



Optimalisasi Model *Ensemble Learning* dengan Augmentasi dan SMOTE pada Sistem Pendeteksi Kualitas Buah

Syahroni Hidayat ^{1*}, Taofan Ali Achmadi ², Hanif Ardhihsyah ³, Hanif Hidayat ⁴, Rian Febriyanto ¹, Abdulloh ¹, dan Intan Ermawati ⁵

- ¹ Prodi Teknik Elektro, Fakultas Teknik, Universitas Negeri Semarang
- ² Prodi Pendidikan Kesejahteraan Keluarga, Fakultas Teknik, Universitas Negeri Semarang
- ³ Prodi Teknik Kimia, Fakultas Teknik, Universitas Negeri Semarang
- ⁴ Prodi Pendidikan Teknik Otomotif, Fakultas Teknik, Universitas Negeri Semarang
- ⁵ Prodi Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik, Universitas Negeri Semarang

* Korespondensi: syahronihidayat@mail.unnes.ac.id

Sitasi: Hidayat, S.; Achmadi, T. A.; 2, Ardhihsyah, H.; Hidayat, H.; Febriyanto, R.; Abdulloh; dan Ermawati, I. (2024). Optimalisasi Model *Ensemble Learning* dengan Augmentasi dan SMOTE pada Sistem Pendeteksi Kualitas Buah. JTIM: Jurnal Teknologi Informasi Dan Multimedia, 6(1), 27-36 <https://doi.org/10.35746/jtim.v6i1.406>

Abstract: Fruit quality is an important factor in selecting fruit for consumption because it affects consumer health and satisfaction. Identification of fruit quality has become the focus of research, and one of the approaches used is a non-destructive approach through measuring the gases produced by the fruit. Machine learning can be used to process this gas data and build system models that can classify fruit quality. This research discusses the application of the DCS-OLA and Stacking dynamic ensemble learning algorithms to build a fruit quality detection system model. The basic methods used to build models are Logistic Regression, Decision Tree, Gaussian Naïve Bayes, and Multi-Layer Perceptron. The fruit used is mango with a shelf life of 7 days and Srikaya (sugar apple) with a shelf life of 4 days. The condition of the initial dataset is unbalanced. The research results show that trimming the mango dataset to only 4 days according to the shelf life of sugar apple helps reduce the difference in shelf life between the two. Then jittering and balancing techniques are used to increase and balance the number of datasets between the two types of fruit. High accuracy is achieved by the DCS-OLA ensemble and stacking ensemble by combining the basic methods of Logistic Regression and Decision Tree, especially in balanced dataset conditions. In conclusion, the use of ensemble learning in detecting fruit quality has great potential for real-world applications. However, further validation is needed with larger datasets and a wider variety of conditions.



Copyright: © 2024 oleh para penulis. Karya ini dilisensikan di bawah Creative Commons Attribution-ShareAlike 4.0 International License. (<https://creativecommons.org/licenses/by-sa/4.0/>).

Keywords: Fruit Quality, Machine Learning, Ensemble Learning, Augmentation, SMOTE

Abstrak: Kualitas buah merupakan faktor penting dalam pemilihan buah untuk dikonsumsi, karena berpengaruh pada kesehatan dan kepuasan konsumen. Identifikasi kualitas buah telah menjadi fokus penelitian, dan salah satu pendekatan yang digunakan adalah pendekatan non-destruktif melalui pengukuran gas yang dihasilkan oleh buah. Pembelajaran mesin dapat digunakan untuk mengolah data gas ini dan membangun model sistem yang dapat mengklasifikasikan kualitas buah. Penelitian ini membahas penerapan algoritma ensemble learning dynamic DCS-OLA dan Stacking untuk membangun model sistem pendeteksi kualitas buah. Metode dasar yang digunakan untuk membangun model adalah Logistic Regresion, Decision Tree, Gaussian Naïve Bayes, dan Multi-Layer Perceptron. Buah yang digunakan adalah buah mangga dengan masa simpan 7 hari dan srikaya dengan masa simpan 4 hari. Kondisi dataset awal tidak seimbang. Hasil penelitian menunjukkan bahwa pemangkasan dataset mangga menjadi hanya 4 hari sesuai masa simpan srikaya membantu mengurangi perbedaan lama waktu simpan antara keduanya. Kemudian teknik jittering dan balancing digunakan untuk menambah dan menyeimbangkan jumlah dataset antara kedua jenis buah. Akurasi yang tinggi dicapai oleh ensemble DCS-OLA dan ensemble stacking dengan

mengkombinasikan metode dasar Logistic Resregion dan Decision Tree, terutama pada kondisi dataset yang seimbang. Kesimpulannya, penggunaan ensemble learning dalam mendeteksi kualitas buah memiliki potensi besar untuk aplikasi di dunia nyata. Namun, validasi lebih lanjut diperlukan dengan dataset yang lebih besar dan variasi kondisi yang lebih luas.

