ²⁾ fajri071113@gmail.com

Received: 22 October 2024

Accepted: 17 December 2024

Published: 25 December 2024



*zulladasicupak@gmail.com

Kata Kunci: Logistic Regression, Klasifikasi Kematangan, L2 Ridge, Machine Learning, Strawberry.

DSI: Jurnal Data Science Indonesia is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak : Penelitian ini bertujuan untuk mengembangkan model klasifikasi tingkat kematangan buah strawberry menggunakan algoritma Logistic Regression dengan regularisasi L2 Ridge. Dataset yang digunakan berupa gambar strawberry yang terbagi dalam dua kategori: matang dan belum matang. Proses klasifikasi dilakukan melalui platform Orange Data Mining, yang mempermudah pengolahan data visual dengan fitur image embedding untuk mengonversi gambar menjadi vektor numerik. Hasil evaluasi menunjukkan bahwa model mencapai akurasi sebesar 97,7%, dengan AUC sebesar 0,995, serta nilai precision, recall, dan F1-score masing-masing sebesar 0,977. Confusion matrix menunjukkan bahwa model hanya membuat 12 kesalahan prediksi dari 521 gambar, dengan 7 gambar belum matang diprediksi sebagai matang, dan 5 gambar matang diprediksi sebagai belum matang. Temuan ini menunjukkan bahwa Logistic Regression sangat efektif dalam tugas klasifikasi biner, terutama dalam mendeteksi kematangan buah strawberry. Meskipun terdapat beberapa kesalahan prediksi, hasil keseluruhan memperlihatkan potensi besar penerapan machine learning dalam otomatisasi proses deteksi kematangan buah. Penelitian ini menyarankan penggunaan dataset yang lebih luas dan eksplorasi model yang lebih kompleks untuk meningkatkan performa klasifikasi.

PENDAHULUAN

Seiring berkembangnya teknologi dan inovasi dalam bidang pertanian, penggunaan teknologi cerdas untuk meningkatkan efisiensi dan produktivitas menjadi semakin krusial [1], [2], [3]. Salah satu aspek penting dalam industri pertanian adalah proses pemanenan buah yang optimal, yang sangat bergantung pada penentuan tingkat kematangan buah [4]. Kematangan buah yang tepat sangat mempengaruhi kualitas rasa, daya tahan, dan nilai jualnya di pasar [5]. Pada industri pertanian tradisional, penentuan kematangan buah, termasuk strawberry, umumnya dilakukan secara manual oleh petani berdasarkan pengalaman visual dan sensori. Meskipun metode ini efektif untuk skala kecil, ada banyak kelemahan yang muncul ketika diterapkan pada skala besar, seperti ketidakakuratan, subjektivitas, dan konsumsi waktu yang signifikan [6].

Strawberry merupakan salah satu buah yang memiliki nilai ekonomi tinggi di berbagai negara [7]. Namun, salah satu tantangan utama dalam budidaya strawberry adalah proses panen yang membutuhkan ketepatan waktu. Pemanenan yang dilakukan terlalu cepat akan menghasilkan buah yang belum matang, sementara pemanenan yang terlambat dapat mengakibatkan buah menjadi terlalu matang atau busuk, yang tentunya menurunkan kualitas produk. Oleh karena itu, diperlukan metode yang lebih efisien dan akurat untuk membantu petani dalam menentukan waktu panen yang optimal.

Kemajuan dalam bidang kecerdasan buatan (artificial intelligence) dan pembelajaran mesin (machine learning) telah membuka peluang baru dalam pengembangan solusi otomatis untuk tantangan-tantangan seperti klasifikasi tingkat kematangan buah [8]. Salah satu algoritma yang populer dan sering digunakan

dalam klasifikasi biner adalah Logistic Regression. Algoritma ini sederhana namun efektif dalam memodelkan hubungan antara fitur-fitur independen dan label klasifikasi, sehingga cocok untuk tugas-tugas yang melibatkan klasifikasi objek ke dalam dua kategori, seperti tingkat kematangan buah strawberry.

Dalam penelitian ini, algoritma Logistic Regression digunakan untuk mengembangkan model klasifikasi otomatis yang dapat membedakan antara buah strawberry yang sudah matang dan yang belum matang. Dengan menggunakan dataset terbuka yang terdiri dari dua label, yaitu "matang" dan "belum matang," serta total 521 data latih (258 untuk label "belum matang" dan 263 untuk label "matang"), penelitian ini berfokus pada pembuatan model prediktif yang dapat secara akurat mengklasifikasikan tingkat kematangan buah strawberry berdasarkan citra visual. Penggunaan machine learning dalam klasifikasi kematangan buah ini diharapkan dapat menggantikan metode manual yang rentan terhadap kesalahan dan subjektivitas, serta mempercepat proses penentuan waktu panen, terutama pada skala pertanian yang lebih besar.

