

Klasifikasi Penyakit Kardiovaskular Menggunakan Metode *Fuzzy K-Nearest Neighbor* (FKNN)

Ni Komang Santi Cahyani^{a1}, I Ketut Gede Suhartana^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana, Badung, Bali, Indonesia

¹santicaahyani51@gmail.com

²ikg.suhartana@unud.ac.id

Abstract

Cardiovascular disease can be classified as a non-communicable disease but this disease can cause death because it interferes with the function of the heart. The purpose of researching the classification of cardiovascular disease is to facilitate medical personnel in diagnosing cardiovascular disease in patients. Diagnosing cardiovascular disease early to minimize deaths caused by cardiovascular disease. In this study, the classification of cardiovascular disease uses the Fuzzy K-Nearest Neighbor (FKNN) method to determine the optimum nearest neighbor value by producing a good classification. The use of the Fuzzy K-Nearest Neighbor (FKNN) method was carried out seven times in the search for the value of K, the results obtained were the value of K, namely the value of K = 3 of 75.82%, K = 5 of 75.82%, K = 7 of 76.92%, K = 9 of 79.12%, K=11 at 78.02%, K=13 at 80.21%, and K=15 at 80.21%. From the test results obtained optimal results at the value of K = 13 and K = 15 by producing an accuracy of 80.21%, the test in this study uses the confusion matrix method.

Keywords: Cardiovascular, Classification, FKNN, Medical Personnel, Confusion Matrix

1. Pendahuluan

Penyakit kardiovaskular tergolong pada suatu penyakit yang tidak akan menular, namun penyakit kardiovaskular dapat menyebabkan kematian yang dikarenakan penyakit ini akan menyerang fungsi jantung [1]. Fungsi jantung yang diserang oleh penyakit kardiovaskular akan menyebabkan penyempitan maupun penyumbatan pembuluh darah yang menuju ke jantung. Penyebab penyakit kardiovaskular tidak hanya menyebabkan sakit jantung saja, melainkan dapat menyebabkan penyakit seperti, gagal ginjal, hipertensi serta stroke. Menurut *World Health Organization* (WHO), penyakit kardiovaskular telah menyebabkan 17,6 juta kematian setiap tahunnya [2]. Pendapat lain dari *Institute for Health Metrics and Evaluation* (IHME) yang mengatakan bahwa dengan adanya 14,4% kematian di Indonesia yang disebabkan oleh penyakit jantung koroner [3]. Penyakit kardiovaskular telah menyebabkan kematian sekitar 80% pada negara yang memiliki penghasilan rendah sampai menengah seperti negara Indonesia sendiri [4]. Penyakit kardiovaskular dapat disebabkan oleh beberapa faktor secara umum yakni, faktor yang dapat diubah seperti kebiasaan merokok, dan minuman beralkohol yang berlebihan serta faktor yang tidak dapat diubah seperti riwayat penyakit keturunan. [5].

Dari beberapa data yang telah menunjukkan tingkat tingginya penyebab kematian pada penderita penyakit kardiovaskular, maka pemeriksaan penyakit kardiovaskular sangat penting dilakukan. Pemeriksaan penyakit kardiovaskular dapat memanfaatkan berbagai metode *artificial intelligence* (AI) dalam melakukan identifikasi penyakit kardiovaskular. Terdapat salah satu metode *artificial intelligence* (AI), yang memiliki kinerja yang optimal yakni metode *Fuzzy K-Nearest Neighbor* (FKNN). Metode *Fuzzy K-NN* merupakan sebuah metode pengklasifikasian dengan melakukan pencarian nilai keanggotaan pada setiap kelas. Metode FKNN telah banyak digunakan oleh peneliti lainnya dalam mengidentifikasi berbagai penelitian seperti, penelitian yang dilakukan pada tahun 2019 mengenai "Diagnosis Penyakit Cabai Menggunakan Metode *Fuzzy K-Nearest Neighbor* (FKNN)", dengan mendapatkan hasil pendiagnosaan penyakit yang menyerang pada tanaman cabai, dengan menggunakan pengujian terhadap nilai K sebanyak 4 kali yakni nilai K=5, K=10, K=15 dan K=20, dengan hasil pengujian yang didapatkan yakni akurasi yang optimal

