

Klasifikasi Kualitas Biji Kopi Robusta Dengan Metode Naive Bayes

I Kadek Nurcahyo Putra, I Gede Aris Gunadi, I Made Gede Sunarya
Universitas Pendidikan Ganesha

Jl. Udayana No. 11, Singaraja, Buleleng, Bali, telp (0362) 22570
nurcahyo@undiksha.ac.id, igedearisgunadi@undiksha.ac.id,
sunarya@undiksha.ac.id

Abstract

Sorting the quality of coffee beans is a very important process to maintain and improve the quality of production, seeing coffee as one of the most important commodities traded. The author wants to minimize subjective human misclassification by implementing the Naive Bayes method to classify coffee bean quality objectively. The coffee beans are photographed to produce an image of the coffee beans, the HSV color space is used to extract the color features of the beans, the texture of the coffee beans is extracted using the GLCM method, and the size of the coffee beans is calculated by accumulating the pixel values in the binary image object. Testing the classification model built with 480 training data produces an accuracy of 90.8% based on 120 test data. The accuracy results show that when there are 120 test data, 11 data are classified incorrectly or when there are 200 test data, 19 to 20 seeds will be misclassified by the model. Further research is suggested to conduct deeper observations to obtain features that can represent differences in seed quality more representatively, as well as to compare the Naive Bayes classification method with other classification methods to obtain a better classification model in the future.

Keywords: Coffee, Classification, Naive Bayes

Abstrak

Sortasi Kualitas biji kopi merupakan proses yang sangat penting untuk menjaga serta meningkatkan kualitas produksi, melihat kopi sebagai salah satu komoditas paling penting yang diperjual belikan. Penulis ingin meminimalisir kesalahan klasifikasi oleh manusia yang subjektif dengan mengimplementasi metode Naive Bayes untuk melakukan klasifikasi kualitas biji kopi secara objektif. Biji kopi difoto sehingga menghasilkan citra biji kopi, ruang warna HSV digunakan untuk melakukan ekstraksi ciri warna biji, tekstur biji kopi diekstrak dengan metode GLCM, dan ukuran biji kopi dihitung dengan mengakumulasi nilai piksel pada objek citra biner. Pengujian terhadap model klasifikasi yang dibangun dengan 480 data latih menghasilkan akurasi 90.8% berdasarkan 120 data uji. Hasil akurasi menunjukkan ketika ada 120 data uji maka 11 data salah diklasifikasi atau ketika ada 200 data uji maka 19 hingga 20 biji akan salah diklasifikasi oleh model. Penelitian selanjutnya disarankan untuk melakukan pengamatan yang lebih dalam untuk mendapatkan fitur ciri yang dapat merepresentasikan perbedaan kualitas biji dengan lebih representative, serta membandingkan metode klasifikasi Naive Bayes dengan metode klasifikasi lain untuk mendapatkan model klasifikasi yang lebih baik di masa depan.

Kata kunci: Kopi, Klasifikasi, Naive Bayes

1. PENDAHULUAN

Komoditas kopi adalah salah satu mata pencaharian utama bagi petani di Indonesia. Indonesia merupakan salah satu penghasil kopi terbesar di dunia [1]. Lebih dari dua juta gelas kopi diminum setiap hari di dunia, maka dapat dikatakan bahwa kopi adalah salah satu komoditas paling penting yang diperjual belikan [2]. Saat ini konsumsi dan permintaan untuk biji kopi berkualitas terus naik setiap tahunnya [3]. Karena itu, pemilahan biji kopi sesuai kualitas menjadi fokus yang sangat penting untuk menentukan harga, kestabilan penyimpanan, dan untuk memenuhi permintaan pasar [4].

Biji kopi umumnya dipilah dengan melihat warna biji, tekstur, bentuk, dan ukuran biji kopi [5]. Pada penelitian ini penulis berfokus pada klasifikasi kualitas biji kopi. Pemilahan biji kopi sesuai kualitas sangat penting dilakukan untuk menjaga kualitas, serta mengoptimalkan harga jual.

Penelitian oleh [6] melakukan klasifikasi kerusakan biji kopi menggunakan metode Naive Bayes. Klasifikasi dilakukan pada empat jenis kerusakan yakni biji pecah, biji hitam penuh, biji hitam sebagian, dan biji hitam pecah. Peneliti menggunakan empat fitur yaitu Red, Green, Blue, dan Diameter. Dengan menggunakan 100 data biji kopi penelitian ini memperoleh hasil akurasi 82,5%.