Lebih jauh, penelitian ini bertujuan untuk memberikan solusi praktis bagi para petani dan industri agrikultur dengan menghadirkan sistem yang tidak hanya akurat, tetapi juga mudah diterapkan. Teknologi ini berpotensi meningkatkan efisiensi dan kualitas hasil panen, serta mengurangi kerugian yang disebabkan oleh panen yang tidak tepat waktu. Selain itu, pendekatan ini juga dapat menjadi landasan bagi pengembangan lebih lanjut dalam klasifikasi kematangan berbagai jenis buah lainnya, yang memiliki karakteristik visual berbeda. Dalam jangka panjang, penerapan machine learning pada sektor pertanian ini dapat berkontribusi pada terciptanya sistem pertanian cerdas yang lebih berkelanjutan, efisien, dan ramah lingkungan.

TINJAUAN LITERATUR

Dalam dunia pertanian modern, penentuan tingkat kematangan buah secara otomatis menjadi aspek penting untuk meningkatkan kualitas hasil panen dan efisiensi waktu. Berbagai metode berbasis kecerdasan buatan, terutama machine learning, telah diterapkan untuk tugas-tugas klasifikasi ini [9], [10], [11], [12]. Salah satu penelitian terbaru yang relevan dengan topik ini adalah penggunaan model pembelajaran mesin berbasis algoritma Ensemble untuk mengidentifikasi kematangan buah dengan lebih akurat. Penelitian ini mengkombinasikan teknik Support Vector Machine (SVM) dan jaringan saraf dalam (ResNet) untuk meningkatkan efisiensi pengenalan kematangan dan kesehatan buah secara keseluruhan, meskipun tidak spesifik untuk strawberry [13], [14], [15], [16].

Metode lain yang sering digunakan untuk klasifikasi kematangan buah adalah teknik pengolahan citra visual, yang umumnya mengandalkan fitur visual seperti warna dan ukuran buah [17], [18]. Sebuah penelitian menggunakan sistem visi berbasis robot untuk mengklasifikasikan tingkat kematangan strawberry. Dalam penelitian ini, citra strawberry diolah menggunakan teknik segmentasi gambar dan pengklasifikasian bentuk, menghasilkan akurasi lebih dari 90% dalam penentuan kematangan. Selain itu, metode pengolahan citra dan teknik segmentasi digunakan dalam penelitian lain yang lebih luas, di mana berbagai jenis buah termasuk strawberry dianalisis untuk mengidentifikasi kematangan secara akurat.

Dalam penelitian yang lebih spesifik, algoritma Logistic Regression digunakan untuk klasifikasi biner tingkat kematangan buah. Algoritma ini sederhana namun efektif dalam mengklasifikasikan objek ke dalam dua kategori, seperti buah matang dan belum matang. Salah satu pendekatan untuk klasifikasi ini adalah penggunaan data bioimpedansi, di mana Logistic Regression digunakan bersama dengan metode pembelajaran mesin terawasi lainnya untuk mengklasifikasikan kematangan strawberry. Penelitian ini menunjukkan potensi besar penggunaan Logistic Regression dalam mengolah data untuk prediksi tingkat kematangan, meskipun model pembelajaran yang lebih kompleks seperti Multi-Layer Perceptron (MLP) dapat memberikan hasil yang lebih akurat [19].

Lebih lanjut, penelitian lain yang relevan mengembangkan sistem berbasis visi otomatis menggunakan jaringan saraf konvolusi (CNN). Sistem ini mampu mengidentifikasi kematangan strawberry dengan tingkat akurasi sebesar 91.6%, menggunakan fitur-fitur visual seperti warna permukaan dan ukuran. Ini menunjukkan bahwa teknik berbasis CNN atau metode pembelajaran mendalam lainnya memiliki potensi besar dalam aplikasi otomatisasi di sektor pertanian, termasuk untuk buah strawberry [20].

Sementara banyak penelitian telah memfokuskan pada strawberry, ada juga aplikasi yang relevan dalam klasifikasi kematangan buah lainnya. Sebuah studi pada buah mangga menggunakan algoritma Faster R-CNN untuk mendeteksi kematangan buah. Hasilnya menunjukkan bahwa sistem ini mampu menentukan dan

mendeteksi tingkat kematangan buah dengan efisien dan biaya rendah . Selain itu, sistem berbasis kamera yang digunakan dalam robot pertanian juga telah diaplikasikan pada strawberry, dengan menggunakan deteksi warna sebagai cara cepat untuk mengklasifikasikan kematangan .

Studi lain mengembangkan sistem otomatis berbasis microcontroller dan kamera yang mampu mendeteksi tingkat kematangan strawberry dengan tingkat keberhasilan 87.5% . Penerapan teknologi berbasis kamera ini juga menunjukkan bagaimana mesin dapat diintegrasikan ke dalam sistem robot untuk membantu proses pemanenan otomatis, mengurangi ketergantungan pada tenaga kerja manual yang cenderung kurang akurat [21].