sebesar 92% pada nilai $K=5$ [6]. Penelitian lain dalam penggunaan metode FKNN dilakukan pada tahun 2019 mengenai “Implementasi Metode FKNN (*Fuzzy K-Nearest Neighbor*) Untuk Diagnosa Penyakit Tanaman Kentang”, dengan mendapatkan hasil berupa pendeteksi penyakit tanaman kentang, dengan dilakukan pengujian sebanyak 4 kali pada nilai $K=5$, $K=10$, $K=15$ dan $K=20$, dengan hasil akurasi yang optimal yakni sebesar 91,4% pada nilai $K=15$ [7]. Penelitian lain juga dilakukan pada tahun 2017 mengenai “Klasifikasi kualitas Tanaman Cabai Menggunakan Metode *Fuzzy K-Nearest Neighbor* (FKNN)”, dengan menghasilkan hasil akurasi yang optimal sebesar 96,67% dari data sampel sebanyak 100 data dengan pembagian data *training* dan data *testing* sebesar 70 untuk data *training* dan 30 untuk data *testing* [8].

Dari beberapa penelitian yang menggunakan metode FKNN mendapatkan hasil akurasi yang baik dalam melakukan klasifikasi, sehingga dalam mengidentifikasi penyakit kardiovaskular ini menggunakan metode *Fuzzy K-Nearest Neighbor* (FKNN), untuk mendapatkan hasil klasifikasi yang baik. Dengan dilakukan identifikasi penyakit kardiovaskular ini, untuk membantu para tenaga medis dalam mengidentifikasi penyakit kardiovaskular dengan efisien, sehingga dapat meminimalisir kematian yang meningkat yang disebabkan oleh penyakit kardiovaskular. Penggunaan metode *Fuzzy K-Nearest Neighbor* (FKNN) pada identifikasi penyakit kardiovaskular diharapkan mendapatkan hasil klasifikasi yang lebih baik.

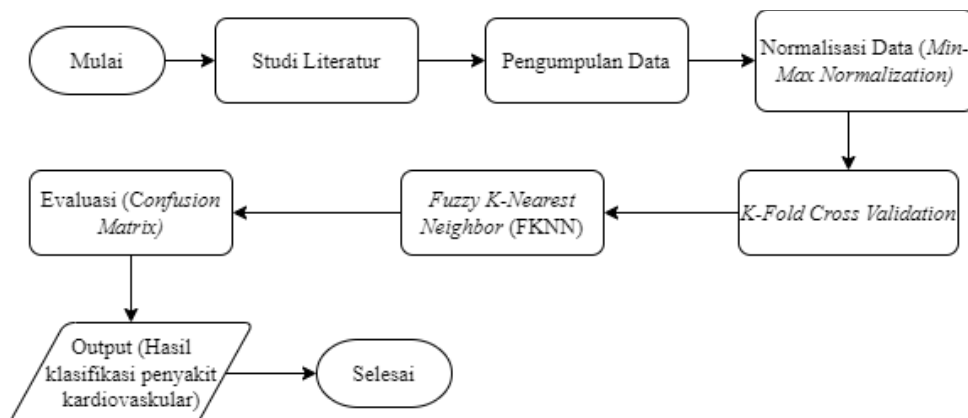
2. Metode Penelitian

Penelitian ini dibagi menjadi beberapa langkah proses yakni, studi literatur, pengumpulan data, klasifikasi menggunakan metode *Fuzzy K-Nearest Neighbor* (FKNN), dan pengujian. Studi literatur digunakan untuk memperkuat permasalahan yang diangkat pada penelitian dengan menelusuri beberapa sumber yang bersangkutan dengan permasalahan. Selanjutnya pengumpulan data, dalam penelitian ini menggunakan data sekunder yang bersumber dari *Website Kaggle.com* dengan jumlah data sebanyak 303 data. Untuk melakukan standarisasi data agar pada saat pengujian, agar data tidak ada yang mendominasi maka data akan melewati proses normalisasi. Normalisasi akan menggunakan metode *Min-Max Normalization* agar skala nilai hanya memiliki rentang 0 sampai 1 pada semua data yang digunakan pada penelitian [9]. Pada persamaan yang dapat dilihat pada persamaan 1 merupakan rumus perhitungan *Min-Max Normalization*.

$$v' = \frac{v - \min_A}{\max_A - \min_A} \quad (1)$$

Dalam normalisasi data, terdapat simbol v' yang mendefinisikan hasil dari data yang telah dinormalisasi, simbol v yang artinya bahwa data akan dilakukan proses normalisasi, simbol \max_A yakni nilai maksimum, serta simbol \min_A yang artinya nilai minimum.

