

Analisis Diskriminan Untuk Klasifikasi Varietas Benih Gandum Berdasarkan Karakteristik Morfologi Biji

Onoy Rohaeni¹, Diyani Fauziyah², Fadhli Jahfal Aufa Maulana³
Fakultas Sains dan Teknologi Universitas Koperasi Indonesia^{1,2,3}
onoyrohaeni@ikopin.ac.id¹

ABSTRAK

Klasifikasi varietas benih gandum merupakan salah satu tahapan penting dalam menjaga mutu benih dan mendukung proses pemuliaan tanaman. Perbedaan karakteristik morfologi biji antar varietas dapat dimanfaatkan sebagai dasar dalam proses identifikasi secara objektif. Penelitian ini bertujuan untuk menganalisis karakteristik morfologi biji yang membedakan varietas gandum serta membangun model klasifikasi menggunakan metode analisis diskriminan. Data yang digunakan berupa data morfologi biji gandum yang terdiri atas tiga varietas, yaitu Kama, Rosa, dan Canadian, dengan variabel pengamatan meliputi *Area*, *Perimeter*, *Compactness*, *Length of Kernel*, *Width of Kernel*, *Asymmetry Coefficient*, dan *Length of Kernel Groove*. Analisis dilakukan melalui pengujian asumsi, pembentukan fungsi diskriminan, pengujian signifikansi model, identifikasi variabel yang berkontribusi terhadap pembentukan fungsi diskriminan, serta evaluasi tingkat ketepatan klasifikasi. Hasil penelitian menunjukkan bahwa model diskriminan yang dibentuk valid dan signifikan secara statistik dengan nilai signifikansi sebesar 0,000, sehingga mampu membedakan ketiga varietas gandum secara nyata. Berdasarkan hasil analisis, variabel *Area* dan *Perimeter* merupakan karakteristik morfologi yang memberikan kontribusi terbesar sebagai pembeda utama antar varietas. Selain itu, model klasifikasi yang dihasilkan memiliki tingkat akurasi sebesar 98,0% yang menunjukkan kemampuan klasifikasi yang sangat tinggi. Dengan demikian, analisis diskriminan terbukti efektif dalam mengklasifikasikan varietas benih gandum berdasarkan karakteristik morfologi biji dan sangat layak diterapkan sebagai metode pendukung dalam sistem penyortiran benih secara otomatis.

Kata Kunci: Analisis diskriminan, klasifikasi, benih gandum, karakteristik morfologi, akurasi klasifikasi

I. PENDAHULUAN

Gandum (*Triticum spp.*) merupakan salah satu komoditas pangan utama di dunia yang memiliki peran penting dalam memenuhi kebutuhan pangan masyarakat. Berbagai varietas benih gandum dikembangkan untuk memperoleh karakteristik unggul, seperti produktivitas tinggi, ketahanan terhadap hama dan penyakit, serta kemampuan beradaptasi terhadap kondisi lingkungan tertentu. Oleh karena itu, identifikasi dan klasifikasi varietas benih gandum menjadi aspek yang penting dalam kegiatan pemuliaan tanaman, pengendalian mutu benih, maupun penelitian di bidang pertanian.

Salah satu cara yang umum digunakan untuk membedakan varietas benih gandum adalah melalui pengamatan karakteristik morfologi biji, seperti panjang, lebar, ketebalan, luas, keliling, serta sifat fisik lainnya. Karakteristik tersebut dapat dijadikan variabel pembeda karena setiap varietas memiliki ciri morfologi yang khas. Namun, apabila jumlah variabel yang diamati cukup banyak dan terdapat kemiripan karakteristik antar varietas, proses klasifikasi secara manual menjadi kurang efektif dan berpotensi menimbulkan kesalahan.

Analisis diskriminan merupakan salah satu metode statistika multivariat yang bertujuan untuk mengelompokkan suatu objek ke dalam kelompok yang telah diketahui berdasarkan beberapa variabel prediktor. Metode ini membentuk fungsi diskriminan yang mampu

memaksimalkan perbedaan antar kelompok sekaligus meminimalkan variasi di dalam kelompok. Selain digunakan sebagai metode klasifikasi, analisis diskriminan juga dapat mengidentifikasi variabel-variabel yang paling berpengaruh dalam membedakan setiap kelompok.