Penggunaan teknologi ini memiliki potensi besar untuk meningkatkan efisiensi di sektor pertanian. Teknologi berbasis machine learning tidak hanya memungkinkan klasifikasi kematangan buah secara lebih akurat, tetapi juga dapat mempercepat proses pemanenan dan mengurangi kerugian yang disebabkan oleh panen yang tidak tepat waktu. Dengan adanya teknologi ini, petani dapat lebih mudah menentukan kapan buah siap dipanen, terutama pada skala besar, yang mana metode manual sulit diterapkan.

Salah satu tantangan dalam penerapan teknologi otomatis ini adalah biaya dan kerumitan penerapannya di lapangan. Namun, seiring dengan kemajuan teknologi, biaya implementasi sistem berbasis kecerdasan buatan diperkirakan akan semakin terjangkau, sehingga memungkinkan lebih banyak petani untuk mengadopsi teknologi ini.

Penelitian mengenai penerapan machine learning, termasuk algoritma Logistic Regression, dalam klasifikasi kematangan buah seperti strawberry telah menunjukkan hasil yang menjanjikan. Dari teknik pengolahan citra hingga penggunaan data bioimpedansi, berbagai pendekatan ini memberikan solusi yang lebih akurat dan efisien dibandingkan metode manual tradisional. Pengembangan lebih lanjut dalam penggunaan kecerdasan buatan di bidang pertanian diharapkan dapat menciptakan sistem yang lebih berkelanjutan, efisien, dan ramah lingkungan.

Penerapan teknologi ini tidak hanya terbatas pada strawberry, namun juga dapat diterapkan pada berbagai jenis buah lain yang memiliki karakteristik visual berbeda. Dengan demikian, dalam jangka panjang, pengembangan metode otomatis ini dapat menjadi landasan bagi sistem pertanian cerdas yang lebih maju dan adaptif terhadap kebutuhan masa depan .

METODE PENELITIAN

Penelitian ini bertujuan untuk membangun model klasifikasi tingkat kematangan buah strawberry menggunakan algoritma Logistic Regression dengan regularisasi L2 Ridge dan parameter regularisasi $C=1$. Untuk melakukan tugas klasifikasi ini, peneliti menggunakan alat Orange Data Mining, yang menyediakan platform visual untuk penerapan machine learning. Langkah-langkah penelitian ini dijelaskan sebagai berikut:

1. Pengumpulan Data

Penelitian ini menggunakan dataset terbuka yang terdiri dari gambar-gambar buah strawberry. Gambar tersebut dikelompokkan dalam dua label:

- a) Matang: Gambar strawberry yang sudah siap dipanen.
- b) Belum Matang: Gambar strawberry yang belum mencapai tingkat kematangan optimal.

Total dataset yang digunakan dalam penelitian ini adalah sebanyak 521 gambar, terdiri dari 263 gambar untuk kelas "matang" dan 258 gambar untuk kelas "belum matang." Dataset ini kemudian digunakan sebagai input untuk pelatihan model klasifikasi.

2. Preprocessing Data

Tahap preprocessing data mencakup beberapa langkah untuk memastikan dataset siap digunakan dalam proses pelatihan model. Langkah-langkah preprocessing meliputi:

- a) Resizing Gambar: Setiap gambar diubah ukurannya menjadi ukuran yang seragam, misalnya 128x128 piksel, untuk mengurangi kompleksitas komputasi dan memastikan konsistensi ukuran input.
- b) Normalisasi: Nilai piksel dalam gambar dinormalisasi ke dalam rentang 0 hingga 1 untuk mempercepat proses pelatihan dan meningkatkan konvergensi model.

- c) Ekstraksi Fitur: Orange Data Mining secara otomatis mengekstraksi fitur yang relevan dari gambar untuk digunakan dalam pelatihan model, seperti representasi vektor piksel.

3. Pembagian Data

Dataset dibagi menjadi dua subset untuk melatih dan menguji model:

- Data Latih (Training Data): Sebanyak 80% dari total dataset digunakan untuk melatih model (416 gambar, terdiri dari 211 gambar "matang" dan 205 gambar "belum matang").
- Data Uji (Testing Data): Sebanyak 20% dari dataset digunakan untuk menguji performa model (105 gambar, terdiri dari 52 gambar "matang" dan 53 gambar "belum matang").

Proses pembagian ini dilakukan secara acak menggunakan fitur bawaan Orange Data Mining untuk memastikan representasi yang seimbang antara kelas "matang" dan "belum matang" dalam setiap subset.