Kata kunci: Kualitas Buah, Pembelajaran Mesin, Ensemble learning, Augmentasi, SMOTE

1. Pendahuluan

Kualitas buah merupakan hal yang sangat penting bagi konsumen dalam memilih buah yang akan dikonsumsi [1]. Kualitas buah yang baik dapat menjamin keamanan dan kelezatan saat dikonsumsi, sehingga dapat memberikan dampak positif pada kesehatan dan kepuasan konsumen [2]. Oleh karena itu, mengetahui kualitas buah di pasaran menjadi sangat penting bagi konsumen. Banyak penelitian telah dilakukan untuk mengidentifikasi faktor-faktor yang mempengaruhi kualitas buah, seperti ukuran, bentuk, warna, tekstur, aroma, rasa, dan kesegaran [3]. Dalam industri pertanian, penilaian kualitas buah secara non-destruktif sangat penting dalam memastikan bahwa buah yang dipasok ke pasar memiliki kualitas yang optimal [4], [5]. Salah satu variabel yang dapat digunakan untuk mendeteksi kualitas buah secara non-destruktif adalah melalui pengukuran gas yang dihasilkan oleh buah. Gas yang dihasilkan oleh buah dapat memberikan informasi tentang kondisi buah, seperti tingkat kematangan, kebusukan, atau kekerasan [6]. Pembelajaran mesin dapat membantu dalam memproses data dari sensor gas dan menghasilkan model pembelajaran mesin yang dapat mengklasifikasikan kualitas buah berdasarkan gas yang dihasilkan [7].

Beberapa penelitian telah dilakukan untuk mengembangkan teknologi pembelajaran mesin dalam mendeteksi kualitas buah secara non-destruktif menggunakan gas yang dihasilkan oleh buah [8]–[11]. Pembelajaran mesin dapat digunakan untuk mengklasifikasikan kualitas buah apel dan tomat berdasarkan gas yang dihasilkan oleh buah tersebut. Sensor gas digunakan untuk mengambil data gas yang dihasilkan oleh buah kemudian dilatih menggunakan algoritma pembelajaran mesin untuk mengklasifikasikan dan memprediksi kualitas buah tersebut [8]–[11]. Saat ini, penggunaan metode *ensemble learning* dalam teknologi pembelajaran mesin telah menjadi fokus penelitian karena dapat meningkatkan akurasi dan keandalan model pembelajaran mesin [12]. Model *ensemble learning* merupakan model yang dibangun dari beberapa algoritma pembelajaran mesin dasar/basis diantaranya *Decision Tree*, *Support Vector Machine*, Jaringan Syaraf Tiruan, *Logistic Regression*, dan *k-NN* [7], [13]–[15] baik dengan metode *stacking* ataupun *dynamic*. Dalam penelitian terbaru, metode *ensemble learning* telah diterapkan untuk meningkatkan akurasi dalam mengklasifikasikan kualitas buah tomat berdasarkan gas yang dihasilkan [16], [17]. Penelitian lain menunjukkan bahwa metode *ensemble learning* juga dapat diterapkan dalam memprediksi kualitas buah apel berdasarkan gas yang dihasilkan [17].

Akan tetapi, pada kondisi nyata buah-buahan, terutama buah klimaterik, memiliki jangka waktu pematangan dan pembusukan yang berbeda-beda [18]. Kondisi ini berdampak terhadap proses perekaman data gas buah yang digunakan sebagai dataset. Akibatnya dapat terjadi keterbatasan dan ketidakseimbangan (*imbalance*) dataset yang akan digunakan untuk melatih sistem. Oleh karena itu, teknik rekayasa fitur dataset yang tepat perlu dilakukan untuk mendapatkan fitur yang sesuai dengan tujuan pembentukan sistem. Metode yang dapat digunakan yaitu dengan menerapkan augmentasi dan rekayasa penyeimbangan (*balancing*) dataset. Metode *jittering* adalah salah satu metode augmentasi sederhana yang cocok diterapkan pada dataset deret waktu (*time series*) dengan memberikan noise pada dataset yang dimiliki [19], [20]. Sedangkan metode *balancing* adalah metode untuk menyeimbangkan dataset sehingga memiliki jumlah target yang seimbang.

Salah satu metode balancing yang banyak digunakan adalah *Synthetic Minority Over-sampling Technique* (SMOTE). Metode SMOTE menghasilkan data sintetik dengan menerapkan interpolasi linier antara titik kelas minoritas dan salah satu K tetangga terdekatnya. [21].

Penelitian ini bertujuan untuk membangun sistem yang dapat mendeteksi kualitas buah-buahan dengan menerapkan algoritma ensemble learning dynamic dan stacking untuk dataset terbatas dan imbalance. Evaluasi akurasi akan diterapkan pada model ensemble learning dengan kombinasi jumlah model dasar yang digunakan dan penerapan rekayasa augmentasi dan balancing dataset dengan metode jittering dan SMOTE secara berturut-turut.

2. Bahan dan Metode

Penelitian ini secara garis besar dilakukan dengan mengikuti jalan penelitian seperti ditunjukkan pada Gambar 1. Penelitian dimulai dengan melakukan studi literatur yang relevan tentang sistem untuk mendeteksi kualitas buah-buahan, alat, dan algoritma yang telah digunakan beserta kelebihan dan kekurangannya. Setelah melakukan studi literatur dilakukan rancang bangun alat deteksi kualitas buah menggunakan sensor gas. Alat yang dibangun kemudian digunakan untuk mengakuisisi data gas yang diproduksi oleh buah.



