

MESIN PEMBELAJARAN ENSEMBLE UNTUK IDENTIFIKASI VARIETAS PADI

Ensemble Machine Learning for Rice Varieties Identification

Ariza Ikhlas, Abdullah, Dwi Yuli Prasetyo

Program Studi Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Islam Indragiri
Jl. Provinsi Parit 1 Tembilahan Hulu Indragiri Hilir Riau- Indonesia
Telp. (0768) 324918, Fax. (0768) 324918
E-mail: abdialam@yahoo.com

(Makalah diterima, 04 November 2018 – Disetujui, 07 Desember 2020)

ABSTRAK

Setiap varietas padi memiliki karakter tertentu dengan anjuran tanam berbeda. Petani umumnya kesulitan memilih varietas padi yang cocok untuk ditanam di lahan mereka karena kurangnya kemampuan identifikasi. Algoritma klasifikasi merupakan solusi mengatasi masalah ini karena mampu mengidentifikasi varietas padi melalui citra digital. Tujuan penelitian ini adalah menerapkan dan mengevaluasi beberapa algoritma klasifikasi untuk mengidentifikasi varietas padi menggunakan fitur warna dan tekstur. Penelitian dilakukan di kabupaten Indragiri Hilir Riau pada tahun 2018. Mesin pembelajaran dibangun dengan cara menggabungkan beberapa algoritma klasifikasi (*classifier*), yaitu Support Vector Machine, k-Nearest Neighbors, Logistic Regression, dan Decision Tree. Varietas yang diteliti adalah IR42, Inpara-9, dan Batang Piaman. Berdasarkan tingkat ketelitian masing-masing algoritma, k-Nearest Neighbors memberikan hasil lebih baik dibanding algoritma lainnya, baik dengan maupun tanpa normalisasi data. Terdapat enam sampel Inpara-9 yang diprediksi benar (*true positive*) dan lima sampel diprediksi salah (*false positive*). Pada varietas Batang Piaman terdapat delapan sampel yang diprediksi benar (*true positive*). Pada IR42 terdapat lima sampel yang diprediksi benar.

Kata kunci: padi, varietas, identifikasi, citra digital

ABSTRACT

Each rice variety has certain characteristics with different planting recommendations. Farmers generally find it difficult to select rice varieties suitable for planting on their land due to a lack of identification capabilities. The classification algorithm is a solution to this problem because it is able to identify rice varieties through digital images. The purpose of this study was to apply and evaluate several classification algorithms to identify rice varieties using color and texture features. The research was conducted in Indragiri Hilir Riau district in 2018. Machine learning was built by combining several classification algorithms (classifier), namely Support Vector Machine, k-Nearest Neighbors, Logistic Regression, and Decision Tree. The varieties studied were varieties IR42, Inpara-9, and Batang Piaman. Based on the level of accuracy of each algorithm, k-Nearest Neighbors gives better results than other algorithms, both with and without normalization data. There were six samples of Inpara-9 that were predicted to be true (true positive) and five samples were predicted to be false (false positive). In the Batang Piaman, there were eight samples that were predicted to be true (true positive). In IR42, there were five samples that were predicted to be true.

Key words: rice, varieties, identification, digital image

PENDAHULUAN

Padi (*Oryza sativa* L.) yang setelah diolah menjadi beras merupakan pangan utama sebagian populasi penduduk dunia. Beras mengandung vitamin, mineral, dan amino acid (Sadeghi *et al.*, 2013). Ada beberapa varietas unggul padi yang dikenal antara lain Batang Piaman, IR42, dan Inpara-9. Lama pertumbuhan tanaman padi berkisar antara 3-6 bulan, bergantung pada varietas dan lingkungan tumbuh. Lebih dari 90% produksi beras dunia dihasilkan pada agroekosistem lahan sawah irigasi. Padi mampu berproduksi pada kondisi lingkungan tertentu dan jika tidak sesuai menyebabkan kegagalan panen. Padi umumnya mampu tumbuh pada agroekosistem lahan sawah irigasi, lahan rawa pasang surut, dan lahan dengan lingkungan salinitas (Bouman, Buresh, Dobermann, & Fitzgerald, 2013).