4. Pengembangan Model Logistic Regression

Model Logistic Regression dengan regularisasi L2 Ridge digunakan untuk melakukan klasifikasi. Regularisasi L2 berfungsi untuk mengurangi kompleksitas model dan mencegah overfitting, dengan parameter regularisasi $C=1$. Logistic Regression bekerja dengan memprediksi probabilitas bahwa gambar yang diberikan termasuk dalam salah satu dari dua kelas yang ada. Persamaan matematis yang digunakan adalah:

$$P(y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)}}$$

Dengan regularisasi L2, fungsi kerugian yang dioptimalkan adalah:

$$\text{Loss Function} = -(y \log(p) + (1 - y) \log(1 - p)) + \lambda \sum_{i=1}^n \beta_i^2$$

di mana λ adalah parameter regularisasi, yang dikontrol oleh nilai $C=1$ dalam penelitian ini.

5. Pelatihan Model Menggunakan Orange Data Mining

Pelatihan model dilakukan menggunakan Orange Data Mining, yang memungkinkan peneliti melakukan pelatihan secara visual dan interaktif. Dengan memasukkan data latih ke dalam model Logistic Regression di Orange, proses pelatihan berjalan dengan memperkirakan parameter $\beta_0, \beta_1, \dots, \beta_n$ menggunakan algoritma optimasi yang tersedia dalam platform tersebut. Penggunaan regularisasi L2 dengan parameter $C=1$ memastikan bahwa model tidak terlalu menyesuaikan diri dengan data latih (overfitting), sambil tetap mempertahankan akurasi prediksi.

6. Evaluasi Model

Setelah pelatihan selesai, model dievaluasi menggunakan data uji. Orange Data Mining menyediakan berbagai metrik evaluasi yang relevan untuk klasifikasi biner, termasuk:

- Akurasi: Persentase prediksi yang benar dari keseluruhan data uji.

$$\text{Akurasi} = \frac{\text{Prediksi Benar}}{\text{Total Data Uji}} \times 100\%$$

- Precision: Kemampuan model untuk menghasilkan prediksi positif yang benar.

$$\text{Precision} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Positive (FP)}}$$

- c) Recall: Kemampuan model untuk mendeteksi semua kasus positif.

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negative\ (FN)}$$

- d) F1-Score: Nilai rata-rata harmonis dari precision dan recall.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

7. Analisis Hasil

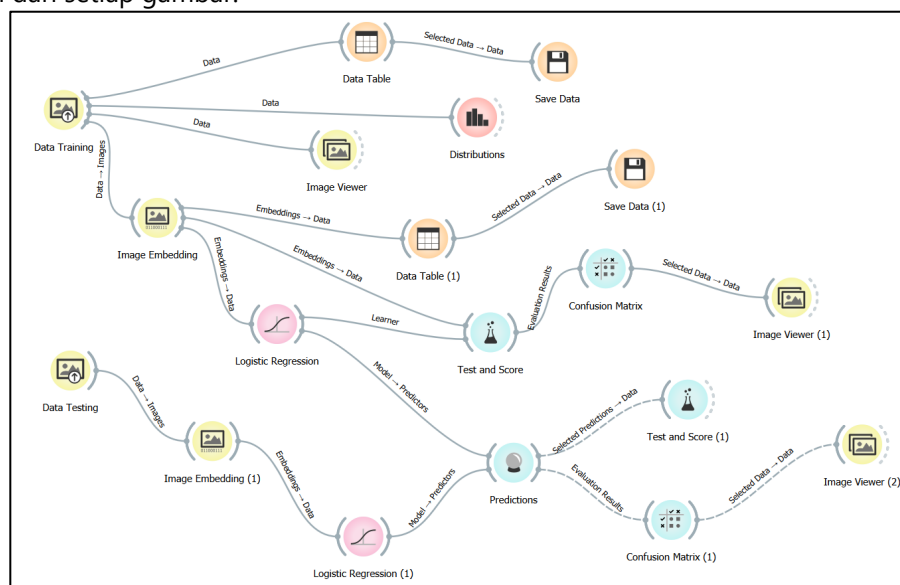
Setelah evaluasi, hasil model dianalisis untuk mengukur kinerja Logistic Regression dengan L2 Ridge regularization. Performa model dibandingkan dengan baseline atau pendekatan lain jika diperlukan. Selain itu, Orange Data Mining juga memfasilitasi pemahaman visual terhadap distribusi hasil prediksi, yang memberikan wawasan lebih lanjut tentang kinerja model dalam klasifikasi tingkat kematangan strawberry.

8. Pelaporan dan Publikasi

Hasil penelitian ini disusun dalam bentuk laporan ilmiah yang mencakup metodologi, hasil, dan analisis. Peneliti berharap hasil penelitian ini dapat dipublikasikan di jurnal ilmiah terkait klasifikasi citra dan aplikasi machine learning dalam pertanian.

HASIL PENELITIAN

Penelitian ini berhasil mengembangkan model klasifikasi tingkat kematangan buah strawberry menggunakan algoritma Logistic Regression dalam platform Orange Data Mining. Proses klasifikasi dimulai dari impor data gambar yang terbagi ke dalam dua subset: data training, yang digunakan untuk melatih model, dan data testing, yang digunakan untuk menguji kinerja model. Pada tahap pertama, gambar-gambar strawberry diimpor dan dikonversi menjadi vektor numerik melalui proses Image Embedding. Model embedding, seperti ResNet atau VGG, digunakan untuk menghasilkan fitur numerik yang merepresentasikan karakteristik visual dari setiap gambar.