Setelah data melewati proses normalisasi, dataset akan diproses untuk pembagian data dengan menggunakan *k-fold cross validation*. Penggunaan *k-fold cross validation* dalam pembagian data dikarenakan hasil pengujian akan lebih optimal, sebab semua dataset yang digunakan pada penelitian akan digunakan sebagai data *training* dan data *testing* pada saat pengujian dilakukan [11]. Setelah dilakukan pembagian data, selanjutnya dilakukan klasifikasi dengan menggunakan metode *Fuzzy K-Nearest Neighbor* (FKNN). Setelah klasifikasi menggunakan metode *Fuzzy K-Nearest Neighbor* (FKNN), dilakukan pengujian menggunakan *confusion matrix* dalam melakukan perhitungan tingkat akurasi klasifikasi penyakit kardiovaskular berdasarkan data yang telah digunakan. Pada gambar yang dapat dilihat pada Gambar 1 merupakan diagram alur dari desain penelitian.



Gambar 1. Diagram alur Penelitian

2.1 Pengumpulan Data

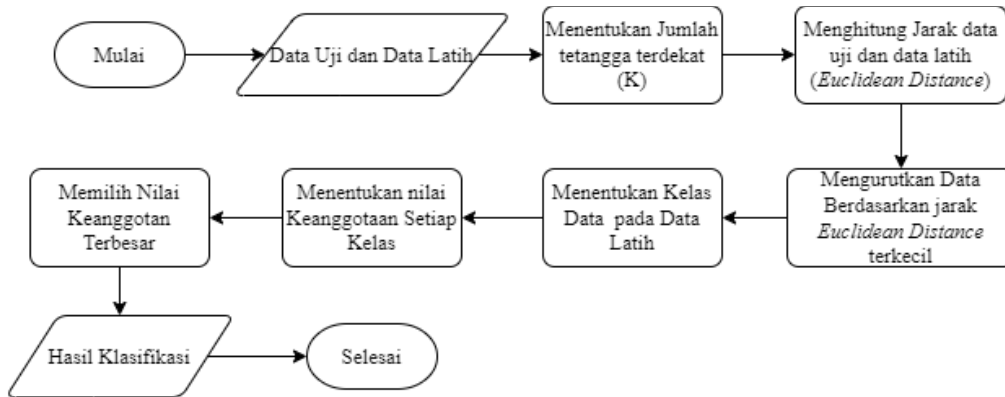
Penelitian ini menggunakan data sekunder yang didapatkan dari *Website Kaggle.com* yang diunduh pada tanggal 23 September 2022 Pukul 01.02 Am yang diterbitkan oleh *Volodymyr Gavrysh*. Data yang didapatkan berjumlah 303 data dengan 14 fitur didalamnya yang terdiri dari *age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal*, serta *target* [10]. Pada tabel yang dapat dilihat pada Tabel 1 merupakan atribut dataset yang digunakan pada penelitian.

Tabel 1. Atribut Dataset Penelitian

No	Nama Atribut	Tipe Data	Keterangan
1	<i>age</i>	<i>Integer</i>	Umur pasien
2	<i>sex</i>	<i>Integer</i>	Jenis kelamin pasien (1 = laki-laki; 2 = perempuan).
3	<i>cp (chestpain)</i>	<i>Integer</i>	Nyeri dada (3 = tipikal; 2 = asimtotik; 1 = nonanginal; 0 = nontipikal)
4	<i>trestbps</i>	<i>Integer</i>	Tekanan darah istirahat
5	<i>chol</i>	<i>Integer</i>	Kolesterol serum dalam mg/dl
6	<i>fbs</i>	<i>Integer</i>	Gula darah puasa > 120 mg/dl (1 = benar; 0 = salah)
7	<i>restecg</i>	<i>Integer</i>	Hasil elektrokardiografi saat istirahat.
8	<i>thalach</i>	<i>Integer</i>	Detak jantung maksimum tercapai
9	<i>exang</i>	<i>Integer</i>	Angina yang diinduksi oleh olahraga (1 = ya; 0 = tidak)
10	<i>oldpeak</i>	<i>Integer</i>	Depresi ST yang diinduksi oleh olahraga relatif terhadap istirahat
11	<i>slope</i>	<i>Integer</i>	Kemiringan segmen ST latihan puncak
12	<i>ca</i>	<i>Integer</i>	Jumlah pembuluh utama yang diwarnai oleh flourosopy (0 – 3)
13	<i>thal</i>	<i>Integer</i>	Cacat yang di alami pasien (3 = normal; 6 = cacat tetap; 7 cacat reversible)
14	<i>target</i>	<i>Integer</i>	Diagnosis penyakit kardiovaskular (1 = ya; 0 = tidak)

