

Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma K-Nearest Neighbor

Yovi Pratama, Anton Prayitno, Defri Nazrian, Nur Aini, Yoga Rizki R, Errissya Rasywir*

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dinamika Bangsa, Kota Jambi, Indonesia
Email: ¹yovi.pratama@gmail.com, ²antonprayitno018@gmail.com, ³defrinazrian069@gmail.com, ⁴nuraini077@gmail.com,
⁵yogarizkiramadan099@gmail.com, ⁶errissya.rasywir@gmail.com.
Email Penulis Korespondensi: errissya.rasywir@gmail.com

Abstrak—Penyakit kardiovaskular (CVDs) adalah penyebab kematian nomor 1 secara global, merenggut sekitar 17,9 juta nyawa setiap tahun, yang merupakan 31% dari semua kematian di seluruh dunia. Empat dari kematian 5CVD disebabkan oleh serangan jantung dan stroke, dan sepertiga dari kematian ini terjadi sebelum waktunya pada orang di bawah usia 70 tahun. Gagal jantung adalah kejadian umum yang disebabkan oleh CVD dan kumpulan data ini berisi 11 fitur yang dapat digunakan untuk memprediksi kemungkinan penyakit jantung. Orang dengan penyakit kardiovaskular atau yang memiliki risiko kardiovaskular tinggi (karena adanya satu atau lebih faktor risiko seperti hipertensi, diabetes, hiperlipidemia, atau penyakit yang sudah ada) memerlukan deteksi dan penanganan dini di mana model pembelajaran mesin dapat sangat membantu.

Kata Kunci: Gagal Jantung; Algoritma K-Nearest Neighbor; Prediksi; Data Mining.

Abstract—Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide. Four out of 5CVD deaths are due to heart attacks and strokes, and one-third of these deaths occur prematurely in people under 70 years of age. Heart failure is a common event caused by CVDs and this dataset contains 11 features that can be used to predict a possible heart disease. People with cardiovascular disease or who are at high cardiovascular risk (due to the presence of one or more risk factors such as hypertension, diabetes, hyperlipidaemia or already established disease) need early detection and management wherein a machine learning model can be of great help.

Keywords: Heart Failure; Algorithm K-Nearest Neighbor; Prediction; Data Mining.

1. PENDAHULUAN

Penyakit kardiovaskular (PKV) menurut definisi dari World Health Organization (WHO) dalam pusdatin Kemenkes RI 2014 yaitu penyakit yang disebabkan oleh gangguan fungsi jantung dan pembuluh darah [1]–[4], seperti penyakit jantung koroner, hipertensi, stroke dan penyakit gagal jantung [5]–[10]. Setiap tahunnya lebih dari 36 juta orang meninggal karena Penyakit Tidak Menular (PTM) 63% dari semua kematian. Secara global PTM menjadi pemicu kematian nomor satu setiap tahunnya. Kematian dini yang disebabkan oleh penyakit jantung terjadi sebesar 4% di negara berpendapatan tinggi dan 42% terjadi di negara berpendapatan rendah [11]–[16]. Data WHO mengatakan, lebih dari 17 juta manusia di dunia meninggal disebabkan penyakit jantung dan pembuluh darah.

Penyakit gagal jantung masuk ke kategori penyakit kardiovaskular [5], [6]. Gagal jantung adalah suatu kondisi ketika jantung tidak bisa mendapatkan cukup darah untuk memenuhi kebutuhan tubuh. Penderita gagal jantung merupakan pasien paling banyak di rumah sakit. Prevalensi gagal jantung di Amerika dan Eropa sekitar 1-2%. Di Indonesia belum ada data epidemiologi untuk gagal jantung, tetapi pada survei kesehatan nasional 2003 ditemukan bahwa penyakit sistem sirkulasi merupakan penyebab kematian utama di Indonesia (26,4%) dan pada profil Kesehatan Indonesia tahun 2003 disebutkan bahwa penyakit jantung berada pada urutan kedelapan. Mengingat berharganya organ vital seperti jantung, memprediksi gagal jantung telah menjadi prioritas bagi dokter dan tenaga medis, namun sampai saat ini prediksi perihail gagal jantung dalam praktik klinis biasanya gagal mendapatkan akurasi yang tinggi.