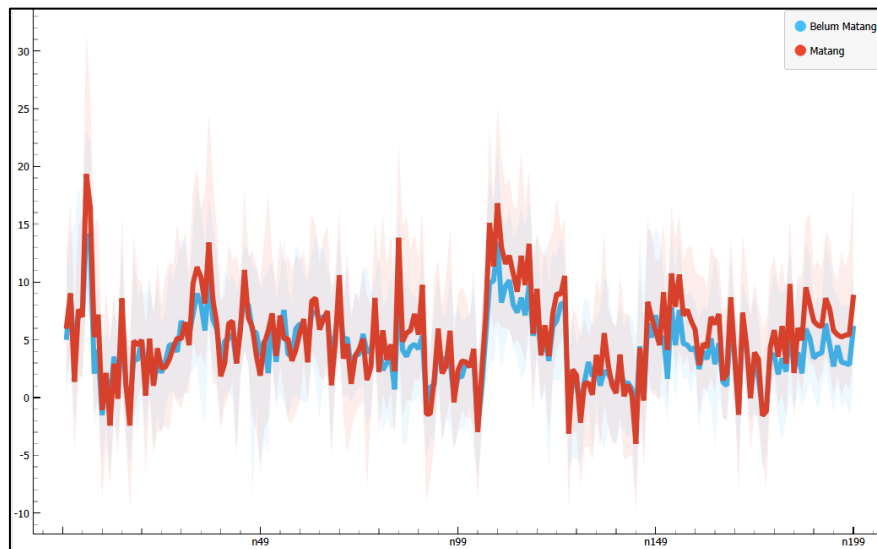
Gambar 1. Alur Kerja Proses Klasifikasi

Setelah proses embedding selesai, model Logistic Regression dilatih menggunakan data training. Regularisasi L2 Ridge dengan parameter $C=1$ diterapkan pada model untuk mencegah overfitting dan mengoptimalkan akurasi prediksi. Dengan menggunakan embedding dari gambar, model Logistic Regression dilatih untuk membedakan antara gambar strawberry yang "matang" dan "belum matang". Setelah model dilatih, langkah selanjutnya adalah mengujinya menggunakan data testing melalui widget Test and Score.

Model	AUC	CA	F1	Prec	Recall	MCC
Logistic Regression	0.995	0.977	0.977	0.977	0.977	0.954

Gambar 2. Hasil Pengujian Matriks Evaluasi

Berdasarkan metrik evaluasi yang dihasilkan, model Logistic Regression mencapai AUC (Area Under the Curve) sebesar 0,995, yang menunjukkan bahwa model memiliki kemampuan yang sangat tinggi untuk membedakan antara buah yang "matang" dan "belum matang." Nilai AUC mendekati 1 menandakan bahwa model hampir selalu membuat prediksi yang benar, baik dalam mendeteksi positif sejati maupun negatif sejati.



Gambar 3. Grafik AUC

Akurasi Klasifikasi (CA) yang diperoleh adalah 97,7%, yang berarti bahwa dari seluruh data uji, hampir 98% prediksi yang dibuat oleh model adalah benar. Tingkat akurasi ini menunjukkan bahwa model memiliki kinerja yang konsisten dan dapat diandalkan dalam tugas klasifikasi biner ini. Akurasi dihitung sebagai:

$$Akurasi = \frac{True\ Positives\ (TP) + True\ Negatives\ (TN)}{Total\ Data}$$

Selanjutnya, metrik F1-Score yang diperoleh adalah 0,977, yang menunjukkan bahwa model Logistic Regression memiliki keseimbangan yang baik antara precision dan recall. F1-Score dihitung sebagai rata-rata harmonis dari precision dan recall, yang penting untuk memastikan bahwa model tidak hanya akurat dalam prediksi positif, tetapi juga dapat mendeteksi kelas yang benar dengan baik. Nilai Precision sebesar 0,977 menunjukkan bahwa dari semua prediksi buah yang dianggap matang, 97,7% di antaranya benar-benar matang. Nilai Recall yang juga sebesar 0,977 menandakan bahwa model berhasil mendeteksi 97,7% dari total buah yang sebenarnya matang.

Metrik lain yang penting adalah MCC (Matthews Correlation Coefficient), yang memberikan ukuran kualitas klasifikasi biner dengan mempertimbangkan keseimbangan antara kelas positif dan negatif. Nilai

MCC sebesar 0,954 menegaskan bahwa model ini dapat menangani dataset yang seimbang dengan sangat baik dan memberikan hasil klasifikasi yang hampir sempurna.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Di mana:

- TP adalah jumlah True Positives (buah matang yang diprediksi dengan benar sebagai matang),
- TN adalah jumlah True Negatives (buah belum matang yang diprediksi dengan benar sebagai belum matang),
- FP adalah jumlah False Positives (buah belum matang yang salah diprediksi sebagai matang),
- FN adalah jumlah False Negatives (buah matang yang salah diprediksi sebagai belum matang).