Penerapan analisis diskriminan pada data karakteristik morfologi biji gandum diharapkan mampu menghasilkan model klasifikasi yang akurat untuk membedakan varietas gandum kama, rosa, dan canadian sehingga dapat membantu proses identifikasi varietas benih secara objektif dan efisien. Hasil penelitian ini juga diharapkan memberikan informasi mengenai karakteristik morfologi yang paling berperan dalam membedakan varietas benih gandum, sehingga dapat menjadi referensi bagi peneliti, pemulia tanaman, maupun pihak yang bergerak di bidang pengawasan mutu benih.

II. TINJAUAN PUSTAKA

A. Gandum (*Triticum spp.*)

Gandum (*Triticum spp.*) merupakan salah satu tanaman serealia yang menjadi sumber pangan utama di berbagai negara. Gandum memiliki kandungan karbohidrat, protein, vitamin, dan mineral yang tinggi sehingga banyak dimanfaatkan sebagai bahan baku berbagai produk pangan, seperti tepung terigu, roti, mi, dan biskuit. Berdasarkan data Organisasi Pangan dan Pertanian Dunia (FAO), gandum termasuk tiga komoditas pangan utama dunia bersama padi dan jagung.

Secara umum, gandum dibedakan menjadi beberapa spesies, di antaranya *Triticum aestivum* (gandum roti), *Triticum durum* (gandum pasta), dan *Triticum compactum*. Masing-masing spesies memiliki karakteristik morfologi dan kualitas biji yang berbeda sehingga memengaruhi pemanfaatannya dalam industri pangan.

Dalam penelitian klasifikasi varietas, karakteristik morfologi biji menjadi salah satu parameter penting karena relatif mudah diamati, bersifat objektif, dan mampu mencerminkan perbedaan genetik antar varietas. Varietas merupakan kelompok tanaman dalam satu spesies yang memiliki karakteristik tertentu sehingga dapat dibedakan dari kelompok lainnya. Perbedaan varietas dipengaruhi oleh faktor genetik maupun lingkungan tumbuh.

Pada penelitian ini digunakan tiga varietas gandum, yaitu Kama, Rosa, dan Canadian. Ketiga varietas tersebut memiliki karakteristik morfologi biji yang berbeda, baik dari ukuran, bentuk, maupun teksturnya. Perbedaan tersebut dapat dimanfaatkan sebagai dasar dalam proses klasifikasi menggunakan pendekatan statistik multivariat. Karakteristik morfologi merupakan sifat fisik yang dapat diamati secara langsung. Pada benih gandum, karakteristik morfologi meliputi ukuran, bentuk, dan struktur permukaan biji.

B. Analisis Multivariat

Analisis multivariat merupakan sekumpulan metode statistik yang digunakan untuk menganalisis data yang memiliki lebih dari satu variabel respon atau variabel bebas secara simultan. Tujuan analisis multivariat adalah memperoleh informasi yang tidak dapat diperoleh melalui analisis univariat maupun bivariat.

Metode analisis multivariat meliputi Analisis Komponen Utama (Principal Component Analysis), Analisis Faktor, Analisis Cluster, Regresi Logistik Multivariat, Analisis Kanonik,

dan Analisis Diskriminan. Pemilihan metode bergantung pada tujuan penelitian serta jenis data yang dianalisis.

Dalam penelitian klasifikasi varietas benih gandum, metode yang digunakan adalah analisis diskriminan karena variabel respon berupa kelompok atau kategori yang telah diketahui sebelumnya.

C. Analisis Diskriminan

Interpretasi Analisis diskriminan adalah teknik statistika multivariat yang termasuk dalam metode dependensi, di mana tujuannya adalah memisahkan atau mengklasifikasikan objek-objek pengamatan ke dalam dua atau lebih kelompok yang berbeda secara tegas (*mutually exclusive*) berdasarkan serangkaian variabel independen bersifat metrik atau kuantitatif (Ghozali, 2021). Prinsip dasar metode ini adalah membentuk kombinasi *linear* dari variabel-variabel independen (disebut fungsi diskriminan) yang dapat memaksimumkan perbedaan antar kelompok (Santoso & Rahmawati, 2023).

