

Klasifikasi Penyakit Stroke Menggunakan Support Vector Machine (SVM) dan Particle Swarm Optimization (PSO)

Yuwike Ayuningtyas¹, I Made Suartana²,

^{1,3} S1 Teknik Informatika, Universitas Negeri Surabaya

yuwike.19009@mhs.unesa.ac.id

madesuartana@unesa.ac.id

Abstrak— Stroke adalah penyakit yang tidak menular tetapi menjadi salah satu penyakit yang paling sering menyebabkan kematian dan kecacatan di dunia termasuk negara kita Indonesia. Ada beberapa faktor yang dapat menyebabkan penyakit stroke yaitu faktor risiko stroke yang dapat diubah dan tidak dapat diubah. Faktor risiko stroke yang tidak dapat diubah yakni jenis kelamin, usia, genetik dan ras. Untuk faktor risiko stroke yang dapat diubah yakni obesitas, merokok, hipertensi (tekanan darah), diabetes militus, dan gaya hidup tidak sehat. Menurut Kementerian Kesehatan RI penderita penyakit stroke mengalami peningkatan yang signifikan sebesar 10.9% pada tahun 2018 dibandingkan dengan tahun 2013 jumlah penderita penyakit stroke sebanyak 7.0% (Kementerian Kesehatan RI, dalam Abdul 2022). Tujuan dari penelitian ini yaitu dapat mengetahui hasil preprocessing dataset, dapat mengetahui hasil klasifikasi penyakit stroke menggunakan metode Support Vector Machine, dan dapat mengetahui hasil perbedaan akurasi dari klasifikasi penyakit stroke menggunakan metode Support Vector Machine dan Support Vector Machine berbasis Particle Swarm Optimization. Sebelum melakukan klasifikasi dataset stroke dilakukan preprocessing terlebih dahulu seperti mengatasi data tidak seimbang menggunakan teknik SMOTE sehingga data siap dimasukkan kedalam model klasifikasi. Dataset stroke yang digunakan memiliki 9 fitur atau atribut dan berjumlah 5.510 data sebelum di upsampling dan menjadi 9.772 data setelah di upsampling. Hasil penelitian ini memperoleh sebuah model untuk sistem sederhana prediksi untuk klasifikasi penyakit stroke, serta mendapatkan hasil nilai akurasi yaitu 88% dengan metode SVM dengan teknik SMOTE dan 95% untuk pengujian menggunakan metode PSO-SVM dengan teknik SMOTE.

Kata Kunci— Stroke, klasifikasi, support vector machine, particle swarm optimization, smote.

I. PENDAHULUAN

Stroke adalah penyakit yang tidak menular tetapi menjadi salah satu penyakit yang paling sering menyebabkan kematian dan kecacatan di dunia termasuk negara kita Indonesia (Jonhson dalam Suci, 2022) [1]. Organisasi Kesehatan Dunia atau WHO mengatakan penyakit stroke terjadi karena adanya gangguan fungsi otak secara tiba-tiba. Wolfe (dalam Haris, 2022) mengatakan perkiraan jumlah kematian di seluruh dunia sebanyak 4.5 juta setiap tahunnya hal ini diakibatkan oleh penyakit stroke. Penyakit stroke sendiri adalah penyebab kematian urutan nomor tiga di Amerika [2]. Menurut Kementerian Kesehatan RI penderita penyakit stroke mengalami peningkatan yang signifikan sebesar 10.9% pada tahun 2018 dibandingkan dengan tahun 2013 jumlah penderita

penyakit stroke sebanyak 7.0% (Kementerian Kesehatan RI, dalam Abdul 2022) [3].