Klasifikasi kualitas biji kopi robusta oleh [7] dengan 1300 gambar biji kopi dari empat kelas yakni biji baik, sekam, gigitan serangga, dan tidak diketahui. Menggunakan CNN memperoleh akurasi 93,33%. Selanjutnya penelitian oleh [8] melakukan klasifikasi terhadap 7 grade kualitas biji kopi, RGB dan HSI digunakan sebagai fitur warna, GLCM digunakan sebagai metode ekstraksi fitur tekstur biji kopi. Menggunakan Fuzzy Logic untuk klasifikasi memperoleh akurasi 92.85%.

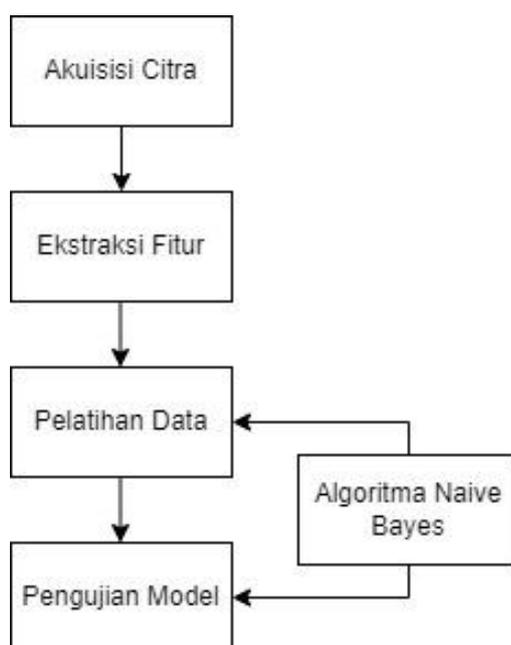
Pada penelitian ini biji kopi yang digunakan sebagai set data yakni biji kopi besar bersih tanpa cacat, biji kopi yang memiliki cacat berlubang, hitam, bertutul, memudar, serta biji kopi pecah. Biji kopi diperoleh dari perusahaan CV. Kaki Lima Solid pengolah kopi didaerah Kabupaten Tabanan, Provinsi Bali.

Kualitas biji kopi akan diklasifikasi berdasarkan warna, tekstur, dan ukuran. Biji kopi besar, bersih tanpa cacat dikategorikan sebagai produk unggulan kualitas A (sangat baik) yang disiapkan untuk pasar ekspor kedepannya, biji kopi berukuran kecil, bersih tanpa cacat dikategorikan sebagai produk kualitas B (baik), dan Biji kopi cacat berlubang, hitam, bertutul, serta memudar dikategorikan sebagai produk C (kurang baik) yang akan diproses sebagai kopi bubuk.

Warna biji kopi akan diekstrak pada ruang warna HSV. Warna HSV merupakan ruang warna yang lebih kokoh terhadap variasi cahaya putih dibandingkan RGB. Metode Gray Level Co-Ocurrence Matrix (GLCM) digunakan untuk mengekstrak fitur tekstur pada biji kopi. Ukuran biji kopi akan diekstrak dengan mengakumulasi nilai piksel pada objek citra biji. Setelah biji kopi diekstrak fiturnya, metode klasifikasi Naive Bayes digunakan untuk melakukan klasifikasi kualitas biji kopi.

2. METODOLOGI PENELITIAN

Penelitian ini terdiri dari empat tahapan yang digambarkan pada Gambar 1, meliputi akuisisi data citra biji kopi, ekstraksi fitur untuk menghasilkan dataset, pelatihan data, dan pengujian terhadap model klasifikasi yang dibangun dengan pelatihan data.



Gambar 1. Alur Penelitian Klasifikasi Kualitas Biji Kopi Robusta

2.1. Akuisisi Citra

Biji kopi yang diperoleh dari produsen kopi berjumlah 300 gram biji untuk setiap kualitas. Biji yang diperoleh telah dipilah sesuai kualitas biji oleh penyortir kopi, meliputi biji kualitas A yakni biji kopi hijau berukuran besar, tanpa cacat, biji kualitas B adalah biji yang bersih tanpa warna hitam, memudar, pecah, hitam namun lebih kecil dari ukuran biji kualitas A, dan biji kualitas C memiliki cacat seperti pecah, berlubang, berwarna hitam, serta memudar.