Dalam budi daya padi, beberapa faktor yang perlu diperhatikan selain agroekosistem dan iklim adalah varietas. Di Indonesia terdapat berbagai jenis varietas padi seperti Inbrida Padi Sawah Irigasi (Inpari), Hibrida Padi (Hipa), Inbrida Padi Gogo (Inpago), dan Inbrida Padi Rawa (Inpara). Sebelum dibudidayakan, varietas-varietas padi tersebut perlu dikenal dan dipilih kesesuaiannya dengan kondisi lahan dan musim tanam. Kesalahan dalam pemilihan varietas dapat menyebabkan penurunan produksi dan bahkan gagal panen karena tidak sesuai dengan agroekosistem pengembangan dan musim tanam.

Identifikasi varietas padi umumnya berdasarkan deskripsi secara visual, namun teknik tersebut sulit mengenali varietas yang mempunyai morfologi benih serupa (Adnan *et al.*, 2015). Teknik yang berpotensi digunakan untuk identifikasi benih padi antara lain Simple Sequence Repeat (SSR) (Mulsanti *et al.*, 2013), khusus untuk membedakan tetua padi hibrida, varietas padi yang berbeda atau campuran varietas lain, bahkan pada benih yang tertular penyakit. Kelemahan teknik SSR adalah relatif mahal diterapkan dalam pengujian rutin.

Metode lain yang dapat digunakan adalah uji BUSS (baru, unik, seragam, dan stabil) (Sitaresmi *et al.*, 2013). Kelemahan uji BUSS adalah sebagian karakter penciri varietas bersifat kualitatif sehingga terbuka kemungkinan perbedaan penilaian secara subyektif. Identifikasi varietas juga bisa dilakukan dengan cara mendeskripsikan morfologi benih (Koutroubas *et al.*, 2004). Identifikasi varietas padi berdasarkan morfologi seperti ukuran gabah biasanya dilakukan secara manual menggunakan caliper sehingga memerlukan banyak waktu dan tidak efisien (Thind & Sogi, 2005).

Alternatif lain yang dapat digunakan adalah memanfaatkan teknologi komputer yang didesain khusus untuk mampu mengidentifikasi jenis varietas

menggunakan algoritma klasifikasi (*classifier*) berdasarkan fitur tertentu dengan citra digital. Dalam hal ini diperlukan pemilihan fitur yang tepat. Fitur digunakan untuk mengidentifikasi obyek dan mampu membedakan dengan obyek lain, sehingga pengelompokan obyek dapat dilakukan dengan baik dan memberikan akurasi yang baik pula. Oleh karena itu pemilihan fitur sangat menentukan keberhasilan pengenalan pola (Chen *et al.*, 2020)

Keberhasilan proses identifikasi citra digital juga ditentukan oleh pemilihan classifier yang tepat. Beberapa classifier yang sering digunakan dalam melakukan prediksi yaitu Decision Tree, k-Nearest Neighbors classifier (KNN), Logistic Regression, Support Vector machine. Meskipun sudah banyak algoritma klasifikasi yang diusulkan namun sulit menetapkan algoritma yang sesuai untuk tugas klasifikasi tertentu. Masing-masing algoritma memiliki kelebihan dan kekurangan. Ensemble learning machine merupakan metode yang menggabungkan beberapa algoritma klasifikasi menggunakan aturan tertentu sedemikian rupa sehingga dapat menggabungkan kekuatan masing-masing algoritma klasifikasi tunggal dan memiliki performa generalisasi yang lebih baik daripada klasifikasi tunggal (Raschka, 2015).

Tujuan penelitian adalah menerapkan dan mengevaluasi beberapa algoritma klasifikasi untuk mengidentifikasi varietas padi menggunakan fitur warna dan tekstur.

BAHAN DAN METODE

Penelitian dilakukan di kabupaten Indagiri Hilir Riau pada tahun 2018. Benih padi sampel diperoleh dari Dinas Pertanian Tanaman Pangan dan Holtikultura kabupaten Indragiri Hilir, Riau. Machine learning dibangun dengan cara menggabungkan beberapa algoritma klasifikasi (*classifier*). Algoritma yang digunakan adalah Support Vector Machine, K-Nearest Neighbors, Logistic Regression, dan Decision Tree. Benih padi yang diteliti yaitu dari varietas IR42, Inpara-9. dan Batang Piaman. Fitur yang digunakan adalah rata-rata warna (Red, Green, Blue - RGB), deviasi standar RGB, pemanfaatan fitur tekstur GLCM (Grey Level Co-occurrence Matrix) berdasarkan contrast, dissimilarity, homogeneity, energy, correlation, dan ASM.