Agar hasil analisis diskriminan valid dan tidak bias, terdapat beberapa asumsi dasar yang harus dipenuhi, antara lain:

1. Normalitas Multivariat: Variabel independen dalam setiap kelompok harus berdistribusi normal.
2. Homogenitas Matriks Varians-Kovarians: Varians antar kelompok data harus sama atau homogen. Pelanggaran asumsi ini dapat memengaruhi tingkat akurasi klasifikasi, meskipun analisis diskriminan dikenal cukup tangguh (*robust*) jika ukuran sampel antar kelompok relatif seimbang (Ibrahim, 2024).
3. Tidak Ada Multikolinearitas: Tidak boleh ada korelasi yang sangat tinggi antar variabel independen.

Apabila asumsi-asumsi tersebut terpenuhi, maka fungsi diskriminan yang dihasilkan akan memiliki kemampuan klasifikasi yang optimal.

III. METODE PENELITIAN

A. Sumber Data Penelitian

Sumber data yang digunakan dalam penelitian ini adalah data sekunder pengukuran geometris biji gandum yang terdiri dari tiga varietas yaitu Kama, Rosa, dan Canadian. Total sampel yang dianalisis mencakup representasi seimbang dari ketiga varietas tersebut.

B. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri dari enam variabel numerik, yaitu:

1. *Area* (Luas permukaan biji)
2. *Perimeter* (Keliling biji)
3. *Compactness* (Kepadatan bentuk)
4. *Kernel Length* (Panjang biji)
5. *Kernel Width* (Lebar biji)
6. *Asymmetry Coefficient* (Tingkat ketidaksimetrisan)
7. *Kernel Groove* (Panjang alur biji)

Ketujuh karakteristik tersebut mampu menggambarkan bentuk geometris biji secara kuantitatif, sehingga banyak digunakan dalam penelitian klasifikasi maupun pengenalan pola.

C. Teknik Pengolahan dan Analisis Data

Tahapan analisis data dilakukan menggunakan perangkat lunak SPSS dengan langkah-langkah sebagai berikut:

1. Pengujian Asumsi (*Assumption Testing*): Tahap ini krusial untuk memastikan model yang dihasilkan bersifat *Best Linear Unbiased Estimator* (BLUE). Uji normalitas dilakukan menggunakan metode *Shapiro-Wilk* mengingat jumlah sampel per grup kurang dari 50, sedangkan uji homogenitas varians-kovarians menggunakan statistik *Box's M*.
2. Estimasi Model Diskriminan: Jika asumsi terpenuhi, langkah selanjutnya adalah mengestimasi fungsi diskriminan. Uji signifikansi dilakukan menggunakan *Wilks' Lambda* untuk melihat apakah variabel independen secara simultan mampu membedakan kelompok. Kekuatan hubungan diukur menggunakan *Canonical Correlation*.
3. Interpretasi Variabel Pembeda: Melakukan analisis pada *Structure Matrix* untuk mengetahui korelasi antara variabel independen dengan fungsi diskriminan, guna menentukan variabel mana yang menjadi ciri khas utama tiap varietas.
4. Validasi Model (*Classification*): Langkah terakhir adalah menghitung akurasi prediksi model (*Hit Ratio*) dengan membandingkan hasil prediksi model terhadap kategori varietas yang sebenarnya menggunakan tabel klasifikasi (*confusion matrix*).

IV. HASIL PENELITIAN DAN PEMBAHASAN

A. Pengujian Asumsi Klasik

Analisis diskriminan merupakan bagian dari statistik parametrik yang menuntut dipenuhinya beberapa asumsi dasar yaitu normalitas data dan homogenitas matriks varian-kovarians.

1. Uji Normalitas

Berdasarkan hasil uji normalitas menggunakan metode *Shapiro-Wilk* pada Gambar 1, terlihat bahwa mayoritas variabel pada ketiga kelompok varietas memiliki nilai signifikansi (*Sig.*) di atas 0.05. Hal ini mengindikasikan bahwa data berdistribusi normal secara univariat. Meskipun terdapat sedikit penyimpangan pada variabel *Asymmetry Coeff* untuk varietas Kama (*Sig.* 0.026), analisis diskriminan diketahui memiliki sifat *robust* (tangguh) terhadap sedikit pelanggaran normalitas selama ukuran sampel cukup besar.

