

# Perancangan Aplikasi Sistem Pakar Penyakit *Diabetes Mellitus* Menggunakan Algoritma *Random Forest*

Benedicto Reinaldy Pramananditya<sup>a1</sup>, Ryan Putranda Kristanto<sup>a2</sup>

<sup>a</sup> Ilmu Informatika, Universitas Katolik Darma Cendika

Jl. Dr. Ir. H. Soekarno No. 201, Klampis Ngasem, Kec. Sukolilo, Surabaya, Jawa Timur 60117

<sup>1</sup>benedicto.reinaldy@student.ukdc.ac.id

<sup>2</sup>ryan@ukdc.ac.id

## Abstract

*Diabetes Mellitus is a disease that infects many people both domestically and abroad. Based on data from the International Diabetes Federation (IDF), in 2021 Indonesia will be ranked number 5 in the number of people suffering from Diabetes Mellitus, namely 19.47 people. Diabetes Mellitus is a disorder of insulin function in the body which usually has general symptoms, namely an increase in blood sugar in humans. In general, Diabetes Mellitus is categorized into two types, namely Diabetes Mellitus type 1 and Diabetes Mellitus type 2. However, many people, especially in Indonesia, have a low level of awareness and awareness of this disease. This is caused by a lack of knowledge about this disease and its risks as well as limited time or costs in consulting a doctor. Therefore, it is necessary to implement artificial intelligence which is applied to the expert system application for Diabetes Mellitus. The design of this expert system application is intended to obtain results in the form of diagnosis, prediction and consultation. This research applies the Random Forest algorithm as a classification algorithm. In its application, this expert system application uses a combination of datasets from the Gotong Royong Surabaya Hospital and publik references with a total of 70 rows of data. This algorithm model training uses a ratio of training data to training data, namely 80:20 with an accuracy obtained of 100% and from the confusion matrix evaluation the results obtained were precision of 1.00, recall of 1.00, and f1 score of 1.00. From the results of the accuracy of model training and algorithm evaluation using a confusion matrix, it can be said that the implementation of the Diabetes Mellitus expert system using the Random Forest algorithm is suitable and accurate.*

**Keywords:** *Diabetes Mellitus, Diabetes, Expert System, Random Forest, Classification*

## Abstrak

Penyakit *Diabetes Mellitus* adalah penyakit yang banyak menjangkit masyarakat baik dalam negeri maupun luar negeri. Berdasarkan data dari International Diabetes Federation (IDF), pada tahun 2021 Negara Indonesia sebagai Negara dengan peringkat nomor 5 dalam jumlah pengidap penyakit *Diabetes Mellitus* yaitu sebanyak 19,47 jiwa. Penyakit *Diabetes Mellitus* adalah penyakit kelainan pada fungsi insulin pada tubuh yang biasanya memiliki gejala umum yaitu kenaikan gula darah pada manusia. Pada umumnya, penyakit *Diabetes Mellitus* dikategorikan mejadi dua macam yaitu *Diabetes Mellitus* tipe 1 dan *Diabetes Mellitus* tipe 2. Namun, banyak masyarakat, khususnya di Indonesia memiliki tingkat kewaspadaan dan tingkat kesadaran yang rendah terhadap penyakit ini. Hal ini disebabkan oleh minimnya pengetahuan terhadap penyakit ini dan resikonya serta keterbatasan waktu atau biaya dalam melakukan konsultasi kepada dokter. Oleh karena itu dibutuhkan implementasi dari ilmu kecerdasan buatan yang diterapkan pada aplikasi sistem pakar penyakit *Diabetes Mellitus*. Perancangan aplikasi sistem pakar ini ditujukan untuk mendapatkan hasil berupa diagnosis, prediksi, dan konsultasi. Penelitian ini menerapkan algoritma *Random Forest* sebagai algoritma klasifikasinya. Pada penerapannya, aplikasi sistem pakar ini menggunakan gabungan dataset dari Rumah Sakit Gotong Royong dan referensi publik dengan total baris data yaitu 70. Pelatihan model algoritma ini menggunakan perbandingan data training dan data training yaitu 80:20 dengan akurasi yang didapat sebesar 100% dan dari evaluasi confusion matrix didapatkan hasil precision sebesar 1.00, recall sebesar 1.00, dan f1 score sebesar 1.00. Dari hasil akurasi pelatihan model dan evaluasi algoritma menggunakan

confussion matrix dapat dikatakan bahwa implementasi sistem pakar penyakit *Diabetes Mellitus* menggunakan algoritma *Random Forest* adalah cocok dan akurat