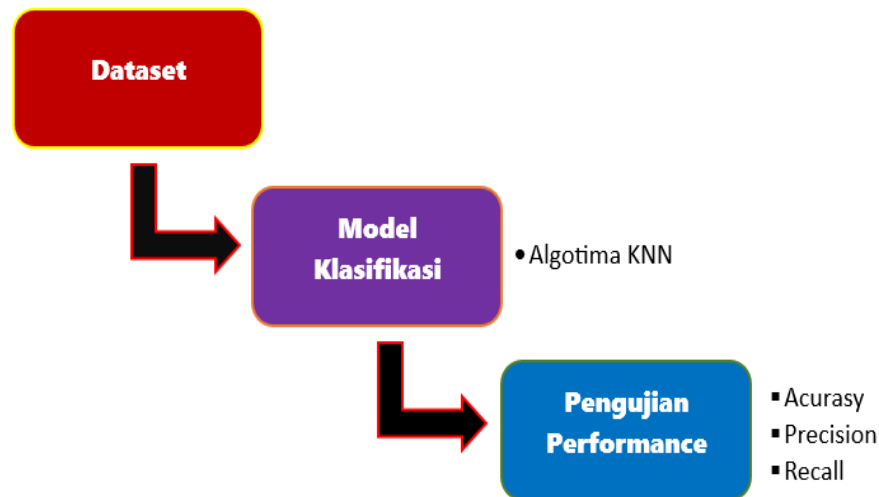
Algoritma K-Nearest Neighbor (KNN) yakni algoritma klasifikasi berlandaskan kedekatan jarak suatu data dengan data yang lain [11], [17]–[20]. Pada algoritma K-Nearest Neighbor (KNN), data berdimensi q , jarak dari data tersebut ke data yang lain dapat dihitung. Nilai jarak inilah yang ditetapkan sebagai nilai kedekatan atau kemiripan antara data uji dengan data latih. Nilai K pada K-Nearest Neighbor (KNN) berarti K -data terdekat dari data uji. Telah banyak penelitian [21]–[24] tentang prediksi dengan memakai algoritma K-Nearest Neighbor (KNN) salah satunya penelitian yang dilakukan oleh Gultom dan Simanjuntak (2020), yaitu prediksi tingkat kelancaran pembayaran kredit bank. Algoritma yang digunakan yaitu naïve bayes dan K-Nearest Neighbor (KNN), dataset yang digunakan mengandung 363 data dan 14 atribut. Hasil yang diperoleh menyatakan bahwa pengujian yang dilakukan dengan menggabungkan algoritma Naïve Bayes dan K-Nearest Neighbor (KNN) mampu melakukan prediksi dengan baik dengan akurasi 80%. Sehingga gabungan kedua algoritma tersebut cocok digunakan untuk prediksi kelancaran pembayaran kredit pada bank, dan dapat membantu petugas bank dalam menentukan nasabahnya.

2. METODOLOGI PENELITIAN

2.1. Prosedur Penelitian

Penelitian ini dilakukan dengan menggunakan implementasi algoritma K-Nearest Neighbor dalam prediksi dan identifikasi kematian akibat gagal jantung. Algoritma KNN digunakan dengan melakukan beberapa ujicoba terhadap

nilai k dengan ukuran jarak Euclidian. Pengujian kinerja KNN menggunakan akurasi, presisi, dan recall. Tahapan penelitian seperti yang ditunjukkan pada Gambar 1 di bawah ini :



Gambar 1. Tahapan Penelitian Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma K-Nearest Neighbor

2.2. Dataset Penyakit gagal jantung

Kumpulan data penyakit gagal jantung yang di dapat dari website kaggle.com, <https://www.kaggle.com/dataset/fedoriano/heart-failure-prediction>. Dataset ini terdiri dari 918 data dan 12 atribut.