Biji kopi selanjutnya akan difoto sehingga menghasilkan citra biji kopi. Citra biji kopi akan disimpan dan dibagi menjadi dua set data, 80% dari data akan dijadikan sebagai data latih untuk membangun model klasifikasi, 20% sisanya akan dijadikan sebagai data uji.

2.2. Ekstraksi Fitur

Citra biji kopi akan di pre-proses terlebih dahulu sehingga siap untuk diekstrak ciri warna, tekstur, dan ukurannya. Citra asli RGB hasil akuisisi data akan dikonversi terlebih dahulu menjadi citra HSV, dan citra aras keabuan (grayscale). Citra HSV akan diekstrak fiturnya sehingga menghasilkan tiga fitur warna yakni Hue, Saturation, Value. GLCM membangun matriks kookurensi untuk mengekstrak fitur tekstur Energi, homogeneiti, kontras, serta korelasi pada citra grayscale. Citra aras keabuan akan dikonversi menjadi citra biner untuk ekstraksi luas dan keliling biji kopi.

2.3. Metode Klasifikasi Naive Bayes

Metode klasifikasi Naive Bayes merupakan salah satu metode *machine learning* yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya [9]. Kualitas biji kopi diprediksi dengan menghitung probabilitas terjadinya suatu kualitas ketika serangkaian evidence muncul pada biji kopi [10]. Algoritma Naive Bayes hanya membutuhkan data yang kecil dalam melakukan klasifikasi, karena dalam perhitungannya hanya membutuhkan rata-rata serta deviasi dari data latih. Walau tampak sederhana Naive Bayes sering bekerja melebihi harapan pada banyak situasi dunia nyata [11].

Pada penelitian ini dataset berupa data kontinu, maka perhitungan probabilitas terjadinya suatu kondisi cacat biji kopi didapat dari sekumpulan probabilitas evidence yang muncul pada data. Probabilitas evidence dihitung dengan rumus pendistribusian probabilitas atau distribusi normal (gaussian).

$$P(E_n) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

Keterangan:

- $P(E_n)$ = Probabilitas Evidence ke-n
- X = Nilai Evidence fitur

Setelah mendapat nilai probabilitas dari setiap evidence, nilai probabilitas terjadinya kualitas biji kopi diperoleh dengan rumus:

$$P(H|E) = P(E_1) * P(E_2) * \dots * P(E_n) * P(H) \quad (2)$$

Keterangan:

- $P(H|E)$ = Probabilitas Kualitas Biji Kopi
- $P(H)$ = Probabilitas awal Kualitas Biji Kopi

2.3.1 Pelatihan Data

Data latih berjumlah 480 data akan ditraining menggunakan algoritma Naïve Bayes. Naive Bayes akan membangun model klasifikasi yang mempelajari pola pada data latih, sehingga ketika suatu data baru masuk, model dapat melakukan klasifikasi terhadap data tersebut berdasarkan pola atau pengalaman sebelumnya berdasarkan data latih.

2.3.2 Pengujian Model Klasifikasi

Data uji digunakan untuk mengetahui performa terhadap model klasifikasi Naive Bayes yang dibangun. Akurasi yang merepresentasikan berapa persen suatu algoritma dapat memprediksi secara benar dari 120 data yang diuji. Pada multi-class classification problem, akurasi dihitung dengan menjumlahkan data asli yang diprediksi dengan benar lalu dibagi dengan jumlah data uji.