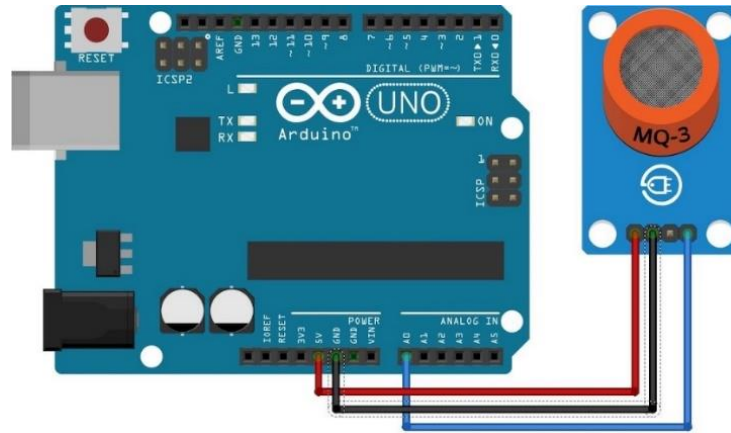
Gambar 1. Jalan Penelitian

Data yang diperoleh dari proses akuisisi kemudian ditingkatkan kualitasnya. Peningkatan kualitas data meliputi penghilangan sinyal gangguan (noise) pada sinyal dan normalisasi data. Data yang telah ditingkatkan kualitasnya kemudian digunakan pada proses pengembangan sistem deteksi kualitas buah. Sistem yang dikembangkan menggunakan bahasa pemrograman python. Sistem yang dibangun mengakomodasi seluruh kebutuhan penelitian berupa pembelajaran dengan metode *ensemble*.

2.1. Rancang Bangun Alat Akuisisi Data Gas

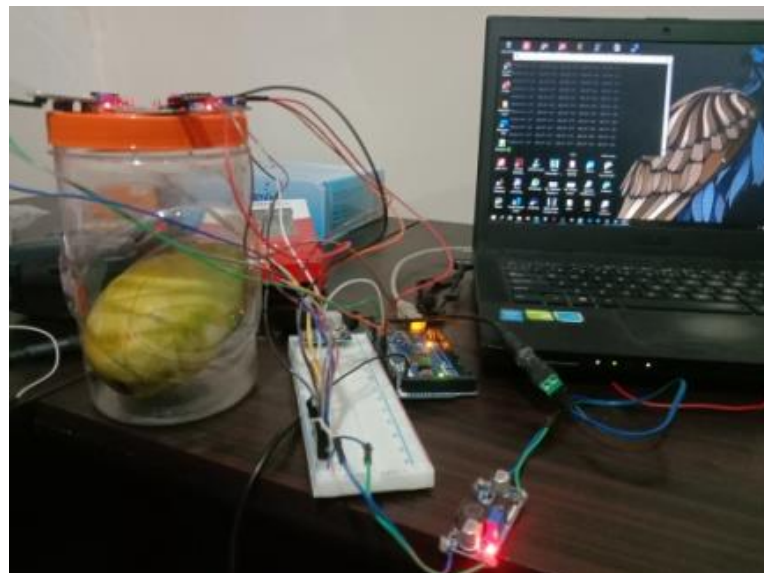
Pada penelitian ini digunakan alat TGS 2600 dengan range kerja 1-30 ppm, MQ3 dengan range kerja 25-500 ppm, dan MQ8 dengan range kerja 100~1000 ppm. Jumlah setiap sensor masing-masing satu buah. Ketiga sensor dihubungkan dengan Arduino Mega dan sumber tegangan luar sebesar 5 V. Skema umum rangkaian yang digunakan

seperti ditunjukkan pada gambar 4. Setiap pin analog sensor dihubungkan dengan pin analog Arduino, kemudian pin VCC dan Gnd setiap sensor dihubungkan dengan sumber tegangan 5-volt dc.



Gambar 2. Skema umum rangkaian alat penelitian

Untuk mengukur gas yang dihasilkan oleh buah mangga dan srikaya dibutuhkan alat perangkat gas universal. Alat perangkat gas di desain dapat buka tutup/lepas pasang dengan posisi sensor di letakkan pada bagian penutupnya. Dengan demikian dapat mempermudah dalam proses pengukuran gas. Peralatan yang disebutkan pada bagian sebelumnya kemudian dirangkai mengikuti skema alat yang ditunjukkan pada gambar 2 dan digabungkan dengan alat perangkat gas. Bentuk jadi alat yang digunakan setelah dirangkai ditunjukkan pada Gambar 3.