#	A	S	C	R	C	F	R	E	A	O	S	H
1	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
2	40	M	ATA	140	289	0	Normal	172	N	0	Up	Normal
3	49	F	NAP	160	180	0	Normal	156	N	1	Flat	Tervonis Jantung
4	37	M	ATA	130	283	0	ST	98	N	0	Up	Normal
5	48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	Tervonis Jantung
6	54	M	NAP	150	195	0	Normal	122	N	0	Up	Normal
7	39	M	NAP	120	339	0	Normal	170	N	0	Up	Normal
8	45	F	ATA	130	237	0	Normal	170	N	0	Up	Normal
9	54	M	ATA	110	208	0	Normal	142	N	0	Up	Normal
10	37	M	ASY	140	207	0	Normal	130	Y	1.5	Flat	Tervonis Jantung
11	48	F	ATA	120	284	0	Normal	120	N	0	Up	Normal
12	37	F	NAP	130	211	0	Normal	142	N	0	Up	Normal
13	58	M	ATA	136	164	0	ST	99	Y	2	Flat	Tervonis Jantung
14	39	M	ATA	120	204	0	Normal	145	N	0	Up	Normal
15	49	M	ASY	140	234	0	Normal	140	Y	1	Flat	Tervonis Jantung
16	42	F	NAP	115	211	0	ST	137	N	0	Up	Normal
17	54	F	ATA	120	273	0	Normal	150	N	1.5	Flat	Normal
18	38	M	ASY	110	196	0	Normal	166	N	0	Flat	Tervonis Jantung
19	43	F	ATA	120	201	0	Normal	165	N	0	Up	Normal
20	60	M	ASY	100	248	0	Normal	125	N	1	Flat	Tervonis Jantung
21	36	M	ATA	120	267	0	Normal	160	N	3	Flat	Tervonis Jantung
22	43	F	TA	100	223	0	Normal	142	N	0	Up	Normal
23	44	M	ATA	120	184	0	Normal	142	N	1	Flat	Normal
24	49	F	ATA	124	201	0	Normal	164	N	0	Up	Normal
25	44	M	ATA	150	288	0	Normal	150	Y	3	Flat	Tervonis Jantung
26	40	M	NAP	130	215	0	Normal	138	N	0	Up	Normal
27	36	M	NAP	130	209	0	Normal	178	N	0	Up	Normal
28	53	M	ASY	124	260	0	ST	112	Y	3	Flat	Normal
29	52	M	ATA	120	284	0	Normal	118	N	0	Up	Normal
30	53	F	ATA	113	468	0	Normal	127	N	0	Up	Normal
31	51	M	ATA	125	188	0	Normal	145	N	0	Up	Normal
32	53	M	NAP	145	518	0	Normal	130	N	0	Flat	Tervonis Jantung
917	57	M	ASY	130	131	0	Normal	115	Y	1.2	Flat	Tervonis Jantung
918	57	F	ATA	130	236	0	LVH	174	N	0	Flat	Tervonis Jantung
919	38	M	NAP	138	175	0	Normal	173	N	0	Up	Normal

Gambar 2. Dataset Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma K-Nearest Neighbor

Berikut dataset dan penjelasan pada atributnya :

1. Age : umur
2. Sex : jenis kelamin (M:pria, F:wanita)
3. ChestPainType: tipe nyeri dada [TA: angina khas, ATA: angina atipikal, NAP: nyeri non-anginal, ASY: asimtomatik]
4. RestingBP : tekanan darah [mm Hg]
5. Colestrol: kolestrol serum [mm/dl]
6. FastingBS : gula darah puasa [1: jika fastigBS>120 mg/dl, 0: jika tidak]
7. RestingECG : hasil elektrokardiogram istirahat [Normal: normal, ST: memiliki kelainan gelombang ST-T, LVH: menunjukkan hipertrofi ventrikel]
8. MaxHR : detak jantung maksimum tercapai
9. ExerciseAngina : angin akibar olahraga [Y: ya, No: tidak]
10. Oldpeak : oldpeak [ST: nilai numerik diukur dari depresi]
11. ST_Slope: kemiringan segmen ST latihan puncak [Up: upsloping, Flat: flat, Down:downsloping]
12. HeartDiasease: kelas keluaran

2.3. Algoritma K-Nearest Neighbor (KNN)

Algoritma/metode KNN merupakan algoritma dengan pembelajaran terawasi dan banyak digunakan untuk prediksi dan klasifikasi. Kelebihan dari metode KNN adalah akurasi yang tinggi, intensif pada outlier dan tidak ada asumsi tentang data. Menentukan nilai K menjadi penting. Kesamaan data dengan label digunakan jarak Euclidian dengan rumus:

$$d(x, y) = \sqrt{\sum_i^n (x_i - y_i)^2}$$

Penjelasan :

d(x,y) : jarak data x dan y

xi : sampel data

yi : data uji atau data testing

i : variabel data

n : dimensi data

Tahapan dari algoritma K-Nearest Neighbour (KNN) adalah sebagai berikut :

- Tentukan nilai k, yaitu jumlah banyaknya tetangga terdekat
- Hitung jarak antara data baru dan semua data training
- Mengurutkan jarak tersebut dan menentukan tetangga mana yang terdekat berdasarkan jarak minimum ke-k
- Menentukan kategori dari tetangga terdekat
- Menggunakan kategori mayoritas sederhana dari tetangga yang terdekat tersebut sehingga nilai prediksi dari data yang baru