$$\text{Akurasi} = \frac{TP}{\text{Jumlah Data Uji}} \quad (3)$$

Keterangan: TP = Data asli yang diprediksi benar

3. HASIL DAN PEMBAHASAN

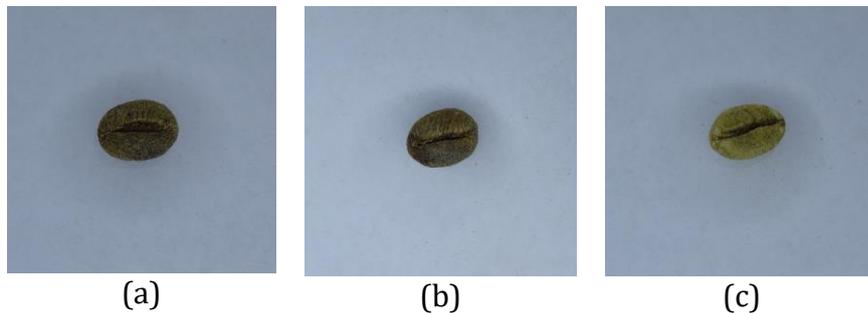
3.1. Akuisisi Citra

Penulis memperoleh 300 gram biji kopi dari setiap kualitas dari pengolah kopi. Pemilahan kualitas biji dibantu oleh pengolah kopi, dari 300 gram setiap kualitas biji didapat, biji kualitas A berjumlah 202, 241 biji kualitas B, dan 236 biji kopi kualitas C. Biji yang digunakan sebagai set data sejumlah 600 biji atau 200 biji untuk setiap kualitas hal ini dilakukan agar set data memiliki jumlah yang seimbang.

Biji kopi diletakkan satu persatu diatas alas berwarna putih, selanjutnya di foto menggunakan kamera digital dengan bantuan pencahayaan ringlight. Jarak objek biji kopi dengan kamera adalah 15 cm, proses foto biji kopi dilakukan sama untuk semua biji kopi agar mendapatkan data citra yang representative.



Gambar 2. Akuisisi Citra Biji Kopi Robusta

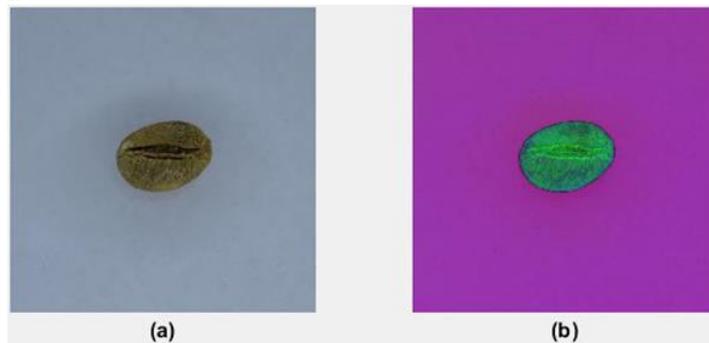


(a) (b) (c)
Gambar 3. Citra Biji Kopi Robusta, (a) Biji Kualitas A, (b) Biji Kualitas B, (c) Biji Kualitas C

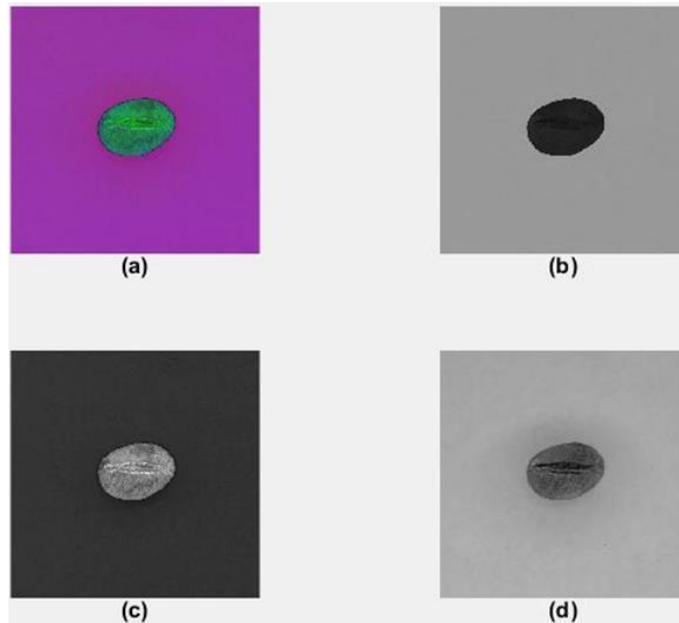
Citra biji kopi yang telah didapat dibagi menjadi dua set data, 80% dari jumlah data dijadikan data latih dan 20% sisanya digunakan sebagai data uji.