Tests of Normality							
	Type	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Area	Kama	.059	66	.200 [*]	.994	66	.992
	Rosa	.112	68	.033	.965	68	.050
	Canadian	.091	65	.200 [*]	.971	65	.127
Perimeter	Kama	.048	66	.200 [*]	.989	66	.826
	Rosa	.089	68	.200 [*]	.976	68	.216
	Canadian	.060	65	.200 [*]	.987	65	.749
Compactness	Kama	.061	66	.200 [*]	.995	66	.996
	Rosa	.082	68	.200 [*]	.981	68	.375
	Canadian	.061	65	.200 [*]	.986	65	.668
Kernel.Length	Kama	.068	66	.200 [*]	.987	66	.723
	Rosa	.042	68	.200 [*]	.984	68	.532
	Canadian	.062	65	.200 [*]	.991	65	.924
Kernel.Width	Kama	.065	66	.200 [*]	.990	66	.895
	Rosa	.115	68	.025	.976	68	.221
	Canadian	.105	65	.074	.964	65	.054
Asymmetry.Coeff	Kama	.077	66	.200 [*]	.958	66	.026
	Rosa	.056	68	.200 [*]	.983	68	.460
	Canadian	.081	65	.200 [*]	.987	65	.715
Kernel.Groove	Kama	.082	66	.200 [*]	.986	66	.656
	Rosa	.089	68	.200 [*]	.964	68	.045
	Canadian	.083	65	.200 [*]	.983	65	.490

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Gambar 1. Test of Normality

2. Uji Homogenitas (*Box's M*)

Uji kesamaan matriks varians-kovarians dilakukan menggunakan uji Box's M. Hasil pada Gambar 2 berikut menunjukkan nilai statistik Box's M sebesar 624.395 dengan nilai signifikansi 0.000.

Test Results		
Box's M		624.395
F	Approx.	14.218
	df1	42
	df2	113749.624
	Sig.	.000
Tests null hypothesis of equal population covariance matrices.		

Gambar 2. Box's M

Nilai $Sig. < 0.05$ menunjukkan bahwa hipotesis nol ditolak, yang berarti matriks varians-kovarians antar kelompok tidak homogen. Dalam kondisi ideal, homogenitas diperlukan. Namun, pelanggaran asumsi ini dapat diatasi dengan menggunakan matriks kovarians kelompok terpisah (*separate-groups covariance matrix*) dalam proses klasifikasi, terutama

karena jumlah sampel antar kelompok dalam penelitian ini relatif seimbang yaitu: *Kama* = 66, *Rosa* = 68, *Canadian* = 65.

B. Estimasi Fungsi Diskriminan dan Signifikansi Model

Untuk mengetahui apakah model yang terbentuk layak digunakan, dilakukan pengujian Wilks' Lambda. Pada Gambar 3, nilai Wilks' Lambda tercatat sebesar 0.033 dengan nilai signifikansi (*Sig.* 0.000).

Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.033	659.341	14	.000
2	.250	267.625	6	.000

Gambar 3. Wilks' Lambda

Nilai Signifikansi yang jauh di bawah 0.05 ($p < 0.001$) menunjukkan bahwa terdapat perbedaan rata-rata yang sangat nyata antar kelompok varietas gandum. Artinya, variabel morfologi yang digunakan mampu mendiskriminasi atau membedakan ketiga jenis gandum tersebut secara signifikan.

Kekuatan hubungan model juga tercermin dari nilai Eigenvalues pada Gambar 4. Fungsi diskriminan pertama memiliki nilai *Canonical Correlation* sebesar 0.932. Jika dikuadratkan (R^2), nilai ini menjelaskan bahwa 86.8% variasi antar kelompok dapat dijelaskan oleh model, yang menunjukkan hubungan yang sangat kuat.

Eigenvalues				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	6.611 ^a	68.8	68.8	.932
2	3.001 ^a	31.2	100.0	.866

a. First 2 canonical discriminant functions were used in the analysis.

Gambar 4. Eigenvalues

C. Interpretasi Struktur Variabel Pembeda

Setelah model dinyatakan signifikan, langkah selanjutnya adalah menentukan variabel mana yang paling berpengaruh. Hal ini dapat dilihat pada *Structure Matrix* (Gambar 5) yang menunjukkan korelasi antara variabel independen dengan fungsi diskriminan yang terbentuk.