2.2 Fuzzy K-Nearest Neighbor (FKNN)

Fuzzy K-Nearest Neighbor (FKNN) merupakan sebuah metode yang digunakan dalam pengklasifikasian yang menggabungkan teknik *fuzzy* dengan *K-Nearest Neighbor classifier*. Teori himpunan *fuzzy* dalam menggeneralisasikan teori K-NN klasik yakni mendefinisikan nilai keanggotaan pada sebuah data dalam masing-masing kelas. FKNN yakni sebuah metode yang akan mencari nilai derajat keanggotaan data uji pada setiap kelasnya, yang kemudian akan mengambil nilai derajat keanggotaan terbesar yang dihasilkan. Nilai derajat keanggotaan terbesar akan digunakan sebagai kelas hasil dari klasifikasi. Pada gambar yang dapat dilihat pada Gambar 2 merupakan diagram alur perhitungan dari metode *Fuzzy K-Nearest Neighbor (FK-NN)*.



Gambar 2. Diagram alur perhitungan metode *Fuzzy K-Nearest Neighbor* (FKNN)

Proses perhitungan dari metode FKNN akan dimulai dengan menginputkan data latih serta data uji yang akan digunakan dalam proses klasifikasi FKNN, kemudian dilakukan penentuan terhadap jumlah nilai K tetangga terdekat yang akan dilakukan dalam perhitungan. Kemudian dilakukan perhitungan jarak dengan menggunakan fungsi *euclidean distance* pada data uji dan data latih. Pada persamaan yang dapat dilihat pada (2) merupakan rumus perhitungan jarak *euclidean distance* [12].

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \quad (2)$$

Dalam perhitungan jarak terdekat ada beberapa simbol yakni d_i yang memiliki arti jarak kedekatan (*euclidean distance*), kemudian adanya simbol p yang memiliki arti bahwa jumlah atribut, simbol x_1 yang artinya data latih serta x_2 yang artinya data uji.

Dalam perhitungan FKNN adanya perhitungan pada data latih dalam penentuan kelas data. Pada persamaan yang dapat dilihat pada persamaan (3) merupakan rumus perhitungan penentuan kelas pada data latih [13].

$$u_{ij} = \begin{cases} 0,51 + \left(\frac{n_j}{n}\right) * 0,49, & \text{jika } j = i \\ \frac{n_j}{n} * 0,49, & \text{jika } j \neq i \end{cases} \quad (3)$$

Pada penentuan suatu data latih adanya simbol n_j yang mengartikan bahwa jumlah anggota pada kelas j pada data latih n , simbol N mendefinisikan jumlah dari data latih yang digunakan pada penelitian, serta simbol J artinya kelas pada data.

Proses perhitungan terakhir pada metode FKNN yakni menentukan keanggotaan setiap kelas. Pada persamaan yang dapat dilihat pada persamaan (4) merupakan perhitungan keanggotaan setiap kelas [14].

$$u_i(x) = \frac{\sum_{j=1}^k u_{ij} \left(\|x - x_j\|^{\frac{-2}{(m-1)}} \right)}{\sum_{j=1}^k \left(\|x - x_j\|^{\frac{-2}{(m-1)}} \right)} \quad (4)$$

Pada rumus perhitungan penentuan keanggotaan setiap kelas terdapat simbol $u_i(x)$ yang mendefinisikan nilai keanggotaan data x dalam kelas ke i , kemudian adanya simbol k yang mendefinisikan jumlah tetangga terdekat yang dimiliki, simbol u_{ij} yang artinya nilai keanggotaan pada kelas i vektor j , simbol $x - x_j$ yang artinya selisih jarak data uji ke data latih, serta simbol m yang artinya bobot pangkat yang dimana bobot pangkat memiliki nilai lebih besar dari pada 1.

