

## KLASIFIKASI MUTU BERAS MENGGUNAKAN K-NN BERBASIS BACKWARD ELIMINATION

### CLASSIFICATION RICE QUALITY USING K-NN BACKWARD-BASED ELIMINATION

Muhammad Faishol Amrulloh<sup>1\*</sup>, Muhammad Mahrus Ali<sup>2</sup>, Mochamad Sirodjudin<sup>3</sup>

<sup>1</sup>Universitas Yudharta Pasuruan, <sup>2</sup>Universitas Merdeka Pasuruan, <sup>3</sup>Universitas Negeri Malang  
Email : <sup>1</sup>\*faishol@yudharta.ac.id, <sup>2</sup>mahrusali@unmerpas.ac.id, <sup>3</sup>mochamadsirodjudin@gmail.com

\*Penulis Korespondensi

**Abstrak** - Beras merupakan salah satu hasil pertanian yang paling utama. Dan termasuk komoditas yang strategis karena hampir semua masyarakat Indonesia membutuhkannya. Karena pentingnya fungsi beras sebagai bahan makanan pokok, maka kualitas beras yang akan dikonsumsi harus dipastikan benar-benar berkualitas atau bermutu. Penentuan kualitas atau mutu beras sampai saat ini sudah banyak dilakukan oleh peneliti terdahulu. Namun dari beberapa metode yang sudah dipilih, mayoritas menggunakan image processing. Pengolahan data kualitas beras menggunakan data mining masih jarang dilakukan. Dalam penelitian ini, dataset yang akan digunakan berasal dari database Dinas Pertanian Kabupaten Probolinggo yang mempunyai 1 atribut khusus dan 9 atribut regular. Adapun atributnya yaitu : varietas, panjang, bentuk, warna, rasa, teknik penanaman padi, musim, hama, PH dan mutu. Metode yang akan digunakan dalam penelitian ini yaitu metode K-NN berbasis *backward elimination*.

Dari hasil analisa dan komputasi beberapa metode, dapat disimpulkan bahwa metode K-NN berbasis backward elimination yang digunakan untuk meningkatkan akurasi klasifikasi beras menghasilkan beberapa kesimpulan. Yaitu atribut yang berpengaruh terhadap penentuan mutu beras adalah varietas, bentuk, warna, teknik, musim, hama dan PH. Sedangkan feature yang dianggap tidak berpengaruh adalah rasa dan panjang. Sedangkan untuk data sebanyak 4952, maka akurasi terbaiknya adalah 83.08% yaitu saat  $k=1$ , sedangkan untuk data sebanyak 2000, maka akurasi terbaiknya adalah 87.70% yaitu saat  $k=1$ , sedangkan untuk data sebanyak 1000, maka akurasi terbaiknya adalah 98.60% yaitu saat  $k=1$ .

**Kata Kunci:** *Backward Elimination; Hybrid; Optimize Selection; K-NN;*

**Abstract** - Rice is one of the most important agricultural products. And it is a strategic commodity because almost all Indonesian people need it. Because of the importance of the function of rice as a staple food ingredient, the quality of the rice to be consumed must be ensured to be of high quality. Determination of the quality or quality of rice until now has been done by many previous researchers. However, of the several methods that have been selected, the majority use image processing. Rice quality data processing using data mining is still rarely done. In this study, the dataset that will be used comes from the Probolinggo Regency Agriculture Service database which has 1 special attribute and 9 regular attributes. The attributes are: variety, length, shape, color, taste, rice planting technique, season, pests, PH and quality. The method that will be used in this study is the backward elimination-based K-NN method.

From the results of the analysis and computation of several methods, it can be concluded that the backward elimination-based K-NN method used to increase the accuracy of rice classification resulted in several conclusions. Namely the attributes that affect the determination of rice quality are variety, shape, color, technique, season, pests and PH. While the features that are considered to have no effect are taste and length. As for the data as much as 4952, the best accuracy is 83.08% when  $k = 1$ , while for the data as much as 2000, then the best accuracy is 87.70% at  $k = 1$ , while for the data as much as 1000, the best accuracy is 98.60%, namely when  $k=1$ .