Berdasarkan hasil ini, dapat disimpulkan bahwa Logistic Regression dengan regularisasi L2 Ridge adalah model yang sangat efektif untuk tugas klasifikasi tingkat kematangan buah strawberry. Dengan akurasi yang tinggi dan keseimbangan yang kuat antara precision dan recall, model ini dapat diandalkan untuk memprediksi kematangan buah dengan tingkat kesalahan yang sangat rendah.

		Predicted		
		Belum Matang	Matang	Σ
Actual	Belum Matang	251	7	258
	Matang	5	258	263
Σ		256	265	521

Gambar 4. Confusion Matrix

Berdasarkan gambar confusion matrix yang dihasilkan dari proses training, dapat dilakukan analisis mengenai performa model Logistic Regression dalam mengklasifikasikan tingkat kematangan buah strawberry. Confusion matrix ini membandingkan hasil prediksi model dengan label asli dari data pengujian, yang terdiri dari dua kategori, yaitu "matang" dan "belum matang." Dalam matriks ini, model berhasil memprediksi dengan benar sebanyak 258 gambar buah strawberry yang matang dan 251 gambar yang belum matang, menunjukkan kemampuan model yang sangat baik dalam mengidentifikasi kedua kategori tersebut. Namun, terdapat 7 gambar yang sebenarnya belum matang, tetapi diprediksi oleh model sebagai matang, yang menunjukkan adanya kesalahan dalam mengklasifikasikan beberapa buah yang belum matang. Selain itu, model juga membuat kesalahan pada 5 gambar yang sebenarnya matang tetapi diprediksi sebagai belum matang.

Dari total 521 gambar yang diuji, model hanya membuat 12 kesalahan prediksi, yang mencakup 7 kesalahan pada buah yang belum matang dan 5 kesalahan pada buah yang matang. Hal ini mengindikasikan bahwa model Logistic Regression menunjukkan performa yang sangat baik dalam klasifikasi ini, dengan hanya sedikit kesalahan. Kesalahan tersebut kemungkinan disebabkan oleh faktor seperti kemiripan visual antara gambar buah matang dan belum matang atau variasi pencahayaan dalam beberapa gambar. Secara keseluruhan, akurasi model dalam mengklasifikasikan tingkat kematangan buah strawberry cukup tinggi, dan dengan beberapa perbaikan seperti peningkatan jumlah data atau penggunaan teknik augmentasi data untuk menangani variasi gambar, akurasi ini dapat lebih ditingkatkan dan kesalahan prediksi dapat diminimalkan.

PEMBAHASAN

Pada penelitian ini, model Logistic Regression digunakan untuk mengklasifikasikan tingkat kematangan buah strawberry dengan hasil yang sangat memuaskan. Berdasarkan hasil evaluasi model menggunakan

Orange Data Mining, akurasi model tercatat sebesar 97,7%. Ini menunjukkan bahwa hampir seluruh prediksi yang dibuat oleh model terhadap data uji adalah benar. Penggunaan regularisasi L2 Ridge dengan parameter $C=1$ terbukti efektif dalam mencegah overfitting, sehingga model mampu melakukan generalisasi dengan baik pada data uji. Selain itu, nilai AUC yang mencapai 0,995 menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan antara gambar buah yang matang dan belum matang. Hal ini menunjukkan bahwa Logistic Regression dapat diandalkan dalam tugas klasifikasi ini.

Lebih lanjut, metrik precision, recall, dan F1-score yang semuanya berada di angka 0,977 menegaskan keseimbangan performa model dalam mendeteksi kedua kelas, baik "matang" maupun "belum matang". Metrik ini menunjukkan bahwa model tidak hanya akurat dalam memprediksi gambar buah matang, tetapi juga mampu mendeteksi sebagian besar gambar buah matang dan belum matang dengan benar. Confusion matrix yang ditampilkan memberikan pandangan lebih jelas mengenai distribusi hasil prediksi. Dari total 521 gambar, model membuat 12 kesalahan prediksi, dengan 7 gambar belum matang diprediksi sebagai matang dan 5 gambar matang diprediksi sebagai belum matang. Kesalahan-kesalahan ini relatif kecil, mengingat jumlah data dan kompleksitas gambar.

Kemungkinan penyebab kesalahan ini bisa berasal dari faktor visual yang sulit dibedakan antara buah matang dan belum matang, atau variasi dalam kondisi gambar seperti pencahayaan dan sudut pengambilan gambar. Meskipun demikian, tingkat kesalahan yang rendah ini menunjukkan bahwa model bekerja sangat baik secara keseluruhan. Keberhasilan model Logistic Regression dalam penelitian ini juga didorong oleh penggunaan image embedding, yang membantu mengonversi gambar menjadi representasi numerik yang efektif untuk proses klasifikasi.