Setelah didapatkan hasil perhitungan keanggotaan setiap kelas, maka nilai keanggotaan yang diambil yakni nilai keanggotaan yang terbesar. Hasil dari nilai keanggotaan terbesar merupakan hasil dari klasifikasi penyakit kardiovaskular dengan menggunakan metode FKNN.

2.3 Pengujian

Pengujian pada penelitian ini menggunakan metode *confusion matrix*. Keluaran dari perhitungan pengujian ini yakni nilai akurasi, dimana jika nilai akurasi semakin besar maka semakin dekat hasil pengklasifikasian terhadap data *testing* yang digunakan. Pada persamaan yang dapat dilihat pada persamaan (5) merupakan rumus perhitungan akurasi pada metode *confusion matrix* [15].

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \times 100\% \quad (5)$$

Dari persamaan diatas adanya *True Positive* (TP) yang merupakan hasil data positif yang diprediksi sistem dengan benar, *True Negative* (TN) yakni data yang negatif yang diprediksi sistem dengan benar, *False Positive* (FP) yakni data negatif yang diprediksi sistem sebagai data yang positif, serta *False Positive* (FN) yakni data positif yang diprediksi sistem sebagai data yang negatif.

3. Hasil dan Pembahasan

Penelitian ini menggunakan bahasa pemrograman Python dalam melakukan klasifikasi penyakit kardiovaskular dengan metode *Fuzzy K-Nearest Neighbor* (FKNN). Dataset yang digunakan yakni dataset dari *Website Kaggle.com* yang dengan 14 fitur didalamnya yang terdiri dari *age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal*, serta *target*. Dataset berjumlah 303 data yang terdiri dari data pasien yang menderita penyakit kardiovaskular dan data pasien yang tidak menderita penyakit kardiovaskular. Pembagian data dalam penelitian klasifikasi penyakit kardiovaskular dengan metode *Fuzzy K-Nearest Neighbor* (FKNN) menggunakan *k-fold cross validation*, dimana dataset dibagi menjadi data *training* dan data *testing*. Pembagian dataset dibagi dengan 70 data *training* serta 30 data *testing*, dimana akan terdapat 212 data *training* dan 91 data *testing* dari 303 jumlah dataset. Dalam penelitian ini dilakukan pengujian sebanyak tujuh kali pengujian terhadap nilai K. Pengujian terhadap nilai K yang digunakan pada penelitian ini yakni pengujian pada K=3, K=5, K=7, K=9, K=11, K=13 dan K=15. Pada tabel yang dapat dilihat pada Tabel 2 merupakan hasil pengujian yang dilakukan terhadap nilai K yang digunakan.

Tabel 2. Hasil Pengujian Nilai K

No	Nilai K	Hasil Akurasi
1	K=3	75,82%
2	K=5	75,82%
3	K=7	76,92%
4	K=9	79,12%
5	K=11	78,02%
6	K=13	80,21%
7	K=15	80,21%

Perhitungan hasil akurasi yang didapatkan pada penelitian menggunakan perhitungan metode *confusion matrix*. Dari beberapa hasil akurasi yang didapatkan penggunaan metode *Fuzzy K-Nearest Neighbor* (FKNN) mendapatkan hasil yang baik dalam klasifikasi penyakit kardiovaskular.

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa penggunaan metode *Fuzzy K-Nearest Neighbor* (FKNN) pada klasifikasi penyakit kardiovaskular, dengan menggunakan pembagian dataset 70 data *training* serta 30 data *testing* pada 303 jumlah dataset mendapatkan hasil akurasi yang optimal pada K=13 dan K=15 dengan hasil akurasi sebesar 80,21% pada pengujian. Hasil yang didapatkan telah melebihi hasil dari penggunaan metode KNN klasik yang dimana hasil akurasi yang didapatkan pada penggunaan metode KNN klasik

sebesar 75.75% [16]. Dari hasil yang telah didapatkan maka penambahan teknik *fuzzy* ke dalam KNN menghasilkan nilai akurasi yang lebih baik daripada penggunaan KNN klasik.