Tingginya penderita dan kematian yang disebabkan oleh penyakit stroke disebabkan oleh ketidaktahuan masyarakat mengenai penyakit stroke dan akhirnya banyak masyarakat yang mengabaikan gejala-gejala penyakit stroke ini, padahal jika masyarakat dapat mengetahui gejala penyakit stroke secepat mungkin dan segera ditangani maka kemungkinan para penderita stroke dapat diobati dengan tepat dan berangsur pulih. Disamping itu ada beberapa faktor yang dapat menyebabkan penyakit stroke yaitu faktor risiko stroke yang dapat diubah dan tidak dapat diubah. Faktor risiko stroke yang tidak dapat diubah yakni jenis kelamin, usia, genetik dan ras. Untuk faktor risiko stroke yang dapat diubah yakni obesitas, merokok, hipertensi (tekanan darah), diabetes militus, dan gaya hidup tidak sehat (Kenny,2022) [4].

Dengan adanya masalah di atas penulis ingin menerapkan perkembangan teknologi di bidang artificial intelligence atau kecerdasan buatan. Kecerdasan buatan terdiri dari beberapa cabang salah satunya yakni cabang pembelajaran mesin atau machine learning. Klasifikasi penyakit stroke ini perlu dilakukan agar dapat memperoleh hasil secara akurat. Keakuratan dari hasil prediksi ini diharapkan dapat membantu tenaga kesehatan untuk mengambil Tindakan yang tepat.

Pada penelitian sebelumnya mengenai klasifikasi penyakit stroke dalam penelitiannya yang menggunakan metode Random Forest hasil dari penelitiannya yaitu kinerja rata-rata akurasi mencapai 90% (Diah,2021) [5]. Kemudian penelitian yang menggunakan metode Support Vector Machine dan Particle Swarm Optimization untuk memprediksi penyakit kanker payudara dalam penelitiannya menghasilkan akurasi sebesar 97.61% (Ari,2021) [6].

Berdasarkan penelitian yang telah dilakukan oleh peneliti sebelumnya penulis ingin menggunakan metode Support Vector Machine dan Particle Swarm Optimization serta menggunakan Teknik SMOTE untuk mengatasi data tidak seimbang pada dataset penyakit stroke untuk meningkatkan hasil kinerja akurasi pada klasifikasi penyakit stroke.

II. METODOLOGI

Menurut Ibnu Sina (2022) metode penelitian merupakan cara atau metode ilmiah yang digunakan untuk memperoleh informasi yang valid dengan maksud agar ditemukan [7], dikembangkan, maupun dibuktikan, suatu pengetahuan tertentu sehingga pada saat tertentu dapat digunakan untuk

memahami, memecahkan, dan mengantisipasi masalah dalam suatu kajian.

Menurut pengertian diatas penulis menyimpulkan metode penelitian yaitu metode ilmiah yang dapat digunakan untuk memecahkan atau mendapatkan informasi yang valid. Untuk melakukan klasifikasi dan dimudahkan dalam penelitian ini penulis membuat langkah-langkah atau tahapan yang akan dilakukan di dalam penelitian. Berikut tahapan dari penelitian:

Gambar 1 Tahap Penelitian

A. Pengumpulan Data

Pada pengumpulan data penulis menggunakan cara pengumpulan data secara sekunder. Data sekunder sendiri adalah metode pengumpulan data yang hasil datanya diperoleh secara tidak langsung kepada sumbernya. Data yang digunakan penulis sumbernya berasal dari data publik prediction stroke yaitu merupakan dataset dari Website Kaggle. Dataset terdiri dari gender, age, hypertension, heart disease, residence type, glucose, BMI. Data ini bisa digunakan sebagai data training, sedangkan untuk data testing hanya diambil data hasil atau keluaran. Dataset yang penulis gunakan berjumlah 5.110 data. Berikut detail dari atribut dataset untuk penyakit stroke:

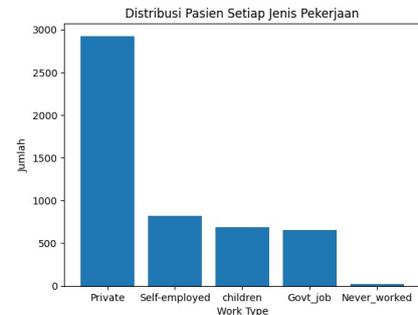
TABEL I
 ATRIBUT DATASET STROKE

No	Atribut	Deskripsi
1	Gender	Jenis Kelamin
2	Age	Umur
3	Hypertension	Tekanan Darah Tinggi
4	Heart Disease	Penyakit Jantung
5	Ever Marriage	Pernikahan
6	Avg Glucose Level	Kadar Gula
7	BMI	Berat Tubuh
8	Smoking Status	Status Merokok
9	Residence Type	Tempat Tinggal