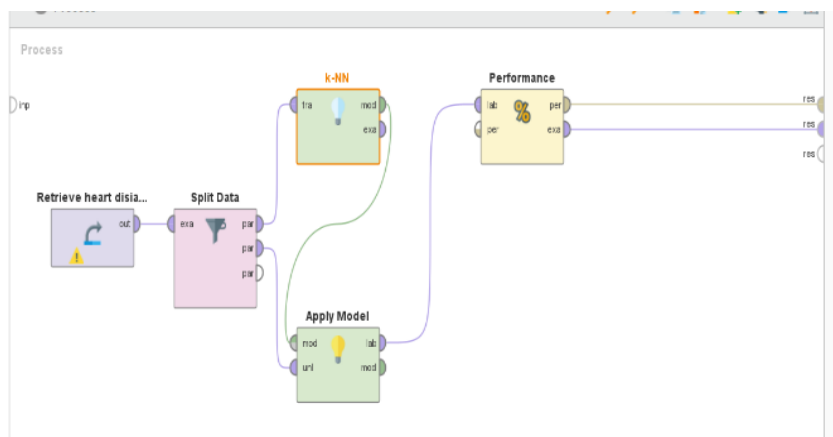
3. HASIL DAN PEMBAHASAN

Pengembangan model klasifikasi dengan machine learning dengan menggunakan metode K-Nearest Neighbor (KNN) menggunakan software Rapidminer. Dataset yang digunakan dataset gagal jantung yang di dapat dari kaggle.com dengan 12 atribut yang digunakan sebagai parameter dengan Heart Disease sebagai label cardio. Dataset berisi 918 data pasien yang terdiri dari penyakit gagal jantung dan data pasien tidak sehat. Implementasi algoritma K-Nearest Neighbor (KNN) untuk model prediktif menggunakan Rapidminer seperti gambar di bawah ini :

Name	Type	Status	Location	Attributes
HeartDisease	General	C	Normal	Tancong Jombang (Amonek, Jember)
Age	Integer	C	30	77
Sex	Nominal	C	M	F
ChestPainType	Nominal	C	ATA (48)	ASY (495) ASY (596), SA
RestingBP	Integer	C	0	200
Cholesterol	Integer	C	0	600
FastingBS	Nominal	C	Normal	1
RestingCG	Nominal	C	SI (176)	Normal (207)
MaxHR	Integer	C	0	180

Gambar 3. Statistic Data Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma K-Nearest Neighbor

Pada gambar 3 di perhatikan statistic pada dataset yang sudah kita olah di Ripedminer tidak ada yang missing value.



Gambar 4. Desain KNN pada Ripedminer Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma K-Nearest Neighbor

Pada gambar 3 merupakan desain untuk menghitung accuracy pada algoritma KNN di Ripedminer pada Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma K-Nearest Neighbor. Pada menu split data kami menggunakan ratio perbandingan 0.8 training dan 0.2 testing. Kemudian kami mencoba menggunakan beberapa ujicoba terhadap nilai k=5, k=7 dan k=9 sebagai perbandingan seperti yang di perlihatkan pada gambar 5,6 dan 7 :

accuracy: 65.76%

	true Normal	true Tervonis Jantung	class precision
pred. Normal	53	34	60.92%
pred. Tervonis Jantung	29	68	70.10%
class recall	64.63%	66.67%	

Gambar 5. Tabel Performance K=5 Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma K-Nearest Neighbor

Pada gambar 5 itu merupakan sampel data KNN dengan k=5 pada Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma K-Nearest Neighbor mendapatkan accuracy = 65,76%, precision = 70,10% dan recall = 66,67%.

accuracy: 68.48%

	true Normal	true Tervonis Jantung	class precision
pred. Normal	56	32	63.64%
pred. Tervonis Jantung	26	70	72.92%
class recall	68.29%	68.63%	

Gambar 6. Tabel Performance K=7 Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma K-Nearest Neighbor

Pada gambar 6 ini kami menggunakan k=7 dengan mendapatkan accuracy = 68,48%, precision = 72,92% dan recall = 68,63% pada Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma K-Nearest Neighbor.

accuracy: 70.65%

	true Normal	true Tervonis Jantung	class precision
pred. Normal	58	30	65.91%
pred. Tervonis Jantung	24	72	75.00%
class recall	70.73%	70.59%	

Gambar 7. Tabel Performance K=9 Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma K-Nearest Neighbor

Pada gambar 6 ini digunakan k=9 pada Klasifikasi Penyakit Gagal Jantung dengan Algoritma K-Nearest Neighbor Beberapa percobaan kepada semua nilai k namun hanya k=9 yang mempunyai accuracy tertinggi yaitu accuracy = 70,65%, precision = 75,00 % dan recall = 70,73%. Dari hasil pengujian dengan beberapa nilai k=5, k=7 dan k=9 hanya k=9 yang mendapat nilai tertinggi maka model klasifikasi dengan algoritma KNN pada dataset penyakit jantung terbilang cukup baik.