**Kata Kunci:** *Diabetes Mellitus*, Diabetes, Sistem Pakar, *Random Forest*, Klasifikasi

## 1. Pendahuluan

Diabetes atau *Diabetes Mellitus* (DM) adalah sebuah penyakit global yang banyak diderita oleh masyarakat baik dalam negeri ataupun luar negeri. IDF (*International Diabetes Federation*) mencatat bahwa pada tahun 2021 lalu, sebanyak 537 juta jiwa orang dewasa terjangkit penyakit *Diabetes Mellitus* dengan angka kematian sebanyak 6,7 juta jiwa. Sedangkan Negara Indonesia mencapai posisi peringkat nomor 5 dalam pengidap penyakit *Diabetes Mellitus* dengan angka sebanyak 19,47 juta jiwa [1].

Penyakit *Diabetes Mellitus* adalah sebuah penyakit kelainan fungsi insulin pada tubuh manusia yang biasanya memiliki gejala umum yaitu naiknya kadar gula darah. Penyakit *Diabetes Mellitus* pada umumnya dikategorikan menjadi dua yaitu *Diabetes Mellitus* tipe 1 (DM 1) dan *Diabetes Mellitus* tipe 2 (DM 2). Gejala dari penyakit *Diabetes Mellitus* yang dialami oleh orang yang terjangkit dengan penyakit ini sangat bervariasi, sehingga gejala penyakit ini sulit untuk dikenali dan seringkali banyak orang tidak menyadari bahwa dirinya sedang dalam resiko terjangkit penyakit *Diabetes Mellitus* [2]. Adapun contoh-contoh dari gejala penyakit ini diantaranya adalah meningkatnya rasa haus (*polyuria*), meningkatnya rasa lapar (*polifagia*), seringnya buang air kecil, sering merasa kelelahan walaupun asupan makan dan minum sudah normal, daya penglihatan berkurang, dan lain sebagainya [3].

Banyak masyarakat, khususnya di Negara Indonesia memiliki tingkat kewaspadaan dan tingkat kesadaran yang rendah terhadap penyakit *Diabetes Mellitus* sehingga tidak sedikit masyarakat tidak menyadari bahwa dirinya sedang dalam resiko atau sudah terjangkit penyakit *Diabetes Mellitus* [4]. Hal ini disebabkan karena kurangnya wawasan atau pengetahuan masyarakat mengenai penyakit *Diabetes Mellitus*. Selain itu, ada factor lainnya seperti keterbatasan waktu dan biaya dalam melakukan konsultasi kepada dokter. Oleh karena itu, diperlukan sebuah pendekatan untuk mengatasi masalah tersebut dengan cara pembuatan sistem pakar yang berfungsi untuk membantu masyarakat melakukan konsultasi dan mendapatkan hasil diagnosis. Perancangan sistem pakar ini akan mengklasifikasikan pasien atau calon pasien penyakit *Diabetes Mellitus* ke dalam kategori *Diabetes Mellitus* tipe 1 atau *Diabetes Mellitus* tipe 2 berdasarkan inputan nilai indikator dari pengguna.

Untuk mengklasifikasikan *Diabetes Mellitus* tipe 1 atau *Diabetes Mellitus* tipe 2 diperlukan sebuah algoritma klasifikasi yang baik. Berdasarkan penelitian terdahulu, telah banyak algoritma klasifikasi yang digunakan dalam bidang medis dan klasifikasi penyakit yang telah diimplementasikan oleh beberapa peneliti diantaranya algoritma *Random Forest* [5] [6], *Support Vector Machine* (SVM) [7] [8], *K-Nearest Neighbor* (KNN) [9] [10], *Naïve Bayes* [9] [4], dan *Decision Tree C4.5* [11] [12].