Gambar 3. Alat penelitian dengan tiga sensor dan perangkat gas universal

2.2. Akuisisi Data Gas Buah

Pada proses pengambilan data gas ini terdapat dua jenis data yang diukur yaitu data set poin yang diukur pada kondisi udara bersih dan data gas yang dihasilkan oleh sampel buah mangga. Pada proses pengukuran ini digunakan sebuah komputer untuk menjalankan program pengambilan data dan menyimpan hasilnya. Software yang digunakan adalah Arduino IDE untuk membangun koding program. Adapun pada proses recording

data menggunakan aplikasi PuTTY dan disimpan dengan format *.csv. Pengukuran gas untuk set poin dan dataset dilakukan selama kurang lebih 5 menit (± 300 detik). Proses pengukuran data set poin dilakukan saat wadah perangkat masih kosong. Hasil pengukuran set poin ditampilkan pada Tabel 1. Dari hasil pengukuran ini diambil nilai rata-rata ADC hasil pengukuran selama satu menit, dengan kondisi steady state untuk dijadikan sebagai set poin. Nilai set poin ini akan digunakan sebagai nilai referensi jika pada proses pengukuran sampel terdapat perubahan nilai dasar udara bersih yang terukur.

Tabel 1. Nilai set poin array sensor gas.

Variabel	Nama Sensor		
	Tgs 2600	MQ3	MQ8
Nilai rata-rata ADC terukur	70	120	95

Setelah melakukan pengukuran data set poin dilanjutkan dengan proses pengukuran gas sampel buah mangga dan srikaya. Jumlah buah mangga yang digunakan adalah 40 buah dan srikaya 3 buah dengan kondisi penyimpanan di dalam ruang terbuka. Buah mangga dan srikaya yang digunakan dengan kondisi awal mengkal. Penyimpanan dilakukan selama 7 hari untuk buah mangga dan 4 hari untuk buah srikaya dan dilakukan pengukuran produksi gas setiap harinya setelah 24 jam. Pada hari terakhir dilakukan uji persepsi untuk menentukan kondisi buah normal (1) atau rusak (0). Data yang terukur pada proses perekaman data ini berupa data ADC. Data raw hasil pengukuran ditunjukkan pada Tabel 2.

Tabel 2. Data raw ADC dari pengukuran gas untuk sampel buah mangga

Waktu (detik)	PuTTY log		
1	222	MQ-8= 85	
2		8= 84	
3	TGS2600= 95	MQ-3= 310	MQ-8= 84
4	TGS2600= 95	MQ-3= 311	MQ-8= 85
5		8= 84	
6	TGS2600= 95	MQ-3= 310	MQ-8= 84
7	TGS2600= 95	MQ-3= 311	MQ-8= 85
8	TGS2600= 95	MQ-3= 313	MQ-8= 88
.		.	
.		.	
.		.	
305	TGS2600= 116	MQ-3= 345	MQ-8= 131
306	TGS2600= 0	MQ-3= 0	MQ-8= 0
307	TGS2600= 0	MQ-3= 0	MQ-8= 0
308	TGS2600= 0	MQ-3= 0	MQ-8= 0
309	TGS2600= 0	MQ-3= 0	MQ-8= 0

2.3. Rekayasa Dataset

Data raw ADC terukur yang diperoleh terdapat data yang tidak dibutuhkan yaitu data yang terbaca terpotong atau bernilai nol semua. Kondisi ini mengharuskan adanya proses pembersihan (cleaning) data sebelum digunakan pada proses berikutnya. Setelah data dibersihkan kemudian dikonversi untuk mendapatkan nilai ppm gas dari nilai ADC terukur dengan persamaan (1) – (4) [22]. Nilai ADC yang akan dikonversi pertama-pertama dinormalisasi. Tujuannya untuk membuat nilai awal setiap sensor yang terukur bernilai 0. Normalisasi dilakukan dengan mengurangi seluruh nilai ADC terukur dengan nilai set poin pada tabel 1 menggunakan persamaan (1), selanjutnya hasilnya dikurangi oleh nilai minimum untuk setiap hasil pengukuran menggunakan persamaan (2).

$$\text{deltaADC} = \text{ADC}_{\text{terukur}} - \text{set_poin}_{\text{sensor}}, \quad (1)$$

$$\Delta\text{ADC} = \text{deltaADC} - \min(\text{deltaADC}), \quad (2)$$

Selanjutnya ditentukan nilai DAC dengan persamaan (3):

$$\text{DAC} = \frac{\text{ADC}_{\text{terukur}}}{1024} \times V_{\text{maks}}, \quad (3)$$

Setelah diperoleh nilai DAC kemudian diubah ke ppm dengan persamaan (4):

$$\text{ppm}_{\text{sensor}} = \frac{\text{range}_{\text{maks_sensor}} - \text{range}_{\text{min_sensor}}}{V_{\text{maks}}} \times \text{DAC}, \quad (4)$$

Dataset gas mangga dan srikaya yang terbentuk dalam nilai ppm ditunjukkan pada tabel 3. Dataset berukuran 43x13 dengan label kondisi 1 dan 0, dimana 43 adalah jumlah total sampel mangga dan srikaya dan 13 adalah jumlah rekaman dataset oleh 3 sensor berbeda selama 4 hari ditambah dengan 1 label. Dataset dalam kondisi tidak seimbang karena jumlah sampel yang berbeda. Oleh karena itu, diterapkan algoritma jittering dan SMOTE untuk menambah dan menyeimbangkan dataset. Penerapan algoritma jittering ditujukan untuk augmentasi data sehingga menambah jumlah dataset namun dengan tetap mempertahankan komposisi jumlah fitur mangga 40 dan srikaya 3 untuk setiap prosesnya [23]–[25]. Sedangkan algoritma SMOTE ditujukan untuk menyeimbangkan perbandingan jumlah dataset mangga dan srikaya sehingga perbandingannya seimbang.