**Keywords :** *Backward Elimination; Hybrid; Optimize Selection; K-NN;*

This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).



## 1. PENDAHULUAN

Beras merupakan salah satu hasil pertanian yang paling utama. Dan termasuk komoditas yang strategis karena hampir semua masyarakat Indonesia membutuhkannya. Karena pentingnya fungsi beras sebagai bahan makanan pokok, maka kualitas beras yang akan dikonsumsi harus dipastikan benar-benar berkualitas atau bermutu. Penentuan kualitas atau mutu beras sampai saat ini sudah banyak dilakukan oleh peneliti terdahulu. Namun dari beberapa metode yang sudah dipilih, mayoritas menggunakan image processing. Pengolahan data kualitas beras menggunakan data mining masih jarang dilakukan.

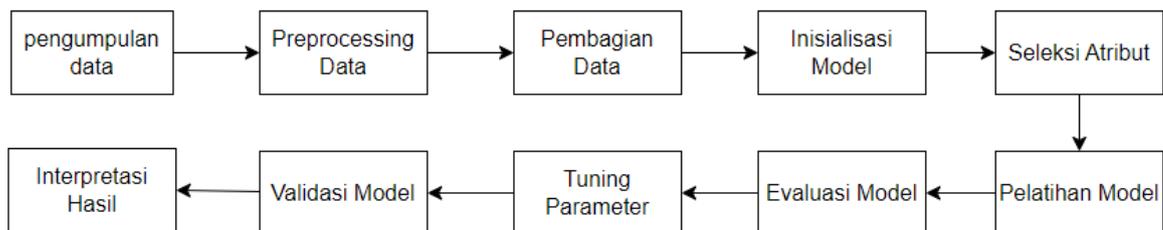
Fungsi utama dari data mining adalah untuk klasifikasi, klustering, asosiasi, prediksi dan estimasi. Dalam penerapannya, seringkali data mining menghadapi banyak kendala, antara lain jumlah data yang tidak seimbang, dimensi data yang terlalu banyak dan jumlah atribut data yang besar. Agar data mining bisa menghasilkan akurasi atau model yang lebih baik lagi dan lebih cepat, maka ada beberapa tahapan yang bisa dilalui. Salah satunya yaitu *optimize selection*. Salah satu metode *optimize selection* adalah *backward elimination* yang merupakan salah satu teknik pengolahan data awal. Dalam teknik ini akan dilakukan seleksi terhadap beberapa *feature* dari dataset. Seleksi ini diperlukan untuk memilih *feature* yang dianggap relevan dan menghapus *feature* yang dianggap kurang relevan. Agar seleksi *feature* bisa menghasilkan dataset yang terbaik maka perlu dilakukan optimasi. Setelah melalui pengolahan data awal, kemudian data akan diolah lebih lanjut. Pada penelitian ini akan diusulkan metode K-NN.

Ruang lingkup dalam penelitian implementasi teknik data mining khususnya dalam bidang *feature selection* ini, dibatasi pada:

- Sumber data atau dataset untuk penelitian ini, diperoleh dari database Dinas Pertanian Kabupaten Probolinggo pada tahun 2019.
- Pendekatan dalam analisis data dalam penelitian ini akan menggunakan metode K-NN berbasis *backward elimination*.

Dalam penelitian ini, dataset yang akan digunakan berasal dari database Dinas Pertanian Kabupaten Probolinggo tahun 2011 yang mempunyai 1 atribut khusus dan 9 atribut reguler. Adapun atributnya yaitu : varietas, panjang, bentuk, warna, rasa, teknik penanaman padi, musim, hama, ph dan mutu. Metode yang akan digunakan dalam penelitian ini yaitu K-NN berbasis *backward elimination*.