Pada penelitian ini, sistem pakar dalam mengklasifikasikan pasien *Diabetes Mellitus* ke dalam kategori *Diabetes Mellitus* tipe 1 atau *Diabetes Mellitus* tipe 2 mengimplementasikan algoritma *Random Forest*. Algoritma ini dinilai baik dalam melakukan klasifikasi berdasarkan studi literatur yang digunakan dengan hasil nilai akurasi yang baik dibandingkan dengan algoritma klasifikasi lainnya. Dari hasil penelitian ini, diharapkan institusi kesehatan dapat secara mudah dan cepat mengklasifikasikan pasien terduga *Diabetes Mellitus* ke dalam kategori *Diabetes Mellitus* tipe 1 atau *Diabetes Mellitus* tipe 2. Pada akhirnya sistem pakar yang dirancang adalah sistem yang mampu mengklasifikasikan pasien terduga *Diabetes Mellitus* ke dalam kategori *Diabetes Mellitus* tipe 1 atau *Diabetes Mellitus* tipe 2 berbasis website dengan menggunakan algoritma *Random Forest*. Hasil akhir dari sistem pakar ini adalah sebuah diagnosa kepada pasien dan juga saran yang diberikan sistem kepada pasien berdasarkan hasil diagnosanya serta pendaftaran konsultasi berupa link online menuju Rumah Sakit Gotong Royong Surabaya.

## 2. Metodologi Penelitian

### 2.1. Pengumpulan Dataset

Penelitian ini menggunakan gabungan dataset primer dan dataset tersier dengan menggunakan satu variabel dependen yaitu diagnosa dan lima variabel independen yaitu gender, gula darah puasa (gdp), gula darah 2 jam pp (gd2pp), tekanan darah, dan usia. Dataset primer didapatkan dari pengambilan data secara langsung menggunakan *software* E-Rekam Medik di Rumah Sakit Gotong Royong Surabaya. Sedangkan dataset tersier didapatkan dari referensi publik [13]. Dataset primer yang didapatkan mengandung informasi pasien yang terdiagnosa dengan penyakit *Diabetes Mellitus* tipe 2 dengan total jumlah data sebanyak 40 baris data, sedangkan dataset tersier yang didapatkan mengandung informasi pasien yang terdiagnosa oleh penyakit *Diabetes Mellitus* tipe 1 dengan total jumlah data sebanyak 30 baris data.

**Table 1. Dataset sebelum Pre-Processing**

gender	gdp	gd2pp	tekanandarah	usia	diagnosa
Laki-laki	130 mg	251 mg	142 mm Hg	58	DM 1
Perempuan	126 mg	117 mg	120 mm Hg	69	DM 2
Perempuan	115 mg	256 mg	150 mm Hg	71	DM 2
Laki-laki	100 mg	146 mg	100 mm Hg	62	DM 2
Laki-laki	150 mg	285 mg	160 mm Hg	61	DM 1

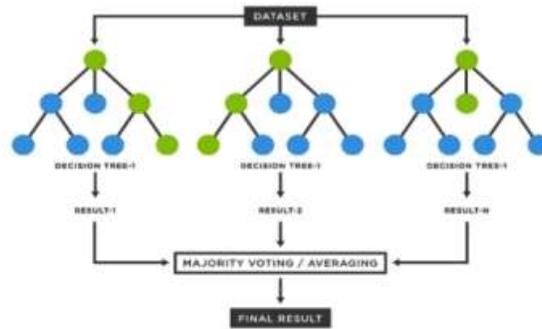
Tabel 1 menunjukkan gabungan antara dataset primer dan dataset tersier sebelum dilakukan *pre-processing* data. Dimana masih ada karakter-karakter yang tidak diperlukan seperti kata "mm", "mm Hg" dan kolom lainnya yang masih bersifat string. Pada Tabel 2 menunjukkan dataset yang sudah dilakukan *pre-processing* dimana semua nilai pada masing-masing kolom bernilai numerik saja. Pada kolom gender ada perubahan nilai dimana kolom yang sebelumnya bernilai "Laki-laki" diubah menjadi nilai numerik yaitu 1 sedangkan untuk "Perempuan" diubah menjadi nilai 2. Begitupun dengan kolom diagnosa, dimana yang sebelumnya bernilai "DM 1" diubah menjadi nilai 0, sedangkan "DM 2" diubah menjadi nilai 1.