Implementasi algoritma jittering pada penelitian ini adalah dengan menggunakan 5 nilai magnitude jittering berbeda, yaitu 0.05, 0.075, 0.1, 0.25, dan 1 [26]. Setiap satu variasi magnitude akan menghasilkan dataset baru dengan ukuran sama dengan ukuran dataset awal 43x13. Sehingga hasil akhirnya berupa sebuah dataset berukuran 258 x 13 yang merupakan gabungan antara dataset awal dengan dataset jittering dan masih dalam kondisi tak seimbang. Berikutnya adalah penerapan SMOTE untuk membentuk dataset seimbang dengan ukuran 504 x 13. Sehingga terdapat 3 dataset yang digunakan pada penelitian ini seperti ditunjukkan pada tabel 4.

Tabel 3. Sampel dataset hasil konversi ppm

Sampel ke	tg26	mq3	mq8	kondisi
0	0.15	5.20	4.51	0
1	0.25	11.58	10.46	1
2	0.15	8.02	7.27	0
3	0.22	11.67	12.92	0
4	0.73	43.79	60.64	0
5	0.59	23.42	30.03	1
6	0.48	25.89	37.29	1

Tabel 4. Dataset penelitian

No	Nama Dataset	Ukuran
1	Dataset asli	43 x 13
2	Dataset Tak Imbang (Jittering)	258 x 13
3	Dataset Imbang (Jittering + SMOTE)	504 x 13

2.4. Penerapan Algoritma Ensemble Learning dan Evaluasinya

Algoritma ensemble learning yang diterapkan pada penelitian ini untuk membangun model ensemble terbaik adalah ensemble learning dynamic dan stacking. Metode ensemble dynamic, tidak seperti algoritme ensemble statis yang menggabungkan model basis tetap, memilih satu model terbaik atau menggabungkan subset model dari

kumpulan menggunakan kondisi just-in-time yang bergantung pada pola input tertentu dari mana prediksi akan dibuat, dilakukan saat membuat prediksi. Algoritma ini menerapkan teknik dimana hanya algoritma pengklasifikasi dasar yang memperoleh level kompetensi tertinggi yang dipilih untuk proses pembelajaran klasifikasi. Ada dua metode pemilihan model dasar metode dynamic yaitu teknik *Overall Local Accuracy* (OLA) atau *Local Class Accuracy* (LCA) [27], [28].

Metode Stacking (Stacked Generalization) melibatkan pelatihan model meta untuk membuat prediksi berdasarkan output dari beberapa model dasar. Model dasar biasanya merupakan jenis model yang berbeda, atau model yang sama dilatih dengan parameter atau sub-kumpulan data yang berbeda. Gagasan di balik stacking adalah bahwa setiap model dasar mungkin memiliki kekuatan dan kelemahan, dan dengan menggabungkan prediksi mereka, meta-model dapat belajar menimbang kekuatan dan kelemahan ini dengan tepat. Ini dapat menghasilkan kinerja keseluruhan yang lebih baik daripada model individual mana pun [29]. Model dasar yang diterapkan untuk membangun model ensemble diantaranya adalah Logistic Regression (LR), Decision Tree (DT), Gaussian Naïve Bayes (GNB), dan Multi-Layer Perceptron (MLP).

Pada penelitian ini rekayasa kombinasi dataset, model dasar, data sensor, metode ensemble, dan jumlah dataset uji ditunjukkan pada tabel 5. Pertama-tama akan dilakukan simulasi model yang hanya menggunakan sampel mangga pada dataset asli dan metode ensemble dynamic. Hasilnya akan dievaluasi untuk menemukan kombinasi model dasar terbaik yang akan diterapkan pada proses berikutnya. Evaluasi hanya akan dititikberatkan pada akurasi model. Selanjutnya pada tahap kedua hanya akan menerapkan kombinasi seperti yang ditampilkan pada tabel 5 namun hanya menggunakan kombinasi model dasar sesuai hasil pengujian simulasi model pertama. Pada tahap kedua ini juga hanya akan menggunakan akurasi model sebagai metrik evaluasi.

Tabel 5. Rekayasa kombinasi jumlah model dasar pembelajaran mesin dan jumlah sensor untuk membentuk model Ensemble learning

Dataset	Model dasar	Jumlah Sensor	Metode Ensemble learning	Variasi dataset uji
Dataset asli	Kombinasi 2 (C2)	Kombinasi 3	Stacking Dynamic-OLA	20 % 30 %
Dataset takimbang	Kombinasi 3 (C3)			
Datasetimbang	Kombinasi 4 (C4)			

3. Hasil

3.1. Penentuan Kombinasi Model Dasar Terbaik

Hasil simulasi pada tabel 6 menunjukkan bahwa pemilihan metode dasar dapat memiliki dampak besar pada akurasi dalam deteksi kualitas buah. Model Dasar Logistic Regression memiliki kinerja yang stabil dan tinggi. Begitupula dengan Ensemble Dynamic OLA menunjukkan potensi untuk meningkatkan kinerja lebih lanjut terutama saat metode dasar Logistic Regression dikombinasikan dengan model dasar Decision Tree. Oleh karena itu pada tahap simulasi selanjutnya digunakan kombinasi metode dasar Logistic Regression dan Decision Tree. Hasil simulasi tahap berikutnya diberikan pada tabel 6 dan tabel 7.