## 2. METODE PENELITIAN



Gambar 1. Tahapan Penelitian

- Pengumpulan Data**  
Kumpulkan data yang relevan untuk penelitian ini. Data dapat mencakup atribut-atribut yang digunakan untuk mengklasifikasikan mutu beras, seperti warna, ukuran, kekerasan, kelembutan, dan lain-lain. Pastikan data yang dikumpulkan sudah lengkap dan representatif.
- Preprocessing Data**  
Lakukan preprocessing data untuk memastikan data yang dikumpulkan berkualitas. Langkah-langkah preprocessing yang umum dilakukan antara lain membersihkan data dari nilai-nilai yang hilang atau tidak valid, menormalkan skala data, dan menghilangkan atribut yang tidak relevan atau redundan.
- Pembagian Data**  
Bagi data yang dikumpulkan menjadi dua bagian, yaitu data pelatihan (training data) dan data pengujian (testing data). Data pelatihan digunakan untuk melatih model K-NN, sedangkan data pengujian digunakan untuk menguji kinerja model yang telah dilatih.
- Inisialisasi Model**  
Inisialisasikan model K-NN dengan memilih nilai K (jumlah tetangga terdekat yang akan digunakan) dan menghitung jarak antara setiap sampel data dalam data pelatihan.
- Seleksi Atribut**  
Lakukan seleksi atribut menggunakan metode Backward Elimination untuk memilih subset atribut yang paling relevan dalam mengklasifikasikan mutu beras. Metode ini melibatkan pengujian secara berulang, di mana atribut-atribut yang tidak signifikan secara bertahap dihapus dari model, dan performa model diukur setiap kali atribut dihapus. Atribut yang memiliki kontribusi rendah atau tidak signifikan akan dieliminasi.

- F. Pelatihan Model  
Latih model K-NN menggunakan data pelatihan yang telah disiapkan. Model akan belajar dari pola-pola yang terdapat dalam data pelatihan untuk melakukan klasifikasi.
- G. Evaluasi Model  
Evaluasi kinerja model menggunakan data pengujian. Hitung akurasi, presisi, recall, atau metrik evaluasi lainnya untuk mengukur seberapa baik model K-NN dalam mengklasifikasikan mutu beras.
- H. Tuning Parameter  
Jika diperlukan, lakukan tuning parameter untuk mencari nilai K yang optimal atau mengoptimalkan parameter lain dalam model K-NN. Hal ini dapat dilakukan dengan menggunakan teknik cross-validation atau teknik optimasi lainnya.
- I. Validasi Model  
Validasi model K-NN yang telah dilatih dengan menggunakan data yang belum pernah dilihat sebelumnya. Hal ini untuk memastikan bahwa model yang dihasilkan dapat digeneralisasi dengan baik dan mampu melakukan klasifikasi mutu beras dengan akurasi yang tinggi.
- J. Interpretasi Hasil  
Analisis hasil yang diperoleh dari model K-NN dan interpretasikan untuk mendapatkan wawasan tentang atribut-atribut yang paling berpengaruh dalam mengklasifikasikan mutu beras.