**Table 2. Dataset setelah Pre-Processing**

gender	gdp	gd2pp	tekanandarah	usia	diagnosa
1	130	251	142	58	0
2	126	117	120	69	1
2	115	256	150	71	1
1	100	146	100	62	1
1	150	285	160	61	0

### 2.2. Algoritma *Random Forest*

Secara garis besar, algoritma *Random Forest* adalah sebuah algoritma yang meng-generate banyak *decision tree* dari parameter data yang disediakan. Kemudian dari banyak *decision tree* tersebut, algoritma ini akan melakukan voting dari setiap hasil *decision tree* untuk mendapatkan hasil yang terbaik [14] [15]. Dalam algoritma *Random Forest*, ada beberapa langkah-langkah yang dilakukan. Langkah pertama adalah melakukan *bootstrap sampling*. *Bootstrap sampling* adalah sebuah proses pengambilan sebanyak 2/3 data secara acak dari seluruh dataset. Sedangkan sisa 1/3 dari seluruh dataset akan digunakan untuk sampel *out-of-bag* (OOB) [5]. Selanjutnya dilakukan langkah kedua yaitu memilih sebuah indikator atau variabel terkuat yang paling tepat sebagai *root node*. Ulangi langkah pertama dan langkah kedua hingga terbentuk *decision tree* sebanyak *k*. Dari banyaknya *k decision tree* yang dibuat, akan dihitung nilai yang muncul terbanyak sebagai hasil akhir dari *Random Forest*.



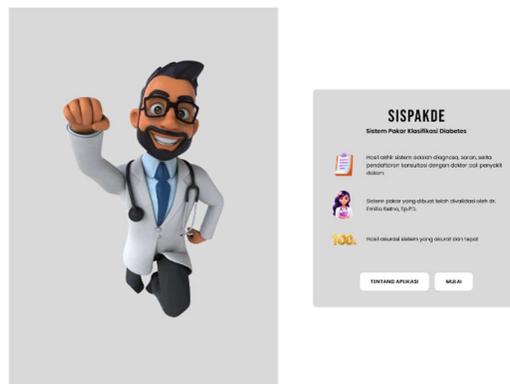
Gambar 1. Konsep *Random Forest* [8].

Sampel *out-of-bag* (OOB) adalah sebuah sampel yang tidak terpilih secara acak pada langkah *bootstrap sampling*. Sampel OOB tidak dibuang namun digunakan sebagai tolak ukur dalam memprediksi keakuratan *decision tree*. Jikalau ada sampel OOB yang tidak sesuai dengan prediksi pada *decision tree* yang dibuat, maka ini disebut dengan *error rate OOB*. *Error rate* dihitung dari jumlah klasifikasi yang salah pada hasil prediksi algoritma dari seluruh data [5].

### 3. Hasil dan Pembahasan

#### 3.1. Tampilan Desain Website

Perancangan aplikasi sistem pakar untuk mengklasifikasikan penderita *Diabetes Mellitus* ke dalam kategori *Diabetes Mellitus* tipe 1 atau *Diabetes Mellitus* tipe 2 adalah berbasis website. Tampilan yang dirancang tidak memerlukan login, dengan kata lain user bisa langsung memasukkan data-data yang dibutuhkan berdasarkan indikator atau variabel. Gambar 2. menunjukkan tampilan halaman *home* website. Gambar 3. menunjukkan halaman yang berfungsi untuk pengisian *form* berdasarkan variabel yang dibutuhkan. Gambar 4. Dan Gambar 5. adalah tampilan ketika *form* yang sudah diisi telah dianalisa oleh sistem dan menampilkan hasil diagnosa pasien, saran yang diberikan berdasarkan hasil diagnosa pasien, serta link pendaftaran konsultasi ke Rumah Sakit Gotong Royong Surabaya.



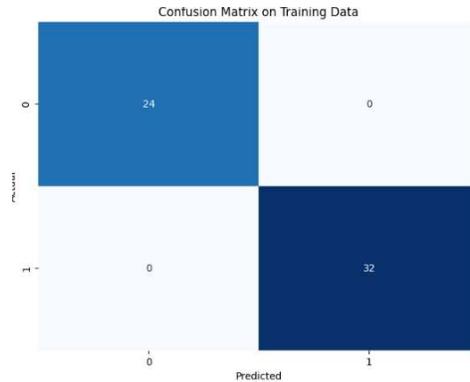
Gambar 2. Halaman Home Website



**Table 3.** Confussion Matrix

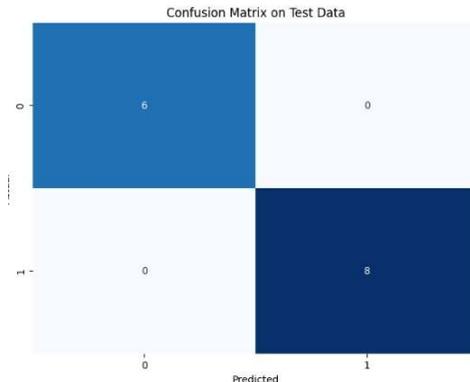
	Actual Class 0	Actual Class 1
Predicted Class 0	TN	FP
Predicted Class 1	FN	TP

Diketahui dataset yang digunakan dalam penelitian adalah sebanyak 70 baris data yang sudah melalui tahapan pre-processing data. Dari 70 baris data tersebut dilakukan pembagian data training sebanyak 80% atau sebanyak 56 baris data dan data testing sebanyak 20% atau sebanyak 14 baris data.