B. Analisa Data

Hasil analisa data ini diperoleh dari hasil pengolahan dataset stroke yang berasal dari website kaagle.com. Analisa data disini untuk mengetahui informasi-informasi yang lebih detail terkait dataset yang akan digunakan untuk penelitian ini. Analisa data dalam penelitian ini memiliki beberapa tahapan sebagai berikut:

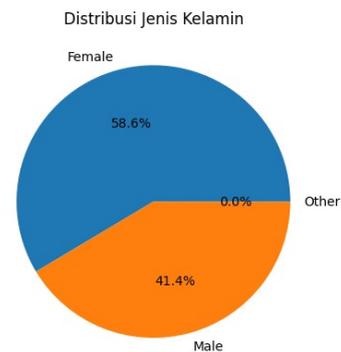
1. Diagram Distribusi Pasien Stroke Setiap Jenis Pekerjaan



Gambar 2 Diagram Distribusi Pasien Stroke Setiap Pekerjaan

Pada gambar 2 merupakan gambar diagram distribusi jumlah pasien stroke pada fitur *worktype* yang ada pada dataset stroke.

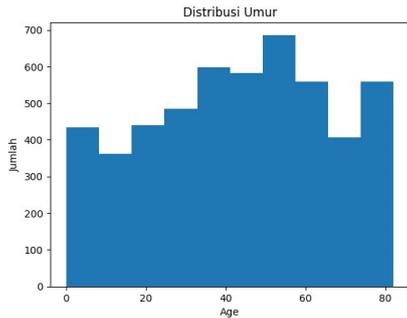
2. Diagram Distribusi Jenis Kelamin Pasien Stroke



Gambar 3 Diagram Distribusi Jenis Kelamin Pasien Stroke

Pada gambar 3 merupakan gambar diagram distribusi jenis kelamin pasien stroke pada fitur *gender* yang ada pada dataset stroke. Dimana untuk *female* 58,6% dan *male* 41,4%.

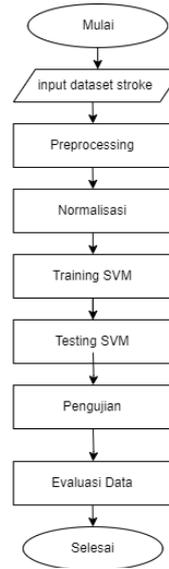
3. Diagram Distribusi Umur Pasien Stroke



Gambar 4 Diagram Distribusi Umur Pasien Stroke

Pada gambar 4 merupakan gambar diagram distribusi umur pasien stroke pada fitur *worktype* yang ada pada dataset stroke

4. Diagram Kemunculan Stroke Berdasarkan Jenis Kelamin



Gambar 6 Alur Klasifikasi SVM-SMOTE

2. Alur Klasifikasi PSO-SVM-SMOTE

Alur untuk tahap klasifikasi penyakit stroke menggunakan metode PSO SVM dengan teknik SMOTE dapat dilihat pada gambar 7.