Citra sampel diambil menggunakan kamera digital dan disimpan dengan format file dalam bentuk JPG. Peralatan yang digunakan adalah komputer Macbook Pro dengan spesifikasi OS (Operating System): MacOS Sierra version 10.12.2, MacBook Pro (15-inch, 2011), Processor 2.2 GHz Inter Core i7, Memory 4 GB 133 MHz DDR3, Graphic AMD Radeon HD 6750M 1024 MB, Intel HD Graphic 3000 384 MB.

Dalam penelitian digunakan sebuah kamera untuk pengambilan gambar sampel dengan spesifikasi EOS 7D dan lensa 18-55 mm (standard). Langkah penerapan dan evaluasi beberapa algoritma mesin pembelajaran dalam mengidentifikasi varietas padi mencakup ekstraksi fitur, normalisasi data, pengembangan sistem, serta implementasi dan evaluasi.

Ekstraksi Fitur

Pada tahapan ini dilakukan ekstraksi fitur atau ciri masing-masing benih padi melalui citranya. Terdapat 30 jenis fitur berupa RGB (rata-rata dan deviasi standar) (Thamizhvani *et al.*, 2018), GLCM (Grey Level Co-occurrence Matrix) (Sharma, Priyanka, Kalsh, & Saini, 2015) yang terdiri atas contrast, dissimilarity, homogeineity, ASM (Angular Single Moment), energy, dan correlation yang masing-masing diambil dari berbagai angle 00, 450,900, 1350.

Fitur pada citra yang telah diekstraksi dinyatakan dalam vektor fitur yang terbagi menjadi beberapa kelompok, sepertivektorfitur rata-rata warna dan deviasi standar warna pada masing-masing kanal RGB. Jika sejumlah piksel pada citra Q adalah *N* sedangkan *T* adalah transpose vektor, (*r_{std}*, *g_{std}*, *b_{std}*) masing-masing adalah rata-rata fitur pada kanal R, G, dan B pada citra Q, maka nilai rata-rata warna *x* dapat diformulasikan pada formula 1-4 berikut:

$$x = (r_{avg}, g_{avg}, b_{avg})^T \dots\dots\dots(1)$$

dimana

$$r_{avg} = \frac{\sum_1^N r(Q)}{N} \dots\dots\dots(2)$$

$$g_{avg} = \frac{\sum_1^N g(Q)}{N} \dots\dots\dots(3)$$

$$b_{avg} = \frac{\sum_1^N b(Q)}{N} \dots\dots\dots(4)$$

Vektor fitur pada nilai deviasi standar warna (*y*) untuk sejumlah piksel yang sama dapat dilihat pada formula 5-8.

$$y = (r_{std}, g_{std}, b_{std})^T \dots\dots\dots(5)$$

dimana

$$r_{std} = \sqrt{\frac{\sum_1^N (r(Q) - r_{avg})^2}{N}} \dots\dots\dots(6)$$

$$g_{std} = \sqrt{\frac{\sum_1^N (g(Q) - g_{avg})^2}{N}} \dots\dots\dots(7)$$

$$b_{std} = \sqrt{\frac{\sum_1^N (b(Q) - b_{avg})^2}{N}} \dots\dots\dots(8)$$

Vektor fitur Grey Level Co-occurrence Matrix (Hall-Beyer, 2017) menggunakan formula 9-14 yang merupakan rumus untuk mengalkulasi tekstur, dimana *P_(i,j)* adalah nilai probabilitas pada sel matrik *co-occurrence matrix* untuk baris ke-*i* dan kolom ke-*j*.