4. KESIMPULAN

Berdasarkan tahapan penelitian yang telah dilakukan pada penerapan algoritma K-Nearest Neighbor pada dataset penyakit gagal jantung. Kesimpulan yang diperoleh dari penelitian ini, yaitu penerapan algoritma KNN terhadap data penyakit gagal jantung dilakukan dengan pengujian akurasi yang dilakukan dengan menggunakan aplikasi rapidminer. Dataset yang digunakan sebanyak 918 data dan 12 atribut. Kemudian hasil implementasi algoritma KNN pada aplikasi ripedminer dilakukan dengan menggunakan pergantian nilai k, dan hasil akurasi tertinggi didapat pada nilai k=9 dengan akurasi sebesar 70,65%, nilai presesinya adalah 75 %, sedangkan recall menghasilkan 70,73%.

REFERENCES

- [1] F. Wulandari, P. A. Jusia, and J. Jasmir, "Klasifikasi Data Mining Untuk Mendiagnosa Penyakit ISPA Menggunakan Metode Naïve Bayes Pada Puskesmas Jambi Selatan," J. Manaj. Teknol. dan Sist. Inf., vol. 2, no. 3, pp. 214-227, 2020.
- [2] M. R. Kumaseh, L. Latumakulita, and N. Nainggolan, "Segmentasi citra digital ikan menggunakan metode Tresholding," J. Ilm. Sains, vol. 13 No., no. 1, p. 6, 2013.
- [3] A. Wanto and A. P. Windarto, "Analisis Prediksi Indeks Harga Konsumen Berdasarkan Kelompok Kesehatan Dengan