Tabel 6. Hasil pengujian simulasi model pertama

No	Metode	Akurasi data uji 20%			Akurasi data uji 30%		
		C2	C3	C4	C2	C3	C4
1	Ensemble Dynamic OLA	0.75	0.625	0.625	0.67	0.67	0.67
2	Model Dasar Logistic Regression	0.75	0.75	0.75	0.75	0.75	0.75
3	Model Dasar Decision Tree	0.625	0.5	0.5	0.5	0.5	0.42
4	Model Dasar Gaussian Naïve Bayes	-	0.5	0.5	-	0.42	0.42

No	Metode	Akurasi data uji 20%			Akurasi data uji 30%		
		C2	C3	C4	C2	C3	C4
5	Model Dasar Multi-Layer Perceptron	-	-	0.625	-	-	0.5

Tabel 7. Pengujian model ensemble dynamic

Metode	Dataset asli		Tak imbang (Jittering)		Imbang (Jittering+SMOTE)	
	Acc.	Acc.	Acc.	Acc.	Acc.	Acc.
	(uji 20%)	(uji 30%)	(uji 20%)	(uji 30%)	(uji 20%)	(uji 30%)
Ensemble DCS-OLA	0.556	0.615	0.962	0.910	0.960	0.954
Model Dasar Logistic Regression	0.778	0.462	0.692	0.641	0.802	0.803
Model Dasar Decision Tree	0.556	0.615	0.942	0.962	1.00	0.967

Tabel 8. Pengujian model ensemble stacking

Metode	Dataset asli		Tak imbang (Jittering)		Imbang (Jittering+SMOTE)	
	Acc.	Acc.	Acc.	Acc.	Acc.	Acc.
	(uji 20%)	(uji 30%)	(uji 20%)	(uji 30%)	(uji 20%)	(uji 30%)
Ensemble Stacking	0.524	0.517	0.964	0.940	0.985	0.972
Model Dasar Logistic Regression	0.504	0.502	0.649	0.643	0.801	0.794
Model Dasar Decision Tree	0.507	0.526	0.959	0.943	0.987	0.973

3.2. Dynamic Ensemble dan Model Dasarnya

Hasil pengujian model ensemble dynamic DCS-OLA menunjukkan perbedaan yang signifikan antara kondisi tak imbang (jittering) dan imbang (jittering+SMOTE). Pada kondisi tak imbang, model dasar Logistic Regression memiliki akurasi yang rendah, terutama pada uji 30%. Namun, ketika dataset seimbang, akurasi model Logistic Regression meningkat secara signifikan, mencapai 80-80.3%. Model Decision Tree juga menunjukkan peningkatan yang signifikan dalam kondisi imbang dengan akurasi mencapai 100% pada beberapa kasus uji. Meskipun model dasar Logistic Regression dan Decision Tree mengalami peningkatan dalam akurasi, Ensemble DCS-OLA tetap unggul dengan akurasi yang sangat tinggi, terutama pada kondisi imbang. Ini menunjukkan bahwa penggabungan model dalam metode ensemble dapat menghasilkan hasil yang jauh lebih baik daripada menggunakan model dasar secara individu.

3.3. Stacking Ensemble dan Model Dasarnya

Hasil pengujian model ensemble stacking juga menunjukkan peningkatan yang signifikan dalam akurasi, terutama pada kondisi imbang dengan penerapan jittering dan SMOTE. Model dasar Logistic Regression dalam ensemble stacking juga mengalami peningkatan akurasi yang cukup besar pada kondisi imbang, mencapai 80.1-80.3%. Namun, Ensemble Stacking masih unggul dengan akurasi yang lebih tinggi, mencapai 97.2-98.5%. Ini menunjukkan bahwa pendekatan stacking dapat menghasilkan model yang lebih baik daripada menggunakan model dasar secara individu, bahkan ketika dataset telah seimbang.

4. Pembahasan

Hasil penelitian ini menunjukkan bahwa sistem yang menggunakan model berbasis metode ensemble, baik itu ensemble dynamic atau stacking, memiliki kemampuan yang sangat baik dalam mendeteksi kondisi buah normal atau rusak. Ensemble DCS-OLA dan Ensemble Stacking keduanya memberikan akurasi yang sangat tinggi, dengan Ensemble Stacking mencapai akurasi tertinggi. Pengaruh pemangkasan dataset mangga menjadi

hanya 4 hari untuk menyesuaikan dengan lama waktu simpan srikaya merupakan langkah yang baik. Hal ini mengurangi perbedaan dalam atribut lama waktu simpan antara kedua jenis buah dan memungkinkan model untuk lebih fokus pada fitur-fitur lain yang mungkin berperan dalam deteksi kondisi normal atau rusak.