### 3. HASIL DAN PEMBAHASAN

- A. Pengumpulan Data  
Data dalam penelitian ini adalah data macam-macam variant beras yang diperoleh dari database Dinas Pertanian Kabupaten Probolinggo pada tahun 2019. Dataset ini terdiri dari 4952 data dengan 10 atribut, yaitu :
  1. Varietas; merupakan *regular atribut* yang berisi pengelompokan jenis beras. Di sini dikelompokkan berdasarkan beras-hitam, beras-merah, ciheran, IR-64, mi-kongga dan pandan-wangi.
  2. Panjang; merupakan *regular atribut* yang berisi informasi panjang beras. Panjang beras antara 6.0 hingga 7.8.
  3. Bentuk; merupakan *regular atribut* yang berisi informasi bentuk beras. Bentuk beras ada yang bulat, ramping dan sedang.
  4. Warna; merupakan *regular atribut* yang berisi informasi warna beras. Warna beras coklat, hitam, merah dan putih.
  5. Rasa; merupakan *regular atribut* yang berisi informasi rasa beras. Rasa beras ada yang pulen dan sangat pulen.
  6. Teknik; merupakan *regular atribut* yang berisi informasi teknik penanaman beras. Teknik penanaman beras berupa SRI dan jajar-legowo.
  7. Musim; merupakan *regular atribut* yang berisi informasi musim, yaitu hujan dan kemarau.
  8. Hama; merupakan *regular atribut* yang berisi informasi hama. Hama beras ada yang berupa burung, penggerek batang, tikus, wereng coklat dan wereng hijau.
  9. PH; merupakan *regular atribut* yang berisi informasi tingkat keasaman. PH beras antara 5.2 hingga 6.5.
  10. Mutu; merupakan *special atribut* atau menjadi *label* yang berisi pengelompokan kualitas beras. Di sini dikelompokkan berdasarkan grade-A, grade-B, grade-C dan grade-D.
- B. Penghitungan K-NN secara umum
  1. Menghitung Jarak
  2. Memilih Tetangga Terdekat
  3. Menentukan Kelas Prediksi :
    - a. Dalam K-NN klasifikasi, kelas prediksi  $y$  akan ditentukan berdasarkan mayoritas kelas dari tetangga terdekat:  $y = \text{argmax}(\sum I(y_i = c))$ , di mana  $I$  adalah fungsi indikator yang bernilai 1 jika kondisi di dalamnya terpenuhi, 0 jika tidak, dan  $c$  adalah kelas yang mungkin
    - b. Dalam K-NN regresi, nilai prediksi  $y$  akan diestimasi berdasarkan nilai rata-rata atau median dari target  $K$  tetangga terdekat:  $y = \text{mean}(y_1, y_2, \dots, y_K)$  atau  $y = \text{median}(y_1, y_2, \dots, y_K)$  Di sini,  $y_i$  adalah nilai target dari tetangga terdekat  $x_i$ .
- C. Konversi Data  
Berikut adalah atribut dan *value* yang memiliki nilai polynomial:
  1. Varietas: IR-64, beras-merah, mi-kongga, ciheran, pandan-wangi, dan beras-hitam
  2. Panjang: berisi data numerik

3. Bentuk: ramping, sedang, bulat
  4. Warna: putih, merah, coklat, dan hitam
  5. Rasa: pulen dan sangat-pulen
  6. Teknik: sri dan jajar-legowo
  7. Musim: hujan dan kemarau
  8. Hama: tikus, wereng-coklat, penggerek-batang, burung, wereng-hijau
  9. PH: berisi data numeric
  10. Mutu: sebagai label dataset
- D. Optimize selection berbasis *backward elimination*
1. Tahap pertama melakukan regresi dengan menggunakan semua atribut
  2. Tahap kedua dengan menggunakan 8 atribut karena atribut rasa sudah dieliminasi
  3. Tahap ketiga dengan menggunakan 7 atribut karena atribut panjang sudah dieliminasi. Langkah-langkahnya sama dengan tahap kedua

Tabel 1 Contoh dataset setelah proses *backward elimination*

	A	B	C	D	E	F	G	H
1	Varietas	Bentuk	Warna	Teknik	Musim	Hama	PH	Mutu
2	3	1	1	2	1	3	5.2	Grade-C
3	5	1	3	2	1	4	5.2	Grade-C
4	4	1	1	1	1	3	5.2	Grade-D
5	3	1	1	1	1	4	5.2	Grade-D
6	6	3	1	1	1	3	5.2	Grade-C
7	1	2	4	2	1	2	5.2	Grade-D
8	3	1	1	2	2	3	5.5	Grade-C
...	...	...	...	...	...	...	...	...
4953	4	1	1	2	2	4	5.5	Grade-B

- E. Klasifikasi dengan metode KNN
- 1) Menentukan ranking tiap atribut
  - 2) Menentukan bobot tiap atribut
  - 3) Menentukan ranking *value* tiap atribut
  - 4) Menentukan bobot tiap *value*
  - 5) Menghitung kedekatan nilai antar *value*
  - 6) Dengan tahapan yang sama seperti tahap ke-3 dan ke-4, maka akan didapatkan ranking *value* tiap atribut dan bobot *value* tiap atribut yang lain
  - 7) Dengan tahapan yang sama seperti tahap ke-5, maka akan didapatkan kedekatan nilai antar *value* tiap atribut yang lain

F. Validasi Hasil Penelitian

Untuk memvalidasi hasil penelitian, digunakan *10-fold cross validation*. Yaitu dengan membagi 4952 dataset ke dalam 10 bagian, yaitu 9 bagian dataset training sebanyak 90% atau 4457 dan 1 bagian dataset testing sebanyak 10% atau 495.