$$contrast = \sum_{i,j=0}^{N-1} P_{i,j} (i - j)^2 \dots\dots\dots(9)$$

$$Dissimilarity = \sum_{i,j=0}^{N-1} P_{i,j} |i - j| \dots\dots\dots(10)$$

$$Homogeineity = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1 + (i-j)^2} \dots\dots\dots(11)$$

$$ASM = \sum_{i,j=0}^{N-1} P_{i,j}^2 \dots\dots\dots(12)$$

$$Energy = \sqrt{ASM} \dots\dots\dots(13)$$

$$Correlation = \sum_{i,j=0}^{N-1} P_{i,j} \left[\frac{(i - \mu_i)(j - \mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right] \dots\dots\dots(14)$$

Normalisasi Data

Pada tahapan ini dilakukan penyeragaman skala ukuran fitur karena terkadang memiliki skala pengukuran yang berbeda. Oleh karena itu, data perlu dinormalisasi. Min-max normalisasi adalah proses pada pengambil pengukuran data pada satuan mesin dan mentransformasikan nilai antara 0,0 hingga 1,0 dimana nilai terkecil (min) menjadi 0,0 dan nilai tertinggi (max) 1,0 (Raschka, 2015). Cara ini memudahkan cara untuk membandingkan nilai yang diukur menggunakan skala atau perbedaan pada pengukuran satuan pada pengukuran. Nilai normalisasi didefinisikan sebagai formula 15.

$$MM(X_{ij}) = \frac{X_{ij} - X_{min}}{X_{max} - X_{min}} \dots\dots\dots(15)$$

Pengembangan Sistem

Pada tahapan ini dilakukan identifikasi komponen dan proses yang akan berinteraksi dalam sistem. Selanjutnya merancang sistem yang diberikan dalam bentuk blok diagram, yang menunjukkan komponen-komponen dan proses-proses berinteraksi.

Implementasi dan Evaluasi

Pada tahap implementasi digunakan *tool development* untuk pengembangan aplikasi sebagai berikut: Phyton, Numpy, Matplotlib, Scikit-learn, Scikit-images, Pandas, PyQT, dan SQLite. Pada tahap evaluasi dilakukan pengujian masing-masing algoritma Support Vector Machine, K-Nearest Neighbors, Decision Tree, Logistic Regression dan Ensemble Majority Vote menggunakan Cross Validation (Varoquaux, Raamana, Engemann, Hoyos-idrobo, AndrésSchwartz, & Thirion, 2017), confusion matrix, Precision, Recall, F1-Score (Kirimu & Moturi, 2016). Pada k-Fold Cross Validation, data

pelatihan dibagi secara acak ke dalam k-Folds tanpa menggantikan, dimana k-1 folds digunakan untuk model pelatihan dan satu fold digunakan untuk pengujian. Prosedur ini dilakukan berulang sebanyak *k* kali sehingga diperoleh model dan perkiraan kinerja (Raschka, 2015).

Confusion matrix adalah matriks per segi sederhana yang melaporkan perhitungan true positive (TP), true negative (TN), false positive (FP), dan false negative (FN) untuk memprediksi classifier (Thomas *et al.*, 2018), sebagaimana ditunjukkan pada Gambar 1.

Error (ERR) dan accuracy (ACC) merupakan informasi tentang sampel yang tidak terprediksi (*miss classified*). Kesalahan dapat dihitung dengan menjumlahkan semua kesalahan prediksi dibagi dengan total prediksi, dan akurasi dikalkulasikan dengan menghitung jumlah kebenaran dalam prediksi dibagi dengan total jumlah prediksi, formula 16 adalah error dan formula 17 adalah akurasinya:

$$ERR = \frac{FP+FN}{FP+FN+TP+TN} \dots\dots\dots(16)$$

$$ACC = \frac{TP+TN}{FP+FN+TP+TN} = 1 - ERR \dots\dots\dots(17)$$

True Positive rate (TPR) dihitung menggunakan formula 18 dan Positive predictive value (PPV) menggunakan formula 19, yaitu matriks performa berhubungan dengan true positive dan true negative. Dalam fakta, recall adalah sinonim true positive.

$$TPR = \frac{TP}{P} = \frac{TP}{FN+TP} \dots\dots\dots(18)$$

$$PPV = \frac{TP}{TP+FP} \dots\dots\dots(19)$$

Precision (PRE) dihitung menggunakan formula 20 dan recall (REC) dihitung dengan formula 21, yaitu matrix performa yang berhubungan dengan true positive dan true negative. Dalam fakta, recall adalah sinonim true positive.

$$PRE = PPV = \frac{TP}{TP+FP} \dots\dots\dots(20)$$

$$REC = TPR = \frac{TP}{FN+TP} \dots\dots\dots(21)$$

Praktek kombinasi precision dan recall digunakan untuk pemanggilan F1-score dengan formula 22.

$$F1 = 2 \frac{PRE \times REC}{PRE+REC} \dots\dots\dots(22)$$

Pada tahap akhir diberikan kesimpulan hasil pengujian terhadap sampel yang mendeskripsikan tingkat akurasi dan sebagainya.