3.2. Ekstraksi Fitur

Penulis melakukan beberapa proses sebelum citra dapat diekstrak fiturnya, citra asli RGB di dikonversi menjadi citra HSV dan citra grayscale. Citra grayscale diubah menjadi citra biner untuk ekstraksi ukuran biji. Ekstraksi fitur warna pada ruang warna HSV dimulai dengan memisahkan tiap channel pada ruang warna HSV. Nilai piksel pada objek biji di tiap channel H, S, V diakumulasikan selanjutnya dibagi dengan jumlah piksel pada objek biji untuk mendapatkan nilai rata-rata dari fitur Hue, Saturation, dan Value. Ekstraksi fitur warna pada citra HSV menghasilkan tiga fitur untuk setiap citra biji kopi.



Gambar 4. Konversi Citra RGB ke Citra HSV, (a) Citra RGB, (b) Citra HSV



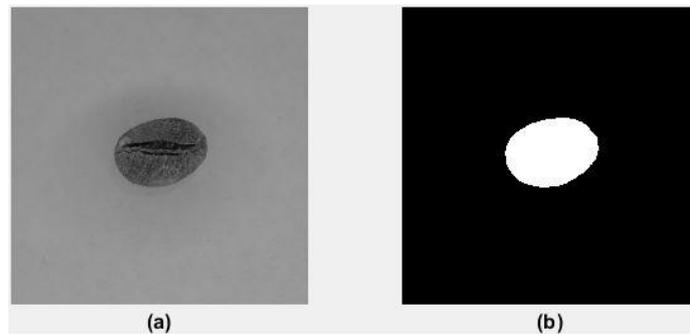
Gambar 5. Citra HSV dan Ketiga Chanel pembentuk Citra HSV, (a) HSV, (b) Hue, (c) Saturasi, (d) Value

Ekstraksi fitur tekstur dengan metode GLCM dimulai dengan membangun matriks kookurensi pada citra grayscale. Penulis menggunakan jarak 1 piksel dengan keempat orientasi arah metode GLCM. Fitur energi, kontras, homogeneiti, dan korelasi menggunakan metode GLCM dengan empat orientasi arah menghasilkan empat nilai untuk setiap fitur, namun akan diambil nilai rata-rata untuk penelitian ini, sehingga ciri tekstur pada biji kopi memiliki empat fitur.



Gambar 6. Konversi Citra RGB menjadi Citra Grayscale, (a) Citra RGB, (b) Citra Grayscale

Ukuran biji kopi diekstrak dengan mengakumulasi nilai piksel pada objek citra biner biji kopi. Luas dihitung dengan mengakumulasi semua nilai piksel pada objek. Nilai piksel pada boundary objek citra biji kopi diakumulasi untuk mendapat keliling citra biji kopi. Ekstraksi ukuran biji menghasilkan dua fitur.



Gambar 7. Konversi Citra Grayscale menjadi Citra Biner, (a) Citra Grayscale, (b) Citra Biner

3.3. Pelatihan Data

Data latih ditraining menggunakan algoritma naive bayes bertujuan membangun model klasifikasi. Naive bayes mempelajari data serta pola suatu kualitas biji kopi terjadi ketika sekumpulan evidence fitur muncul, sehingga ketika suatu saat ada data yang baru masuk, naive bayes mampu memprediksi data yang masuk termasuk kelas biji kualitas A, kualitas B, atau Kualitas C.

3.4. Pengujian Model

Model klasifikasi yang telah dibangun menggunakan algoritma naive bayes berdasarkan 480 data latih diuji untuk mengetahui performa model. Confusion matriks digunakan untuk mengevaluasi model klasifikasi. Akurasi yang merepresentasikan berapa persen suatu metode dapat memprediksi secara benar digunakan sebagai evaluator dalam penelitian ini. Pada multi-class confusion matriks, akurasi dihitung dengan menjumlahkan True Positive (TP) atau data pada kelas asli yang diprediksi dengan benar dibagi dengan jumlah data uji.