Dengan mempertimbangkan hasil ini, penelitian ini memberikan kontribusi penting dalam penggunaan machine learning, khususnya Logistic Regression, untuk aplikasi di bidang agrikultur, khususnya dalam mendeteksi kematangan buah strawberry. Selain itu, penggunaan platform Orange Data Mining juga memberikan kemudahan dalam merancang, melatih, dan mengevaluasi model tanpa memerlukan intervensi manual yang kompleks, membuatnya ideal untuk penelitian berbasis data visual. Namun, untuk penelitian lanjutan, disarankan untuk memperluas dataset, baik dalam hal jumlah gambar maupun variasi dalam kondisi gambar, guna mengurangi kemungkinan kesalahan prediksi lebih lanjut. Teknik augmentasi data atau penggunaan model yang lebih kompleks seperti deep learning juga bisa dieksplorasi untuk meningkatkan performa lebih lanjut. Secara keseluruhan, penelitian ini menunjukkan bahwa Logistic Regression adalah metode yang sangat efektif untuk tugas klasifikasi biner pada dataset citra, dengan potensi besar untuk diterapkan pada aplikasi serupa di bidang lain.

KESIMPULAN

Kesimpulan dari penelitian ini adalah bahwa algoritma Logistic Regression dengan regularisasi L2 Ridge terbukti sangat efektif dalam mengklasifikasikan tingkat kematangan buah strawberry berdasarkan gambar. Dengan akurasi sebesar 97,7%, model menunjukkan kinerja yang sangat baik dalam memprediksi kategori matang dan belum matang. Nilai AUC sebesar 0,995 juga mengindikasikan kemampuan model yang hampir sempurna dalam membedakan kedua kelas, sementara metrik precision, recall, dan F1-score yang mencapai 0,977 menegaskan keseimbangan antara akurasi dan kemampuan deteksi model.

Meskipun terdapat beberapa kesalahan prediksi (12 dari 521 gambar), kesalahan tersebut relatif kecil dan kemungkinan disebabkan oleh faktor visual seperti kemiripan antara gambar buah matang dan belum matang atau variasi dalam kondisi pencahayaan. Secara keseluruhan, penggunaan Logistic Regression dalam penelitian ini menunjukkan bahwa model dapat diandalkan untuk tugas klasifikasi kematangan buah dengan tingkat akurasi yang tinggi.

Keberhasilan model ini juga didorong oleh proses embedding gambar yang efektif dalam mengubah data visual menjadi fitur numerik yang relevan untuk klasifikasi. Dengan demikian, penelitian ini memberikan kontribusi penting dalam penerapan machine learning di bidang agrikultur, khususnya dalam otomatisasi deteksi kematangan buah. Untuk penelitian selanjutnya, disarankan agar dataset diperluas dan model yang lebih kompleks dieksplorasi guna meningkatkan performa lebih lanjut.