HASIL DAN PEMBAHASAN

Akuisisi Data

Pada tahap akuisisi data berhasil dikumpulkan sebanyak 60 sampel berupa citra (images) benih padi dari ketiga varietas. Masing-masing adalah sebagai berikut: varietas IR42 sebanyak 22 sampel, varietas Inpara-9 sebanyak 22 sampel dan varietas Batang Piaman sebanyak 16 sampel.

Normalisasi

Pada tahap normalisasi berhasil dilakukan penyeragaman skala ukuran berkisar antara 0-1 untuk setiap fitur. Tabel 1 menunjukkan contoh normalisasi data. Tidak semua nilai fitur ditampilkan pada tabel, namun mewakili fitur-fitur lain yang ada. Semua nilai fitur berada pada kisaran 0-1. Dengan demikian, semua nilai fitur berada dalam skala yang sama.

Pengembangan Sistem

Pada tahap ini berhasil dibangun rancangan sistem. Gambar 2 adalah tahapan pengembangan sistem, yang terdiri atas tahap awal hingga dilakukan prediksi. Tahap awal dimulai dari penangkapan citra atau image. Setelah itu dilakukan ekstraksi fitur. Sebagian data hasil proses ekstraksi digunakan untuk data training atau data sampel.

		Predicted class	
		P	N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

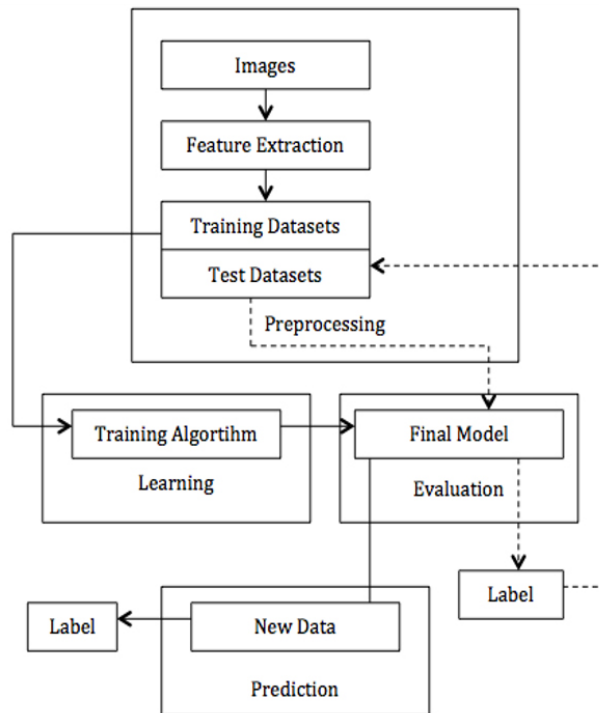
Gambar 1. Confusion matrix pengukur kinerja algoritma

Tabel 1. Normalisasi data

	meanRed	meanGreen	meanBlue	stdevRed	stdevGreen	
1	0.71212	0.24483	0.36556	0.71585	0.60219	0.
2	0.82044	0.33398	0.20644	0.69999	0.37566	0.
3	0.86776	0.18352	0.20249	0.78333	0.14912	0.
4	0.87308	0.07329	0.23434	0.87787	0.13935	0.
5	0.71122	0.18410	0.38758	0.70890	0.24059	0.
6	0.71088	0.03984	0.43790	0.86667	0.45808	0.
7	0.92142	0.36105	0.07756	0.74650	0.28704	0.
8	0.76283	0.23864	0.30762	0.66076	0.11005	0.
9	0.92306	0.16921	0.14200	1.00000	0.65507	0.
10	0.99242	0.10571	0.16700	0.96400	0.50000	0.