		Kelas Prediksi			TOTAL
		Kelas A	B	C	
Kelas Asli	A	38	0	2	40
	B	0	37	3	40
	C	5	1	34	40
	TOTAL	43	38	39	120

Gambar 8. Confusion Matriks Model Naive Bayes

Gambar 8 memperlihatkan hasil dari pengujian model klasifikasi kualitas biji kopi menggunakan metode Naive Bayes. Akurasi dihitung dengan menjumlahkan true positive pada tiap kelas kualitas biji dibagi dengan jumlah data uji.

$$\text{Akurasi Model Klasifikasi Naive Bayes} = \frac{38+37+34}{120} = 0.908$$

4. SIMPULAN

Metode Naive Bayes berhasil diimplementasikan pada klasifikasi kualitas biji kopi berdasarkan ukuran biji, tekstur, dan warna. Ciri warna biji diekstrak dari ruang warna HSV, metode GLCM digunakan untuk mengekstrak ciri tekstur, dan ukuran diekstrak dengan mengakumulasi nilai piksel objek citra biner biji kopi. Hasil klasifikasi kualitas biji kopi memperoleh nilai akurasi 90.8%. Hasil akurasi 90.8% menunjukkan ketika ada 120 data uji maka 11 data salah diklasifikasi atau ketika ada 200 data uji maka 19 hingga 20 biji akan salah diklasifikasi oleh model. Penelitian selanjutnya disarankan untuk melakukan pengamatan yang lebih dalam mengenai fitur ciri yang dapat merepresentasikan perbedaan kualitas pada biji dengan lebih representative, serta membandingkan metode klasifikasi Naive Bayes dengan metode klasifikasi lain untuk mendapatkan model klasifikasi yang lebih baik di masa depan.

DAFTAR PUSTAKA

- [1] M. Saputra, M. P. Kurniawan, dan M. T. Informatika, "Identifikasi Mutu Biji Kopi Arabika Berdasarkan Cacat," *J. Teknol. Inf. dan Komun.*, vol. 10, no. 1, hal. 27–35, 2020.
- [2] D. Giacalone, T. K. Degn, N. Yang, C. Liu, I. Fisk, dan M. Münchow, "Common roasting defects in coffee: Aroma composition, sensory characterization and consumer perception," *Food Qual. Prefer.*, vol. 71, no. March, hal. 463–474, 2019, doi: 10.1016/j.foodqual.2018.03.009.
- [3] N. Bhumiratana, K. Adhikari, dan E. Chambers, "Evolution of sensory aroma attributes from coffee beans to brewed coffee," *LWT - Food Sci. Technol.*, vol. 44, no. 10, hal. 2185–2192, 2011, doi: 10.1016/j.lwt.2011.07.001.
- [4] M. Garcia, J. E. Candelo-Becerra, dan F. E. Hoyos, "applied sciences Quality and Defect Inspection of Green Coffee Beans Using a Computer Vision System," 2019.
- [5] P. Vithu dan J. A. Moses, "Machine vision system for food grain quality evaluation: A review," *Trends Food Sci. Technol.*, vol. 56, hal. 13–20, 2016, doi: 10.1016/j.tifs.2016.07.011.
- [6] J. Aramiko, "KLASIFIKASI KERUSAKAN BIJI KOPI MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER SKRIPSI," 2020.
- [7] R. E. Angelia, K. C. R. Cavan, K. E. Recto, dan R. B. Bactat, "Dried Robusta Coffee Bean Quality Classification Using Convolutional Neural Network Algorithm," *ACM Int. Conf. Proceeding Ser.*, hal. 57–61, 2021, doi: 10.1145/3467707.3467715.

- [8] P. S. Maria dan M. Rivai, "Klasifikasi Kualitas Biji Kopi Menggunakan Pengolahan Citra dan Fuzzy Logic," *Semin. Nas. Menggagas Kebangkitan Komod. Unggulan Lokal Pertan. dan Kelaut.*, 2013.
- [9] S. L. B. Ginting dan R. P. Trinanda, "TEKNIK DATA MINING MENGGUNAKAN METODE BAYES CLASSIFIER UNTUK OPTIMALISASI PENCARIAN PADA APLIKASI PERPUSTAKAAN," 2013.
- [10] T. Wahyono, *FUNDAMENTAL OF PYTHON FOR MACHINE LEARNING*. Yogyakarta: GAVA MEDIA, 2021.
- [11] S. . Pattekari dan A. Parveen, "Prediction system for heart disease using Naïve Bayes," *Int. J. Adv. Comput. Math. Sci.*, vol. 3, no. 3, hal. 290–294, 2012.