Tabel 2 Data training sebanyak 90%

No	Varietas	Bentuk	Warna	Teknik	Musim	Hama	PH
1	3	1	1	2	1	3	5.2
2	5	1	3	2	1	4	5.2
3	4	1	1	1	1	3	5.2
4	3	1	1	1	1	4	5.2
.....	.....	.....	.....	.....	.....	.....	.....
4457	5	2	3	1	1	4	6.5

Tabel di atas merupakan dataset yang digunakan sebagai data training. Data ini berjumlah 4457 dataset atau sejumlah 90% dari keseluruhan dataset. Pada prosesnya nanti tiap *value* dari masing-masing variabel pada data training akan dihitung kedekatannya dengan *value* dari masing-masing variabel yang ada pada data testing.

Tabel 3 Data testing sebanyak 10%

No	Varietas	Bentuk	Warna	Teknik	Musim	Hama	PH
4458	4	1	1	1	1	4	6.5
4459	5	1	3	1	2	2	6.5
4460	2	1	2	2	2	1	6.5
.....	.....	.....	.....	.....	.....	.....	.....
4953	4	1	1	2	2	4	5.5

Tabel di atas merupakan dataset yang digunakan sebagai data testing. Data ini berjumlah 495 dataset atau sejumlah 10% dari keseluruhan dataset. Pada prosesnya nanti tiap *value* dari masing-masing variabel pada data training akan dihitung kedekatannya dengan *value* dari masing-masing variabel yang ada pada data testing.

G. Confusion Matrix Hasil Penelitian

Tabel 4 Confusion matrix backward elimination dan K-NN

	True Grade-C	True Grade-D	True Grade-B	True Grade-A	class precision
pred. Grade-C	2132	0	177	1	92.29%
pred. Grade-D	147	766	11	3	82.63%
pred. Grade-B	443	10	1216	46	70.90%
pred. Grade-A	0	0	0	0	0.00%
class recall	78.32%	98.71%	86.61%	0.00%	

Dari tabel di atas bisa disimpulkan bahwa hasil komputasi metode *backward elimination* dan K-NN adalah sebagai berikut:

- a. Nilai *True Positive* (TP) untuk Grade-A adalah 2132, Grade-B adalah 766, Grade-C adalah 1216 dan Grade-D adalah 0. Dan nilai selain TP disebut dengan *False Negative* (FN).
- b. Sedangkan *class recall* merupakan kolom yang berisi besar nilai klasifikasi yang tepat. Misalnya *class recall* dari Grade-C yang tepat pengklasifikasiannya adalah sebagai berikut:

$$TP = \frac{TP}{TP+FN} = \frac{2312}{2312+147+443+0} = \frac{2312}{2722} = 0.7832 = 78.32\% \tag{1}$$

- c. Sedangkan *class precision* merupakan baris yang berisi besar nilai klasifikasi yang tepat. Misalnya *class precision* dari Grade-C yang tepat pengklasifikasiannya adalah sebagai berikut:

$$TP = \frac{TP}{TP+FN} = \frac{2312}{2312+0+177+1} = \frac{2312}{2310} = 0.9229 = 92.29\% \tag{2}$$

- d. Sedangkan *accuracy* merupakan persentase hasil klasifikasi yang benar yang bisa didapatkan dengan cara berikut:

$$Acc = \frac{TP+TN}{TP+FN+FP+FN} = \frac{2132+766+1216+0}{2132+147+443+0+0+766+10+0+177+11+1216+0+0+1+3+46+0} = \frac{4114}{4952} = 0.8308 = 83.08\% \tag{3}$$

H. Uji Beda (*t-test*)

Secara keseluruhan, hasil percobaan metode K-NN dan penambahan metode *backward elimination* didapatkan persentase akurasi hasil sebagai berikut:

Tabel 5 hasil percobaan metode K-NN dan penambahan metode *backward elimination*

Metode	Jumlah Data		
	4952	2000	1000
K-Means + OWA	35.80	37.45	45.52
K-NN + <i>Backward Elimination</i>	83.08	87.70	98.60

Tabel di atas merupakan nilai akurasi dari dua metode yaitu K-NN dan K-NN yang dikombinasikan dengan *backward elimination*. Kemudian nilai akurasi dari dua metode di atas digunakan sebagai acuan untuk menentukan kuadrat selisih metode pertama dan kedua.