**Gambar 6.** Confussion Matrix pada Data Training

Gambar 6. adalah visualisasi grafik yang menggambarkan hasil confusion matrix dari data training yang mengambil sebanyak 80% dari seluruh dataset yaitu sebanyak 56 baris data. Berdasarkan Gambar 6. bisa dijelaskan bahwa terdapat sebanyak 24 baris data yang termasuk dalam kelas *True Negative* (TN) dan sebanyak 32 baris data termasuk dalam kelas *True Positive* (TP). Sedangkan untuk kelas lainnya yaitu *False Negative* (FN) dan *False Positive* (FP) tidak ada satupun data yang termasuk di dalamnya.



**Gambar 7.** Confussion Matrix pada Data Testing

Gambar 7. adalah visualisasi grafik yang menggambarkan hasil confusion matrix dari data testing yang mengambil sebanyak 20% dari seluruh dataset yaitu sebanyak 14 baris data. Berdasarkan Gambar 7. bisa dijelaskan bahwa terdapat sebanyak 6 baris data yang termasuk dalam kelas *True Negative* (TN) dan sebanyak 8 baris data termasuk dalam kelas *True Positive* (TP). Sedangkan untuk kelas lainnya yaitu *False Negative* (FN) dan *False Positive* (FP) tidak ada satupun data yang termasuk di dalamnya.

Berdasarkan hasil perhitungan evaluasi confusion matrix dengan menggunakan data training sejumlah 56 baris data dan data testing sejumlah 14 baris data maka didapatkan nilai akurasi sebesar 100%, nilai precision sebesar 1.00, nilai recall sebesar

1.00, dan nilai f1 score sebesar 1.00. Dari hasil tersebut disimpulkan bahwa klasifikasi pasien *Diabetes Mellitus* ke dalam kategori *Diabetes Mellitus* tipe 1 atau *Diabetes Mellitus* tipe 2 menggunakan algoritma *Random Forest* dapat dikatakan baik dan akurat.

#### 4. Kesimpulan

Dari hasil penelitian yang dilakukan yaitu perancangan aplikasi sistem pakar penyakit *Diabetes Mellitus* menggunakan algoritma *Random Forest* dan juga berdasarkan uraian dari bab-bab sebelumnya maka diperoleh beberapa kesimpulan sebagai berikut,

- Sistem pakar yang dirancang mampu mengklasifikasikan pasien *Diabetes Mellitus* ke dalam kategori *Diabetes Mellitus* tipe 1 atau *Diabetes Mellitus* tipe 2 menggunakan algoritma *Random Forest*.
- Berdasarkan evaluasi algoritma menggunakan confusion matrix, didapatkan hasil akurasi sebesar 100% dengan nilai precision sebesar 1.00, nilai recall sebesar 1.00, dan nilai f1 score sebesar 1.00 melalui perbandingan data training dan data testing sebanyak 80:20 dengan total baris data sebanyak 70 baris data.