Structure Matrix

	Function	
	1	2
Area	.910*	.163
Perimeter	.909*	.134
Kernel.Width	.765*	.267
Kernel.Groove	.721*	-.329
Kernel.Length	.704*	.031
Asymmetry.Coeff	-.080	-.387*
Compactness	.261	.326*

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
Variables ordered by absolute size of correlation within function.

*. Largest absolute correlation between each variable and any discriminant function

Gambar 5. Structure Matrix

Hasil ini menunjukkan bahwa: Fungsi 1 didominasi oleh variabel Area (0.910) dan Perimeter (0.909). Ini berarti ukuran dimensi (besar/kecilnya biji) adalah factor pembada utama; Fungsi 2 lebih banyak dipengaruhi oleh variabel Asymmetry Coefficient (-0.387), yang berkaitan dengan bentuk simetri biji.

D. Validasi dan Akurasi Klasifikasi

Bagian akhir dan terpenting dari analisis ini adalah melihat ketepatan prediksi model melalui tabel *Classification Results* (Gambar 6).

Classification Results^a

		Type	Predicted Group Membership			Total
			Kama	Rosa	Canadian	
Original	Count	Kama	63	1	2	66
		Rosa	0	68	0	68
		Canadian	1	0	64	65
	%	Kama	95.5	1.5	3.0	100.0
		Rosa	.0	100.0	.0	100.0
		Canadian	1.5	.0	98.5	100.0

a. 98.0% of original grouped cases correctly classified.

Gambar 6. Classification Results

Hasil klasifikasi menunjukkan performa yang sangat memuaskan:

- 1) Tingkat akurasi total (*Hit Ratio*) mencapai 98.0%.
- 2) Varietas Rosa teridentifikasi dengan sempurna 100%.
- 3) Varietas Canadian memiliki akurasi 98.5%.
- 4) Varietas Kama memiliki akurasi 95.5%.

Tingginya angka akurasi ini membuktikan bahwa penggunaan 7 parameter morfologi tersebut sangat efektif untuk menggantikan metode identifikasi manual

V. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan, dapat ditarik beberapa Kesimpulan:

1. Model diskriminan yang dibentuk valid dan signifikan secara statistic (*Sig. 0.000*) untuk membedakan varietas gandum Kama, Rosa, dan Canadian.
2. Variabel Area dan Perimeter merupakan dua karakteristik morfologi yang memiliki kontribusi terbesar sebagai pembeda utama antar varietas.
3. Model klasifikasi menghasilkan tingkat akurasi yang sangat tinggi, yaitu sebesar 98.0%, sehingga metode ini sangat layak diterapkan dalam system penyortiran benih otomatis.

DAFTAR PUSTAKA

- Badan Pusat Statistik. (2024). *Statistik Tanaman Pangan Indonesia 2023*. Jakarta: Badan Pusat Statistik.
- Ghozali, I. (2021). *Aplikasi Analisis Multivariate dengan Program IBM SPSS 26*. Semarang: Badan Penerbit Universitas Diponegoro.
- Hidayat, R., & Kusuma, W. A. (2023). *Penerapan Algoritma Machine Learning untuk Klasifikasi Mutu Benih Padi Berdasarkan Fitur Morfologi*. Jurnal Ilmu Komputer dan Agri-Informatika, 10(2). 115-124.
- Pratama, A. R., & Arifin, Z. (2022). *Identifikasi Varietas Kedelai Menggunakan Ekstraksi Ciri Morfologi dan Jaringan Syaraf Tiruan*. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK), 9(3), 567-574.
- Santoso, S., & Rahmawati, Y. (2023). *Statistika Multivariat Terapan untuk Riset Pertanian*. Malang: UB Press.
- Sujarweni, V. W. (2022). *Metodologi Penelitian: Lengkap, Praktis, dan Mudah dipahami*. Yogyakarta: Pustaka Baru Press.
- Widowati, S., Khumaida, N., & Ardi, S. W. (2020). *Karakteristik Morfologi dan Sifat Kuantitatif Plasma Nutfah Gandum Tropis*. Jurnal Agronomi Indonesia, 48(2), 130-138.
- Yulianto, T. (2021). *Analisis Diskriminan Faktor-Faktor yang Mempengaruhi Produksi Padi di Jawa Tengah*. Jurnal Statistika Industri dan Komputasi, 6(1), 22-30.