If you want to use normalization, please click checkbox ! Normalization

Cancel Next >>



Gambar 2. Blok diagram pembangunan sistem

Sebagian data digunakan sebagai data testing. Model classifier dibentuk berdasarkan hasil algoritma training. Proses klasifikasi menggunakan model classifier yang dipilih. Selanjutnya sistem sudah dapat melakukan prediksi label pada data baru yang belum diketahui kategorinya.

Proses training menggunakan training dataset setelah dilakukan ekstraksi dan disimpan dalam dataset pelatihan. Dataset yang telah disimpan digunakan untuk training algoritma agar machine learning memiliki kemampuan

mengenali pola pada citra. Percobaan dilakukan menggunakan beberapa algoritma terpilih yaitu Support Vector Machine dan K-Nearest Neighbors (Reddy, Prasad, & Mounika, 2017), Decision Tree (Wang, Yu, Liu, & Li, 2018), Logistic Regression (Raschka, 2018), dan Ensemble Majority Vote (Pramesti, Herdiyeni, & Nugroho, 2018) yang merupakan kombinasi empat algoritma sebelumnya.

Sebelum menetapkan model final untuk penentuan model algoritma yang baik, dilakukan pengujian

menggunakan Gridsearch Cross Validation untuk mencari parameter terbaik. Parameter yang digunakan ditunjukkan pada Tabel 2. Pada masing-masing algoritma dilakukan setting parameter untuk memberikan nilai prediksi terbaik.

Setelah menentukan parameter, kemudian dilakukan pengujian algoritma pada data set yang telah terlabel menggunakan cross validation, confusion matrix, precision, recall, dan F1-score. Pada Gambar 3 ditunjukkan hasil pengujian menggunakan cross validation dengan dan tanpa normalisasi.

Pengujian menunjukkan support vector machine, K-nearest neighbors, decision tree, logistic regression, dan ensemble majority vote memberikan hasil berturut-turut 75%, 77%, 75%, 75%, dan 72%. Data yang telah dinormalisasi memberikan hasil masing-masing 80%, 85%, 77%, 82%, 82%. Pada setiap jenis algoritma klasifikasi terjadi peningkatan akurasi dengan adanya proses normalisasi. K-neraest neihbors memberikan hasil yang lebih baik dibanding algoritma lainnya, baik pada data tanpa normalisasi maupun yang telah dinormalisasi.

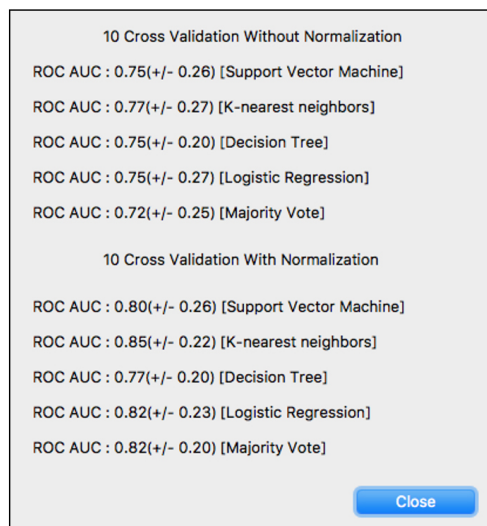
Namun peningkatan akurasi yang terbesar terjadi pada ensemble learning menggunakan aturan majority vote, yaitu 10% dari 0,72 menjadi 0,82.

Pada Gambar 4 dapat dilihat salah satu keluaran sistem berupa confusion matrix. Setelah memasukan data yang belum diketahui untuk dilakukan prediksi diberikan output berupa hasil prediksi atau informasi data yang belum terlabel tersebut.