Tabel 6 acuan menentukan kuadrat selisih

K-MEANS + OWA (x1)	K-NN + BE (x2)	D = x1-x2	D <sup>2</sup>
35.8	83.08	-47.28	2235.3984
37.45	87.7	-50.25	2525.0625
45.52	98.6	-53.08	2817.4864
Jumlah		-150.61	22683.3721

Hasil dari tabel di atas kemudian digunakan untuk menentukan standar deviasi. Standar deviasi ini nantinya akan digunakan untuk mengetahui perbedaan antar metode pertama dan kedua.

#### I. Membandingkan Akurasi Hasil Pengujian Dengan Metode Lain

Tabel 7 Perbandingan Akurasi Hasil Pengujian Dengan Metode Lain

No	Metode	Akurasi
1	<i>K-Means</i> berbasis OWA dengan dataset varietas padi	96.67%
2	<i>K-Means</i> berbasis OWA dengan 4952 dataset mutu beras	35.93%
3	<i>K-Means</i> berbasis OWA dengan 1000 dataset mutu beras	45.52%
4	K-NN k=9 dengan 4952 dataset mutu beras	82.17%
6	<i>Backward elimination</i> + K-NN k=1 dengan 4952 dataset	83.08%
7	<i>Backward elimination</i> + K-NN k=1 dengan 2000 dataset	87.70%
8	<i>Backward elimination</i> + K-NN k=1 dengan 1000 dataset	98.60%

Dari perbandingan tabel di atas, penelitian yang dilakukan oleh Ulya menggunakan metode *K-Means* berbasis OWA untuk klustering dataset varietas padi menghasilkan akurasi sebesar 96.67%. namun setelah diterapkan pada 4952 dataset mutu beras akurasi yang dihasilkan menurun menjadi 35.93%.

Dan juga bisa disimpulkan bahwa *backward elimination* bisa meningkatkan akurasi dengan cukup baik. Terbukti pada metode K-NN dengan k=9 yang menghasilkan akurasi sebesar 82.17% meningkat menjadi 83.08% untuk dataset sebanyak 4952.

Sedangkan untuk 2000 dataset, metode yang menghasilkan akurasi paling baik adalah kombinasi *backward elimination* dan kNN pada saat k=1, yaitu sebesar 87.70%. Dan untuk 1000 dataset, metode yang menghasilkan akurasi paling baik adalah kombinasi *backward elimination* dan kNN pada saat k=1, yaitu sebesar 98.60%.

#### 4. KESIMPULAN

Dari hasil analisa dan komputasi beberapa metode, dapat disimpulkan bahwa metode *K-NN* berbasis *backward elimination* yang digunakan untuk meningkatkan akurasi klasifikasi beras menghasilkan:

- 1) Atribut yang berpengaruh terhadap penentuan mutu beras adalah varietas, bentuk, warna, teknik, musim, hama dan PH.
- 2) Penggunaan metode *K-Means* berbasis OWA untuk klastring varietas padi memiliki akurasi 96.67%. Namun ketika metode ini diterapkan pada dataset mutu beras untuk data sebanyak 4952, akurasinya adalah 35.80%, sedangkan untuk data sebanyak 2000, akurasinya adalah 37.45%, sedangkan untuk data sebanyak 1000, akurasinya adalah 45.52%.
- 3) Penggunaan metode K-NN berbasis *backward elimination* pada dataset mutu beras untuk data sebanyak 4952, akurasinya adalah 83.08% saat k=1, sedangkan untuk data sebanyak 2000, akurasinya adalah 87.70% saat k=1, sedangkan untuk data sebanyak 1000, akurasinya adalah 98.60% saat k=1.
- 4) Dan berdasarkan uji beda *t-test*,  $t_{hitung} < t_{tabel}$ , sehingga  $H_0$  (hipotesis nihil) ditolak dan  $H_a$  (hipotesis alternatif) diterima yang artinya pemilihan metode pertama dan metode kedua memiliki pengaruh peningkatan akurasi yang signifikan. Sehingga metode K-NN berbasis *backward elimination* memiliki peningkatan akurasi yang signifikan dibandingkan dengan metode *K-Means* berbasis OWA