REFERENCES

- [1] I. W. B. Darmawan, I. N. S. Kumara, and D. C. Khrisne, "SMART GARDEN SEBAGAI IMPLEMENTASI SISTEM KONTROL DAN MONITORING TANAMAN BERBASIS TEKNOLOGI CERDAS," *Jurnal SPEKTRUM*, vol. 8, no. 4, 2022, doi: 10.24843/spektrum.2021.v08.i04.p19.
- [2] I. Muslem, "Prototype Kunci RFID (Radio Frequency Identification) dalam Meningkatkan Keamanan Kendaraan Bermotor," *JURNAL TIKA*, vol. 5, no. 3, 2021, doi: 10.51179/tika.v5i3.104.
- [3] I. Muslem R, "Sistem Pendeteksi Kebocoran Gas Rumah Tangga Menggunakan Mq-2 Sensor Dan Mikrokontroler," *JURNAL TIKA*, vol. 6, no. 02, 2021, doi: 10.51179/tika.v6i02.457.
- [4] E. T. Kembaren and Muchsin, "Pengelolaan Pasca Panen Kopi Arabika Gayo Aceh," *Jurnal Visioner dan Strategis*, vol. 10, no. 1, 2021.
- [5] Y. Permadi and . Murinto, "APLIKASI PENGOLAHAN CITRA UNTUK IDENTIFIKASI KEMATANGAN MENTIMUN BERDASARKAN TEKSTUR KULIT BUAH MENGGUNAKAN METODE EKSTRAKSI CIRI STATISTIK," *Jurnal Informatika*, vol. 9, no. 1, 2015, doi: 10.26555/jifo.v9i1.a2044.
- [6] I. G. A. N. Dananjaya, "Strategi Pemasaran Produk Stroberi pada UD Mitra Tani di Desa Pancasari Kabupaten Buleleng," *dwijenAgro*, vol. 4, no. 2, 2014.
- [7] P. B. Timotiwu, T. K. Manik, A. Agustiansyah, and E. Pramono, "FENOLOGI DAN PERTUMBUHAN TANAMAN STRAWBERRY DI DATARAN RENDAH SEBAGAI KAJIAN AWAL DAMPAK PERUBAHAN IKLIM TERHADAP PERTUMBUHAN DAN PRODUKSI TANAMAN," *JURNAL AGROTROPIKA*, vol. 20, no. 1, 2021, doi: 10.23960/ja.v20i1.4596.
- [8] L. Luthfi, R. Imam Muslem, D. Armiady, S. Sriwinar, R. Fajri, and I. Iqbal, "Analysis of CNN Method for Image Classification of Coconut Ripeness Levels," in *2023 Eighth International Conference on Informatics and Computing (ICIC)*, IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/ICIC60109.2023.10381964.
- [9] I. R. Muslem, "KLIK: Kajian Ilmiah Informatika dan Komputer Image Classification pada Kasus American Sign Language Menggunakan Support Vector Machine," *Media Online*, vol. 4, no. 2, pp. 1184–1191, 2023, doi: 10.30865/klik.v4i2.1242.
- [10] D. Armiady and I. M. R, "Klasifikasi Kualitas Buah Pisang Berdasarkan Citra Buah Menggunakan Stochastic Gradient Descent," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 2, 2023.
- [11] D. Armiady and I. Muslem R., "Penetapan Klaster Siswa Unggul Dengan Menggunakan Algoritma Roc-Smarter," *Jurnal TIKA*, vol. 7, no. 2, 2022, doi: 10.51179/tika.v7i2.1229.
- [12] I. R. Muslem and T. M. Johan, "KLIK: Kajian Ilmiah Informatika dan Komputer Klasifikasi Citra Ikan Menggunakan Algoritma Convolutional Neural Network dengan Arsitektur VGG-16," *Media Online*, vol. 4, no. 2, pp. 978–985, 2023, doi: 10.30865/klik.v4i2.1209.
- [13] N. Arifin, C. N. Insani, and M. R. Rasyid, "Klasifikasi Tingkat Kematangan Buah Tomat menggunakan Computer Vision untuk Smart Agriculture," *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)*, vol. 22, no. 2, 2023, doi: 10.53513/jis.v22i2.8387.
- [14] N. Astrianda, "Klasifikasi Kematangan Buah Tomat Dengan Variasi Model Warna Menggunakan Support Vector Machine," *VOCATECH: Vocational Education and Technology Journal*, vol. 1, no. 2, 2020, doi: 10.38038/vocatech.v1i2.27.
- [15] J. Jusrawati, A. Futri, and A. B. Kaswar, "Klasifikasi Tingkat Kematangan Buah Pisang Dalam Ruang Warna RGB Menggunakan Jaringan Syaraf Tiruan (JST)," *Journal of Embedded Systems, Security and Intelligent Systems*, vol. 2, no. 1, 2021, doi: 10.26858/jessi.v2i1.20327.
- [16] Ismail, Nurhikma Arifin, and Prihastinur, "KLASIFIKASI KEMATANGAN BUAH NAGA BERDASARKAN FITUR WARNA MENGGUNAKAN ALGORITMA MULTI-CLASS SUPPORT VECTOR MACHINE," *Jurnal Informatika Teknologi dan Sains*, vol. 5, no. 1, 2023, doi: 10.51401/jinteks.v5i1.2203.
- [17] S. P. Adenugraha, V. Arinal, and D. I. Mulyana, "Klasifikasi Kematangan Buah Pisang Ambon Menggunakan Metode KNN dan PCA Berdasarkan Citra RGB dan HSV," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 1, 2022, doi: 10.30865/mib.v6i1.3287.
- [18] M. F. Ajizi, D. Syauqy, and M. H. H. Ichsan, "Klasifikasi Kematangan Buah Pisang Berbasis Sensor Warna Dan Sensor Load Cell Menggunakan Metode Naive Bayes," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2019.
- [19] M. F. Naufal et al., "Analisis Perbandingan Algoritma Klasifikasi MLP dan CNN pada Dataset American

- Sign Language," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 3, 2021, doi: 10.29207/resti.v5i3.3009.
- [20] W. S. Sidabutar, J. T. Tarigan, and A. Sharif, "Automated Strawberry Ripeness Detection Using Convolution Neural Network," *Univers J Agric Res*, vol. 11, no. 6, 2023, doi: 10.13189/ujar.2023.110614.
- [21] I. K. Leo Puja Artana, D. M. Wiharta, and L. Linawati, "RANCANG BANGUN PERTANIAN STROBERI BERBASIS INTERNET OF THINGS," *Jurnal SPEKTRUM*, vol. 7, no. 4, 2020, doi: 10.24843/spektrum.2020.v07.i04.p1.