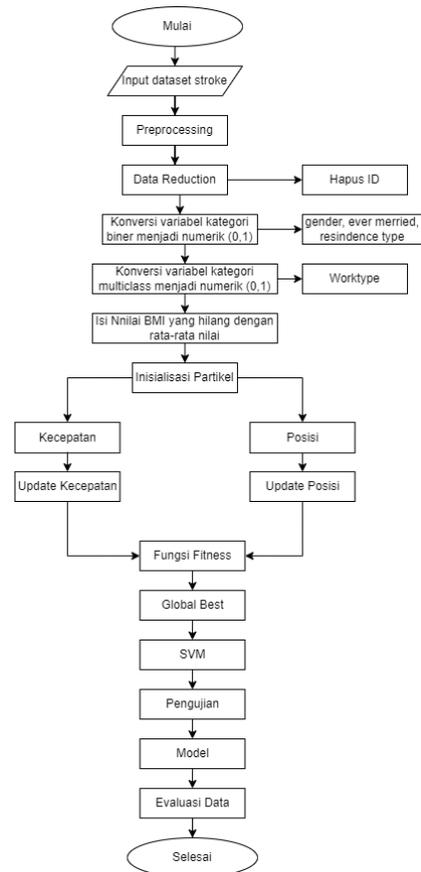
Gambar 5 Kemunculan Penyakit Stroke

Pada gambar 5 merupakan gambar kemunculan penyakit stroke berdasarkan jenis kelamin pada dataset stroke. Dimana lebih banyak jenis kelamin *female* (perempuan) yang terkena penyakit stroke.

C. Desain Sistem

1. Alur Klasifikasi SVM-SMOTE

Alur untuk tahap klasifikasi penyakit stroke menggunakan metode SVM dengan Teknik SMOTE dapat dilihat pada gambar 6.



Gambar 7 Alur Klasifikasi PSO-SVM-SMOTE

D. Preprocessing

Pada tahap preprocessing data ini dilakukan pengolahan data seperti menambahkan data dan menghapus data dengan tujuan agar data yang digunakan menjadi lebih efektif. Ada beberapa tahap yang dilakukan untuk preprocessing data sebagai berikut:

1. Menghapus Kolom ID
Tahap pertama preprocessing data yaitu penghapusan kolom id pada dataset stroke. Penghapusan kolom id ini dikarenakan kolom id pada dataset tidak mempengaruhi hasil dari klasifikasi yang dilakukan pada penelitian ini.
2. Konversi Variabel Biner ke Numerik
Pada tahap ini dilakukan pengkonversian variabel atau fitur yang kategorinya masuk ke dalam biner. Dimana fitur yang berada di dalam dataset memiliki dua jawaban lalu dikonversi menjadi numerik 0 untuk tidak dan 1 untuk iya. Variabel-variabel tersebut yaitu *gender*, *ever married*, dan *residence type*.
3. Konversi Variabel *Multiclass* ke Numerik
Pada tahap ini dilakukan pengkonversian variabel atau fitur yang kategorinya masuk ke dalam *multiclass*. Dimana fitur yang berada di dalam dataset memiliki banyak nilai lalu dikonversi menjadi numerik. Variabel-variabel tersebut yaitu *work type* dan *smoking status*.
4. Mengisi Nilai Hilang BMI
Pada tahap ini dilakukan pengisian nilai yang hilang pada kolom BMI. Nilai yang hilang pada kolom BMI ini diisi dengan rata-rata.
5. Normalisasi
Pada tahap ini dilakukan normalisasi dataset dengan cara mengubah atribut numerik ke rentang yang sama yang bertujuan untuk menghindari perbedaan skala agar setiap atribut memiliki pengaruh yang sebanding pada model.
6. Data *Imbalance*
Pada tahap dilakukan *handle data imbalance* atau mengatasi data yang tidak seimbang pada dataset stroke yang digunakan pada penelitian ini, bertujuan agar model klasifikasi yang akan digunakan menjadi lebih akurat. Teknik yang digunakan untuk mengatasi data yang tidak seimbang yaitu SMOTE (*Synthetic Minority Over-sampling Technique*). Sebelum melakukan SMOTE ada beberapa tahap yang harus dilakukan seperti menginstall package dan mengimport modul.

7. Pembagian Dataset

Pada tahap ini dilakukan pembagian dataset stroke setelah menggunakan Teknik SMOTE menjadi dua yaitu data *training* dan data *testing*. Hasil dari pembagian dataset stroke yaitu 7.777 untuk data *training* dan 1.945 untuk data *testing*.