#### DAFTAR PUSTAKA

- [1] S. M. Boker, R. G. Moulder, and G. R. Sjobeck, "Constrained Fourth Order Latent Differential Equation Reduces Parameter Estimation Bias for Damped Linear Oscillator Models," *Struct. Equ. Model.*, vol. 27, no. 2, pp. 202–218, 2020.
- [2] A. S. Dabaghi, M. H. Khoshtaghaza, M. R. Alizadeh, and H. Zareiforoush, "Identification the appearance

- quality of rice kernels by vision technology and neural network classifier,” *Int. J. Eng. Technol.*, vol. 9, p. 118, 2020.
- [3] D. Setian and I. Seprina, “ANALISIS SENTIMEN MASYARAKAT TERHADAP DATA TWEET LAZADA MENGGUNAKAN TEXT MINING DAN ALGORITMA NAIVE BINA DARMA CONFERENCE ON COMPUTER SCIENCE,” *Bina Darma Conf. Comput. Sci.*, pp. 998–1004, 2019.
- [4] S. H. H. Ilyas, St. Hajrah Mansyur, and Harlinda, “Analysis of Learning Web Application Testing With Weighting Product and Ordered Weighting Averaging on Group Decision Maker,” *J. Tek. Inform.*, vol. 3, no. 6, pp. 1513–1522, 2022.
- [5] S. Sucipto, “Analisa Hasil Rekomendasi Pembimbing Menggunakan Multi-Attribute Dengan Metode Weighted Product,” *Fountain Informatics J.*, vol. 2, no. 1, p. 27, 2017.
- [6] E. Rahm and H. Do, “Data cleaning: Problems and current approaches,” *IEEE Data Eng. Bull.*, vol. 23, no. 4, pp. 3–13, 2020.
- [7] A. R. F. M. L. M. F. Amrulloh, “Klasifikasi jenis tanaman kelengkeng berdasarkan ciri tekstur daun menggunakan metode Adaptive Neuro Fuzzy Inference System (AFIS),” *J. Explor. IT*, vol. 14, no. 36, pp. 29–38, 2022.
- [8] M. Faishol Amrulloh, Moch. Lutfi, and W. Agung Nurcahyo, “Deep Pre-Trained Multi Model Convolution Neural Network Untuk Diagnosa Covid 19 Pada Citra Rontgen Dada,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 649–654, 2022.
- [9] Bhavesh B. Prajapati and Sachin Patel, "Algorithmic Approach to Quality Analysis of Indian Basmati Rice Using Digital Image Processing," *International Journal of Emerging Technology and Advanced Engineering*, vol. 3, no. 3, pp. 503-504, March 2013.
- [10] IBM Corporation, *IBM SPSS Modeler CRISP-DM Guide*.: IBM Corporation, 2011.
- [11] Zenon Gniazdowski and Michal Grabowski, "Numerical coding of nominal data," in *Zeszyty Naukowe WWSI*, Warsaw, 2015, pp. 53-61.
- [12] Choudhary and Nidhi, "A Study over Problems and Approaches of Data Cleansing/Cleaning," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 4, no. 2, pp. 774-779, February 2014.
- [13] Mengchao Wang, Jonathan Wright, Richard Buswell, and Alexander Brownlee, "A comparison of approaches to stepwise regression for global sensitivity analysis used with evolutionary optimization," in *13th Conference of International Building Performance Simulation Association*, Chambéry, France, August 2013, pp. 2551-2558.