Kemudian penerapan algoritma jittering dan balancing pada dataset srikaya sangat efektif dalam menyeimbangkan jumlah dataset antara mangga dan srikaya. Hasilnya adalah peningkatan yang signifikan dalam akurasi deteksi pada kedua jenis ensemble, terutama pada stacking ensemble. Dan terakhir perbandingan performa model dasar dan ensemble dari hasil pengujian menunjukkan bahwa ensemble dynamic (DCS-OLA) dan ensemble stacking memiliki performa yang jauh lebih baik dibandingkan dengan model dasar. Hasil penelitian ini sejalan dengan beberapa hasil penelitian sebelumnya dimana metode jittering memang mampu meningkatkan performa model-model pembelajaran mesin [30]–[32].

Namun, ada beberapa pertimbangan yang perlu diperhatikan. Pertama, sebaiknya dilakukan validasi lebih lanjut menggunakan dataset yang lebih besar dan variasi kondisi yang lebih luas untuk memastikan hasil yang lebih umum. Kedua, metode ensemble dapat menjadi lebih kompleks dan memakan waktu dalam pengembangan dan pelatihan model, sehingga perlu diperhitungkan dalam pengaplikasiannya di dunia nyata. Ketiga, penggabungan model dasar yang berbeda, terutama dalam jumlah yang cukup besar, dapat memberikan hasil yang lebih akurat dan andal. Akan tetapi, perlu dicatat bahwa kinerja ensemble juga tergantung pada seleksi model dasar yang tepat, jumlah model yang digunakan, dan karakteristik data uji yang dipakai dalam evaluasi.

5. Kesimpulan

Penelitian ini berhasil menunjukkan bahwa metode ensemble dynamic dan stacking dapat meningkatkan kinerja model dalam mendeteksi kondisi buah normal atau rusak. Dengan menggunakan teknik pemangkasan dataset, jittering, dan balancing, hasil akurasi yang sangat baik dapat dicapai, terutama pada kondisi dataset yang seimbang. Ensemble DCS-OLA dan Ensemble Stacking merupakan pendekatan yang sangat efektif dalam menggabungkan kelebihan model dasar (Logistic Regression dan Decision Tree) untuk mencapai akurasi yang sangat tinggi. Ini menunjukkan potensi besar dalam penggunaan ensemble dalam pengembangan sistem berbasis machine learning untuk mendeteksi kualitas buah.

Namun, perlu dicatat bahwa hasil ini masih memerlukan validasi lebih lanjut dan penerapan di dunia nyata. Selain itu, perlu diperhatikan bahwa kompleksitas pengembangan dan pelatihan model ensemble dapat menjadi tantangan tersendiri. Dalam pengaplikasiannya, diperlukan perencanaan dan sumber daya dataset yang cukup untuk memastikan keberhasilan implementasi.

Ucapan Terima Kasih: Penelitian ini dibiayai oleh Daftar Isian Pelaksanaan Anggaran (DIPA) Universitas Negeri Semarang Nomor: DPA 023.17.2.690645/2023.05/2023, tanggal 30 bulan Desember tahun 2023 sesuai dengan Surat Perjanjian Penugasan Pelaksanaan Penelitian Dosen Pemula (Fakultas) Dana DIPA UNNES Tahun 2023 Nomor 89.17.4/UN37/PPK.05/2023.

Referensi

- [1] A. U. Alam, P. Rathi, H. Beshai, G. K. Sarabha, and M. Jamal Deen, "Fruit quality monitoring with smart packaging," *Sensors*, vol. 21, no. 4, pp. 1–30, 2021.
- [2] A. Gordon and D. Gordon, *Food safety and quality systems implementation along value chains*. 2020.
- [3] K.-T. Li, "Physiology and Classification of Fruits," in *Handbook of Fruits and Fruit Processing: Second Edition*, Second., N. K. Sinha, J. S. Sidhu, J. Barta, J. S. B. Wu, and M. P. Cano, Eds. Oxford, United Kingdom: Wiley-Blackwell, 2012, pp. 3–12.
- [4] A. M. Bratu, C. Popa, M. Bojan, P. C. Logofatu, and M. Petrus, "Non-destructive methods for fruit quality evaluation," *Sci. Rep.*, vol. 11, no. 1, pp. 1–15, 2021.
- [5] B. Hasanzadeh, Y. Abbaspour-Gilandeh, A. Soltani-Nazarloo, M. Hernández-Hernández, I. Gallardo-Bernal, and J. L. Hernández-Hernández, "Non-Destructive Detection of Fruit Quality Parameters Using Hyperspectral Imaging, Multiple Regression Analysis and Artificial Intelligence," *Horticulturae*, vol. 8, no. 7, 2022.