- Menggunakan Metode Backpropagation,” *J. Penelit. Tek. Inform. Sink.*, vol. 2, no. 2, pp. 37–43, 2017.
- [4] B. H. Hayadi and A. Setiawan, “Sistem Berbasis Pengetahuan Dengan Menggunakan Fuzzy Tsukamoto (Untuk Kesehatan Dan Perawatan Bayi),” in *SENTIKA*, 2016, vol. 2016, no. Sentika, pp. 18–19.
- [5] C. Science, “Prediksi Penyakit Jantung dengan Algoritma Klasifikasi,” vol. 5, no. 1, pp. 978–979, 2019.
- [6] S. Nurmaini, A. Darmawahyuni, A. N. S. Mukti, M. N. Rachmatullah, F. Firdaus, and B. Tutuko, “Deep learning-based stacked denoising and autoencoder for ECG heartbeat classification,” *Electron.*, vol. 9, no. 1, 2020, doi: 10.3390/electronics9010135.
- [7] E. Rasywir, R. Sinaga, and Y. Pratama, “Analisis dan Implementasi Diagnosis Penyakit Sawit dengan Metode Convolutional Neural Network (CNN),” *J. Paradig. UBSI*, vol. 22, no. 2, pp. 117–123, 2020.
- [8] Fachruddin, M. R. Pahlevi, M. Ismail, and E. Rasywir, “Pengujian Implementasi Sistem Pengelolaan Keuangan Masjid Berbasis Web Dan Android,” *J. Paradig. UBSI*, vol. 22, no. 2, pp. 124–131, 2020.
- [9] Saparudin and E. Rasywir, “Pengenalan potensi anak melalui sidik jari menggunakan algoritma voting feature intervals 5 (vfi5) 1,” vol. 1, no. 1, pp. 25–30, 2012.
- [10] M. R. Borroek, E. Rasywir, Y. Pratama, Fachruddin, and M. Istoningtyas, “Analysis on Knowledge Layer Application for Knowledge Based System,” in *Proceedings of 2018 International Conference on Electrical Engineering and Computer Science, ICECOS 2018, 2019*, pp. 177–182, doi: 10.1109/ICECOS.2018.8605262.
- [11] E. Rasywir, R. Sinaga, and Y. Pratama, “Evaluasi Pembangunan Sistem Pakar Penyakit Tanaman Sawit dengan Metode Deep Neural Network (DNN),” vol. 4, pp. 1206–1215, 2020, doi: 10.30865/mib.v4i4.2518.
- [12] H. Kaur and V. Kumari, “Predictive modelling and analytics for diabetes using a machine learning approach,” *Appl. Comput. Informatics*, 2019, doi: 10.1016/j.aci.2018.12.004.
- [13] R. I. Fajri, “Identifikasi Penyakit Daun Tanaman Kelapa Sawit Menggunakan Support Vector Machine,” *J. Teknol. Perkeb.*, 2014, [Online]. Available: <http://repository.usu.ac.id/handle/123456789/42256>.
- [14] Y. Pratama and E. Rasywir, “Eksperimen Penerapan Sistem Traffic Counting dengan Algoritma YOLO (You Only Look Once) V.4.,” *J. Media Inform. Budidarma*, vol. 5, no. 4, p. 1438, 2021, doi: 10.30865/mib.v5i4.3309.
- [15] Y. Hartiwi, E. Rasywir, Y. Pratama, and P. A. Jusia, “Sistem Manajemen Absensi dengan Fitur Pengenalan Wajah dan GPS Menggunakan YOLO pada Platform Android,” *J. Media Inform. Budidarma*, vol. 4, pp. 1235–1242, 2020, doi: 10.30865/mib.v4i4.2522.
- [16] Fachruddin, M. R. Pahlevi, M. Ismail, E. Rasywir, and Y. Pratama, “Analisis Usability Pada Implementasi Sistem Pengelolaan Keuangan Masjid Menggunakan USE Questionnaire,” *J. Media Inform. Budidarma*, vol. 4, pp. 1216–1224, 2020, doi: 10.30865/mib.v4i4.2518.
- [17] V. N. Sari, L. Y. Astri, and E. Rasywir, “Analisis Dan Penerapan Algoritma Naive Bayes Untuk Evaluasi,” *J. Ilm. Mhs. Tek. Inform.*, vol. 2, no. 1, pp. 53–68, 2020.
- [18] Y. Pratama and E. Rasywir, “Automatic Cost Estimation Analysis on Datawarehouse Project with Modified Analogy Based Method,” in *Proceedings of 2018 International Conference on Electrical Engineering and Computer Science, ICECOS 2018, 2019*, pp. 171–176, doi: 10.1109/ICECOS.2018.8605195.
- [19] E. Gho, D. Z. Abidin, and E. Rasywir, “Analisis Dan Penerapan Data Mining Pada Transaksi Penjualan Obat Menggunakan Algoritma Apriori Di Apotek Persijam,” *Tek. Inform. STIKOM Din. Bangsa*, pp. 56–64, 2013.
- [20] Hendrawan, A. Haris, E. Rasywir, and Y. Pratama, “Sistem Pakar Diagnosis Penyakit Tanaman Karet dengan Metode Fuzzy Mamdani Berbasis Web,” *J. Media Inform. Budidarma*, vol. 4, no. 4, pp. 1225–1234, 2020, doi: 10.30865/mib.v4i4.2521.
- [21] E. Rasywir, Y. Pratama, H. Hendrawan, and M. Istoningtyas, “Removal of Modulo as Hashing Modification Process in Essay Scoring System Using Rabin-Karp,” *2018 Int. Conf. Electr. Eng. Comput. Sci.*, pp. 159–164, 2018.
- [22] E. Rasywir and A. Purwarianti, “Eksperimen pada Sistem Klasifikasi Berita Hoax Berbahasa Indonesia Berbasis Pembelajaran Mesin,” *J. Cybermatika*, vol. 3, no. 2, pp. 1–8, 2015, [Online]. Available: <http://cybermatika.stei.itb.ac.id/ojs/index.php/cybermatika/article/view/133>.
- [23] F. Fachruddin, E. Rasywir, Hendrawan, Y. Pratama, D. Kisbianty, and M. R. Borroek, “Real Time Detection on Face Side Image with Ear Biometric Imaging Using Integral Image and Haar- Like Feature,” *2018 Int. Conf. Electr. Eng. Comput. Sci.*, pp. 165–170, 2018.
- [24] D. Zaenal Abidin and E. Rasywir, “Penerapan Data Mining Klasifikasi Untuk Memprediksi Potensi Mahasiswa Berprestasi Di Stikom Dinamika Bangsa Jambi Dengan Metode Naive Bayes,” *J. Ilm. Mhs. Tek. Inform.*, vol. 3, no. 2, 2021.