Referensi

- [1] K. D. I. Yogyakarta, J. Farmako, S. Utara, J. Farmako, and S. Utara, "11096-31793-1-Sm," vol. 13, no. 1, pp. 1–12, 2020.
- [2] "Kenali Penyakit Kardiovaskuler yang Paling Umum Terjadi - Alodokter." <https://www.alodokter.com/kenali-penyakit-kardiovaskuler-yang-paling-umum-terjadi> (accessed Sep. 25, 2022).
- [3] "Penyakit Jantung Koroner Didominasi Masyarakat Kota – Sehat Negeriku." <https://sehatnegeriku.kemkes.go.id/baca/umum/20210927/5638626/penyakit-jantung-koroner-didominasi-masyarakat-kota/> (accessed Sep. 25, 2022).
- [4] "Kementerian Kesehatan Republik Indonesia." <https://www.kemkes.go.id/article/view/17073100005/penyakit-jantung-penyebab-kematian-tertinggi-kemenkes-ingatkan-cerdik-.html> (accessed Oct. 15, 2022).
- [5] R. R. Juslim and F. Herawati, "Penyakit kardiovaskular," 2007.
- [6] W. G. Akbari, N. Hidayat, and N. Santoso, "Diagnosis Penyakit Cabai Menggunakan Metode Fuzzy K-Nearest Neighbor (FKNN)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 3, no. 1, pp. 1070–1074, 2019.
- [7] D. M. F. Shanti, N. Hidayat, and R. C. Wihandika, "Implementasi Metode F-KNN (Fuzzy K-Nearest Neighbor) Untuk Diagnosis Penyakit Anjing," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 12, pp. 7401–7407, 2018.
- [8] A. A. Indra Wiratmaka, I. F. Rozi, and R. A. Asmara, "Klasifikasi Kualitas Tanaman Cabai Menggunakan Metode Fuzzy K-Nearest Neighbor (Fknn)," *J. Inform. Polinema*, vol. 3, no. 3, p. 1, 2017, doi: 10.33795/jip.v3i3.25.
- [9] A. Reza and O. S. Simanjuntak, "Classification of prospective borrowing customers to reduce the risk of bad deposits in sharia cooperatives using the FK-NNC method," *Comput. Inf. Process. Lett.*, vol. 1, no. 1, pp. 8–16, 2021, [Online]. Available: <http://103.23.20.161/index.php/cip/article/view/6125>.
- [10] "Heart_Disease | Kaggle." <https://www.kaggle.com/datasets/volodymyrgavrysh/heart-disease?resource=download> (accessed Sep. 30, 2022).
- [11] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, "Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 81–86, 2020, doi: 10.33096/ilkom.v12i2.507.81-86.
- [12] Dylan Trotsek, "APLIKASI KLASIFIKASI JENIS BAJA BERDASARKAN KOMPOSISI KIMIA DENGAN MENGGUNAKAN METODE FK-NNC," *J. Chem. Inf. Model.*, 2017.
- [13] C. Zheng, M. Shareduwan, M. Kasihmuddin, M. A. Mansor, and J. Chen, "Intelligent Multi-Strategy Hybrid Fuzzy K-Nearest Neighbor Using Improved Hybrid Sine Cosine Algorithm," pp. 1–23, 2022.
- [14] D. W. I. Ispriyanti, A. Prahutama, R. Dian, and I. K. A. Wati, "Analysis classification of households who received 'raskin' in Semarang City using Fuzzy K-Nearest Neighbor (FKNN) and Support Vector Machine (SVM)," *J. Math. Comput. Sci.*, pp. 1–13, 2022, doi: 10.28919/jmcs/7478.
- [15] H. Hozairi, A. Anwari, and S. Alim, "Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes," *Netw. Eng. Res. Oper.*, vol. 6, no. 2, p. 133, 2021, doi: 10.21107/nero.v6i2.237.
- [16] S. Sumarlinda and W. Lestari, "Aplikasi K-Nearest Neighbor (KNN) untuk Klasifikasi Penyakit Kardiovaskuler," no. 55, pp. 259–261.