#### Daftar Pustaka

- [1] R. Pahlevi, "Jumlah Penderita Diabetes Indonesia Terbesar Kelima di Dunia," 2021, [Online]. Available: <https://databoks.katadata.co.id/datapublish/2021/11/22/jumlah-penderita-diabetes-indonesia-terbesar-kelima-di-dunia>
- [2] Sriyanto and A. Ria Supriyatna, "Prediksi Penyakit Diabetes Menggunakan Algoritma *Random Forest*," *Ijccs*, vol. 17 No. 1, no. x, pp. 1–5, 2023, [Online]. Available: <https://jurnal.polsri.ac.id/index.php/teknika/article/view/6808>
- [3] D. Hardianto, "Telaah Komprehensif Diabetes Melitus: Klasifikasi, Gejala, Diagnosis, Pencegahan, Dan Pengobatan," *J. Bioteknol. Biosains Indones.*, vol. 7, no. 2, pp. 304–317, 2021, doi: 10.29122/jbbi.v7i2.4209.
- [4] H. T. Sihotang, "Perancangan Aplikasi Sistem Pakar Diagnosa Diabetes Dengan Metode Bayes," *J. Manik Penusa*, vol. 1, no. 1, pp. 36–41, 2017, [Online]. Available: <https://osf.io/preprints/inarxiv/znj3r/>
- [5] F. Anisha, Dodi Vionanda, Nonong Amalita, and Zilrahmi, "Application of *Random Forest* for The Classification *Diabetes Mellitus* Disease in RSUP Dr. M. Jamil Padang," *UNP J. Stat. Data Sci.*, vol. 1, no. 2, pp. 45–52, 2023, doi: 10.24036/ujsds/vol1-iss2/30.
- [6] A. C. Mawarni, R. Rusdah, L. L. Hin, and D. Anubhakti, "Deteksi Dini Gejala Awal Penyakit Diabetes Menggunakan Algoritma *Random Forest*," *IDEALIS Indones. J. Inf. Syst.*, vol. 6, no. 2, pp. 165–171, 2023, doi: 10.36080/idealis.v6i2.3018.
- [7] S. S. S. L. Dhita Diana Dewi, Nurul Qisthi, Zulfa Hidayah Satria Putri, "PERBANDINGAN METODE NEURAL NETWORK DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI DIAGNOSA PENYAKIT DIABETES," *Cerdika J. Ilm. Indones.*, vol. 3, no. 9, pp. 828–839, 2023, [Online]. Available: <https://cerdika.publikasiindonesia.id/index.php/cerdika/article/view/662>
- [8] A. Ghozali, H. Pratiwi, S. Sulistijowati Handajani Program Studi Statistika, F. Matematika dan Ilmu Pengetahuan Alam, and U. Sebelas Maret, "Implementasi Data Mining Menggunakan Metode *Random Forest* Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes," *J. Ilm. Pendidik. Mat. p*, vol. 11, no. 2, pp. 147–162, 2023, [Online]. Available: <https://jurnal.unikal.ac.id/index.php/Delta/index>
- [9] F. Nur Ikhromr, I. Sugiyarto, U. Faddillah, and B. Sudarsono, "Implementasi Data Mining Untuk Memprediksi Penyakit Diabetes Menggunakan Algoritma Naives Bayes dan K-Nearest Neighbor," *J. Inf. Technol. Comput. Sci.*, vol. 6, no. 1, pp. 416–428, 2023, [Online]. Available: <https://journal.ipm2kpe.or.id/index.php/INTECOM/article/download/5916/3632>
- [10] C. Susanto, T. Taufiq, E. Hasmin, and K. Aryasa, "Sistem Pakar Prediksi Penyakit Diabetes Menggunakan Metode K-NN Berbasis Android," *CogITo Smart J.*, vol. 8, no. 2, pp. 359–370, 2022, doi: 10.31154/cogito.v8i2.406.359-370.

- [11] B. A. R. P. Wahyu, A. F. Farazi, C. P. Mahendra, and R. K. Hapsari, "Klasifikasi Penderita Penyakit Diabetes Berdasarkan Decision Tree Menggunakan Algoritma C4," *INTEGER J. Inf. Technol.*, vol. 8, no. 1, pp. 80–89, 2023, [Online]. Available: <http://ejurnal.itats.ac.id/integer/article/view/4423>
- [12] F. M. Hana, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 32–39, 2020, doi: 10.47970/siskom-kb.v4i1.173.
- [13] I. Syahputra, *Klasifikasi Penyakit Diabetes Melitus Tipe 1 dan Diabetes Melitus Tipe 2 dengan menggunakan Metode Analisis Diskriminan*. 2020.
- [14] R. Annisa, "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung," *J. Tek. Inform. Kaputama*, vol. 3, no. 1, pp. 22–28, 2019, [Online]. Available: <https://jurnal.kaputama.ac.id/index.php/JTIK/article/view/141/156>
- [15] A. B. Wibisono and A. Fahrurrozi, "Perbandingan Algoritma Klasifikasi Dalam Pengklasifikasian Data Penyakit Jantung Koroner," *J. Ilm. Teknol. dan Rekayasa*, vol. 24, no. 3, pp. 161–170, 2019, doi: 10.35760/tr.2019.v24i3.2393.