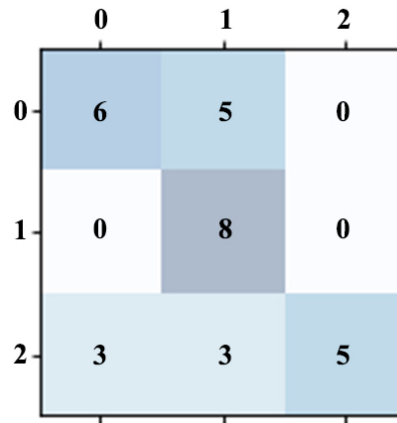
Hasil prediksi menggunakan confusion matrix, dimana skala 0 adalah varietas Inpari-9, skala 1 varietas Batang Piaman, dan skala 2 adalah IR42. Vertikal adalah aktual dan horisontal adalah hasil prediksi. Pengujian ini menggunakan 30 dari 60 sampel, dimana 50% data tidak dinormalisasi diambil secara acak yang digunakan untuk pelatihan dan 50% digunakan untuk pengujian. Dalam pengujian ini memberikan enam sampel benih padi Inpara-9 yang diprediksi benar (*true positive*) dan lima sampel diprediksi salah (*false positive*). Pada benih padi varietas Batang Piaman terdapat delapan sampel yang diprediksi benar (*true positive*). Pada benih IR42 terdapat lima sampel yang diprediksi benar.

Tabel 2. Parameter algoritma

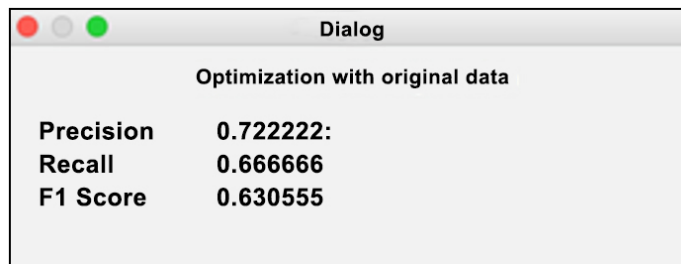
Algoritma	Parameter
<i>Support Vector Machine</i>	$C = 0.1$ $Kernel = Linear$
<i>K-Nearest Neighbors</i>	$Algorithm = auto$ $n_neighbors = 7$
<i>Decision Tree</i>	$Max_depth = 5$ $Random_state = 5$
<i>Logistic Regression</i>	$C = 10.0$ $Penalty = L2$ $Random_state = 0$
<i>Majority Vote</i>	$Estimator = [svm, knn, tree, lr]$ $Voting = hard$



Gambar 3. Hasil pengujian cross validation



Gambar 4. Hasil *confusionmatrix* untuk data tanpa normalisasi



Gambar 5. Precision, recall, dan F1-score

Pada Gambar 5 dapat dilihat salah satu hasil pengujian prediksi berdasarkan precision, recall, dan F1-score menggunakan mesin pembelajaran. Hasil yang didapatkan tanpa normalisasi data masing-masing 72% , 66%, dan 63% untuk precision, recall, dan F1-score.

KESIMPULAN

Penerapan dan evaluasi beberapa algoritma klasifikasi untuk mengidentifikasi varietas padi menggunakan fitur warna dan tekstur, telah dilakukan. Terdapat enam sampel benih padi Inpara-9 yang diprediksi benar (*true positive*) dan lima sampel diprediksi salah (*false positive*). Pada benih padi varietas Batang Piaman terdapat delapan sampel yang diprediksi benar (*true positive*). Pada benih IR42 terdapat lima sampel yang diprediksi benar. Normalisasi data berhasil meningkatkan kinerja semua algoritma. Peningkatan signifikan terjadi pada pembelajaran ensemble menggunakan majority vote.

UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada Dinas Pertanian Tanaman Pangan dan Holikultura Kabupaten Indragiri Hilir, Riau, yang telah memberikan data yang diperlukan sehingga penelitian ini dapat terselenggara.