E. Pengujian

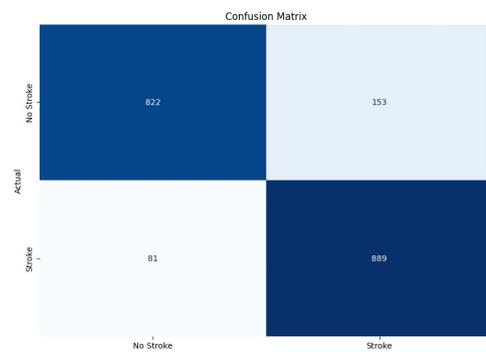
Pada tahap pengujian pada penelitian klasifikasi penyakit stroke menggunakan metode SVM dan PSO-SVM antara lain yaitu menentukan nilai hasil klasifikasi seperti *precision*, *recall*, dan *f1 score*.

III. HASIL

Pada tahap ini merupakan hasil pengujian atau implementasi dari klasifikasi penyakit stroke menggunakan metode Support Vector Machine (SVM) dengan Teknik SMOTE dan metode Particle Swarm Optimization (PSO)-Support Vector Machine (SVM) dengan Teknik SMOTE.

A. Hasil Pengujian SVM-SMOTE

Pada tahap ini melakukan pengujian yaitu klasifikasi penyakit stroke menggunakan metode SVM dengan menggunakan SMOTE sebagai cara mengatasi data yang tidak seimbang. Untuk hasil pengujian metode SVM dengan menggunakan Teknik SMOTE dapat ditampilkan pada gambar 8 yaitu hasil *confusion matrix* dari klasifikasi dataset penyakit stroke.



Gambar 8 Hasil Confusion Matrix SVM-SMOTE

Berdasarkan pada gambar 8 merupakan hasil confusion matrix dapat diketahui untuk nilai *precision*, *recall*, dan *f1 score* untuk pengujian SVM menggunakan SMOTE:

1. Berikut untuk hasil True Positive (TP), False Positive (FP), False Negative (FN), True Negative (TN), dan Total Data.
TP= 822
FP = 153
FN = 81
TN = 889
Total data = 1945

2. Menghitung Precision:
 $= TP / (FP + TP)$
 $= 822 / (153 + 822)$
 $= 0.85$
3. Menghitung Recall:
 $= TP / (FN + TP)$
 $= 822 / (81 + 822)$
 $= 0.92$
4. F1 Score:
 $= 2 * (Recall * Precision) / (Recall + Precision)$
 $= 2 * (7.820 / 177) \times 100 \%$
 $= 88\%$

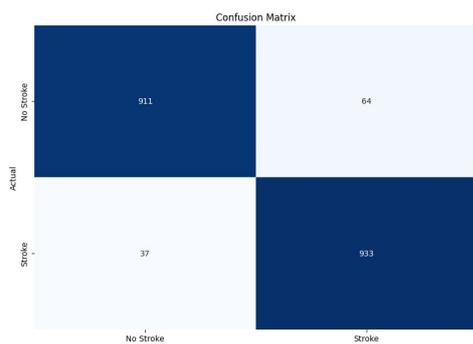
Hasil perhitungan *confusion matrix* dari pengujian SVM menggunakan Teknik SMOTE untuk precision 85%, recall 92% dan f1 score 88%. Gambar 5 merupakan hasil perhitungan klasifikasi menggunakan fungsi python.

	precision	recall	f1-score	support
0	0.91	0.84	0.88	975
1	0.85	0.92	0.88	970
accuracy			0.88	1945
macro avg	0.88	0.88	0.88	1945
weighted avg	0.88	0.88	0.88	1945

Gambar 9 Tampilan Hasil Pengujian SVM-SMOTE

B. Hasil Pengujian PSO-SVM dengan SMOTE

Pada tahap ini melakukan pengujian yaitu klasifikasi penyakit stroke menggunakan metode PSO-SVM dengan menggunakan SMOTE sebagai cara mengatasi data yang tidak seimbang. Untuk hasil pengujian metode SVM dengan menggunakan Teknik SMOTE dapat ditampilkan pada gambar 5 yaitu hasil *confusion matrix* dari klasifikasi dataset penyakit stroke