- [6] M. Murad, S. Sukmawaty, A. Ansar, R. Sabani, and S. Hidayat, "Sistem Pendeteksi Kerusakan Buah Mangga Menggunakan Sensor Gas Dengan Metode DCS - LCA," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 3, no. 4, pp. 186–194, 2021.
- [7] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," *Sensors (Switzerland)*, vol. 18, no. 8, pp. 1–29, 2018.
- [8] Q. Ding, D. Zhao, J. Liu, and Z. Yang, "Detection of fruits in warehouse using Electronic nose," in *MATEC Web of Conferences*, 2018, vol. 232, pp. 1–6.
- [9] N. Geethapriya and S. M. Praveena, "Evaluation of Fruit Ripeness Using Electronic Nose," *Int. J. Adv. Inf. Sci. Technol.*, vol. 6, no. 5, pp. 1–5, 2017.
- [10] N. Aghilinategh, M. J. Dalvand, and A. Anvar, "Detection of ripeness grades of berries using an electronic nose," *Food Sci. Nutr.*, vol. 8, no. 9, pp. 4919–4928, 2020.
- [11] M. Baietto and A. D. Wilson, "Electronic-nose applications for fruit identification, ripeness and quality grading," *Sensors (Switzerland)*, vol. 15, no. 1, pp. 899–931, 2015.
- [12] M. Di Napoli *et al.*, "Machine learning ensemble modelling as a tool to improve landslide susceptibility mapping reliability," *Landslides*, vol. 17, no. 8, pp. 1897–1914, 2020.
- [13] A. S. Britto, R. Sabourin, and L. E. S. Oliveira, "Dynamic selection of classifiers - A comprehensive review," *Pattern Recognit.*, vol. 47, no. 11, pp. 3665–3680, 2014.
- [14] M. Woźniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Inf. Fusion*, vol. 16, no. 1, pp. 3–17, 2014.
- [15] R. M. O. Cruz, R. Sabourin, and G. D. C. Cavalcanti, "Dynamic classifier selection: Recent advances and perspectives," *Inf. Fusion*, vol. 41, pp. 195–216, 2018.
- [16] M. Astani, M. Hasheminejad, and M. Vaghefi, "A diverse ensemble classifier for tomato disease recognition," *Comput. Electron. Agric.*, vol. 198, p. 107054, 2022.
- [17] P. Tyagi, R. Semwal, A. Sharma, U. S. Tiwary, and P. Varadwaj, "E-nose: a low-cost fruit ripeness monitoring system," *J. Agric. Eng.*, vol. 54, no. 1, 2023.
- [18] Y. Fukano and Y. Tachiki, "Evolutionary ecology of climacteric and non-climacteric fruits," *Biol. Lett.*, vol. 17, no. 9, 2021.
- [19] B. K. Iwana and S. Uchida, "An empirical survey of data augmentation for time series classification with neural networks," *PLoS One*, vol. 16, no. 7 July, pp. 1–32, 2021.
- [20] B. Liu, Z. Zhang, and R. Cui, "Efficient Time Series Augmentation Methods," in *Proceedings - 2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISP-BMEI 2020*, 2020, pp. 1004–1009.
- [21] D. Elreedy, A. F. Atiya, and F. Kamalov, "A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning," *Mach. Learn.*, no. 0123456789, 2023.
- [22] I. Daugela, J. Suziedelyte Visockiene, J. Kumpiene, and I. Suzdalev, "Measurements of flammable gas concentration in landfill areas with a low-cost sensor," *Energies*, vol. 14, no. 13, 2021.
- [23] S. Braun and I. Tashev, *Data Augmentation and Loss Normalization for Deep Noise Suppression*, vol. 12335 LNAI. Springer International Publishing, 2020.
- [24] R. Ni, M. Goldblum, A. Sharaf, K. Kong, and T. Goldstein, "Data Augmentation for Meta-Learning," *Proc. Mach. Learn. Res.*, vol. 139, pp. 8152–8161, 2021.
- [25] B. Hanin and Y. Sun, "How Data Augmentation affects Optimization for Linear Regression," *Adv. Neural Inf. Process. Syst.*, vol. 10, no. NeurIPS, pp. 8095–8105, 2021.
- [26] J. P. Teixeira and A. Gonçalves, "Algorithm for Jitter and Shimmer Measurement in Pathologic Voices," *Procedia Comput. Sci.*, vol. 100, pp. 271–279, 2016.
- [27] R. M. O. Cruz, L. G. Hafemann, R. Sabourin, and G. D. C. Cavalcanti, "DESlib: A Dynamic ensemble selection library in Python," *J. Mach. Learn. Res.*, vol. 21, pp. 1–5, 2020.
- [28] M. A. Souza, G. D. C. Cavalcanti, R. M. O. Cruz, and R. Sabourin, "Online local pool generation for dynamic classifier selection," *Pattern Recognit.*, vol. 85, no. 1, pp. 132–148, 2018.
- [29] H. Li, Y. Jin, J. Zhong, and R. Zhao, "A Fruit Tree Disease Diagnosis Model Based on Stacking Ensemble Learning," *Complexity*, vol. 2021, 2021.
- [30] V. Nourani, P. Asghari, and E. Sharghi, "Artificial intelligence based ensemble modeling of wastewater treatment plant using jittered data," *J. Clean. Prod.*, vol. 291, p. 125772, 2021.
- [31] P. J. Azevedo and A. M. Jorge, "Ensembles of jittered association rule classifiers," *Data Min. Knowl. Discov.*, vol. 21, no. 1, pp. 91–129, 2010.
- [32] G. P. Zhang, "A neural network ensemble method with jittered training data for time series forecasting," *Inf. Sci. (Ny)*, vol. 177, no. 23, pp. 5329–5346, 2007.