DAFTAR PUSTAKA

- Adnan, Widiastuti, M. L., & Wahyuni, S. (2015). Identifikasi Varietas Padi Menggunakan Pengolahan Citra Digital dan Analisis Diskriminan Identification of Rice Variety using Image Processing and Discriminant Analyses. *Penelitian Pertanian Tanaman Pangan*, 34(2), 89–96.
- Bouman, B., Buresh, R., Dobermann, A., & Fitzgerald, M. (2013). *Rice Almanac*. Metro Manila: Global Rice Science Partership.
- Chen, R. C., Dewi, C., Huang, S. W., & Caraka, R. E. (2020). Selecting critical features for data classification based on machine learning methods. *Journal of Big Data*, 7(52). <https://doi.org/10.1186/s40537-020-00327-4>
- Hall-Beyer, M. (2017). *GLCM Texture: A Tutorial v. 3.0 March 2017*. <https://doi.org/10.13140/RG.2.2.12424.21767>
- Kirimi, J. M., & Moturi, C. A. (2016). Application of Data Mining Classification in Employee Performance Prediction. *International Journal of Computer Applications*, 146(7), 28–35. <https://doi.org/10.5120/ijca2016910883>
- Koutroubas, S. D., Mazzini, F., Pons, B., & Ntanos, D. A. (2004). Grain quality variation and relationships with morpho-physiological traits in rice (*Oryza sativa* L.) genetic resources in Europe. *Field Crops Research*, 86(2), 115–130. [https://doi.org/https://doi.org/10.1016/S0378-4290\(03\)00117-5](https://doi.org/https://doi.org/10.1016/S0378-4290(03)00117-5)

- Mulsanti, I. W., Surahman, M., Wahyuni, S., & Utami, D. W. (2013). Identifikasi Galur Tetua Padi Hibrida dengan Marka SSR Spesifik dan Pemanfaatannya dalam Uji Kemurnian Benih. *Penelitian Pertanian Tanaman Pangan*, 32(1), 1–8.
- Pramesti, R. P. A., Herdiyeni, Y., & Nugroho, A. S. (2018). Weighted Ensemble Classifier for Plant Leaf Identification. *TELKOMNIKA*, 16(3), 1386–1393. <https://doi.org/10.12928/TELKOMNIKA.v16i3.7615>
- Raschka, S. (2015). *Python Machine Learning: Unlock Deeper Insights Into Machine Learning with this Vital Guide to Cutting-edge Predictive Analytics*. Packt Publishing. Retrieved from <https://books.google.co.id/books?id=HuxuawEACAAJ>
- Raschka, S. (2018). *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*. Retrieved from <http://arxiv.org/abs/1811.12808>
- Reddy, N. C. S., Prasad, K. S., & Mounika, A. (2017). Classification Algorithms on Datamining : A Study. *International Journal of Computer Intelligence Research*, 13(8), 2135–2142.
- Sadeghi, M., Nasrnia, E., Masoumi, A., & Hemmat, A. (2013). Head rice yield response to low and high drying and tempering conditions. *International Agrophysics*, 27, 219–223. <https://doi.org/10.2478/v10247-012-0088-5>
- Sharma, E. K., Priyanka, E., Kalsh, E. A., & Saini, E. K. (2015). GLCM and its Features. *International Journal of Advanced Researc in Electronics and Communication Engineering (IJARECE)*, 4(8), 2180–2182.
- Sitairesmi, T., Yunani, N., Mulsanti, I. W., Utomo, S. T. W., & Daradjat, A. . (2013). Identifikasi Varietas Contoh untuk Karakter Penciri Spesifik sebagai Penunjang Harmonisasi Pengujian BUSS Padi. *Penelitian Pertanian Tanaman Pangan*, 32(3), 148–158.
- Thamizhvani, T. R., Hemalatha, R. J., Babu, B., & Dhivya, A. J. A. (2018). Identification of Skin Tumours using Statistical and Histogram Based Features. *Journal of Clinical and Diagnosis Research*, 12(9), 11–15. <https://doi.org/10.7860/JCDR/2018/36258.12040>
- Thind, G. K., & Sogi, D. S. (2005). Identification of coarse (IR-8), fine (PR-106) and superfine (Basmati-386) rice cultivars. *Food Chemistry*, 91, 227–233. <https://doi.org/10.1016/j.foodchem.2003.10.015>
- Thomas, P., Bril El Haouzi, H., Suhner, M. C., Thomas, A., Zimmermann, E., & Noyel, M. (2018). Using a Classifier Ensemble for Proactive Quality Monitoring and Control. *Computers in Industry*, 99, 193–204.
- Varoquaux, G., Raamana, P. R., Engemann, D., Hoyos-idrobo, AndrésSchwartz, Y., & Thirion, B. (2017). Assessing and tuning brain decoders : cross-validation , caveats , and guidelines. *Neuro Image*, 166–179.
- Wang, L., Yu, Y., Liu, J., & Li, Q. (2018). Region compatibility based stability assessment for decision trees. *Expert Systems with Applications*.