Gambar 10 Hasil Confusion Matrix PSO-SVM-SMOTE

Berdasarkan pada gambar 6 merupakan hasil *confusion matrix* dapat diketahui untuk nilai precision, recall, dan f1 score untuk pengujian PSO-SVM menggunakan SMOTE:

1. Berikut untuk hasil True Positive (TP), False Positive (FP), False Negative (FN), True Negative (TN), dan Total Data.
 TP = 911
 FP = 64
 FN = 37
 TN = 933
 Total data = 1.945
2. Menghitung Precision:
 $= TP / (FP + TP)$
 $= 911 / (64 + 911)$
 $= 0.94$
3. Menghitung Recall:
 $= TP / (FN + TP)$
 $= 911 / (37 + 911)$
 $= 0.96$
4. F1 Score:
 $= 2 \times (Recall \times Precision) / (Recall + Precision)$
 $= 2 \times (96\% \times 94\%) / (96\% + 94\%)$
 $= 95\%$

Hasil perhitungan *confusion matrix* dari pengujian SVM menggunakan Teknik SMOTE untuk precision 94%, recall 96% dan f1 score 95%. Gambar 7 merupakan hasil perhitungan klasifikasi menggunakan fungsi python.

	precision	recall	f1-score	support
0	0.96	0.93	0.95	975
1	0.94	0.96	0.95	970
accuracy			0.95	1945
macro avg	0.95	0.95	0.95	1945
weighted avg	0.95	0.95	0.95	1945

Gambar 11 Tampilan Hasil Pengujian SVM-SMOTE

Selanjutnya dapat dibandingkan hasil dari pengujian untuk klasifikasi penyakit stroke menggunakan SVM-SMOTE dan PSO-SVM-SMOTE hasil perbandingan dari pengujian dapat dilihat pada tabel II:

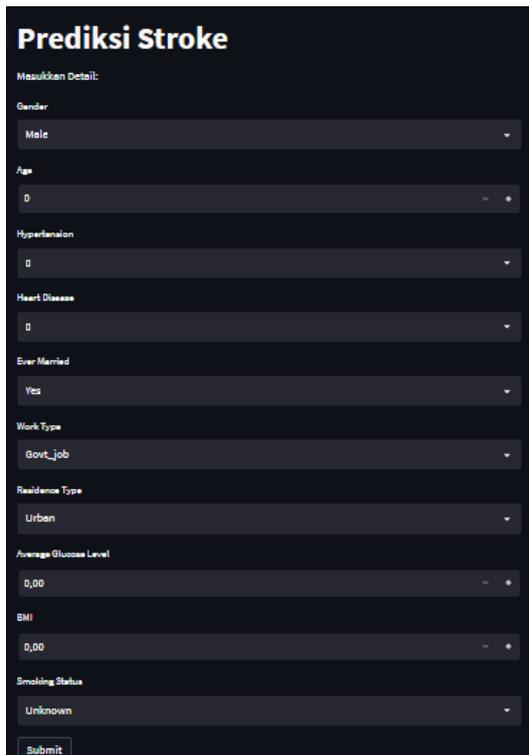
TABEL II
PERBANDINGAN HASIL PENGUJIAN

No	Metode	Hasil
1	SVM-SMOTE	Precision: 85% Recall: 92% F1 Score: 88%
2	PSO-SVM-SMOTE	Precision: 94% Recall: 96% F1 Score: 95%

C. Hasil Implementasi GUI Deteksi Dini Penyakit Stroke

Pada gambar 12 merupakan hasil implementasi GUI untuk deteksi dini penyakit stroke. GUI untuk deteksi dini penyakit

stroke ini diambil dari model yang hasil pengujiannya paling baik yaitu model dari PSO_SVM dengan Teknik SMOTE.



Gambar 12 Tampilan Hasil GUI Deteksi Dini Penyakit Stroke

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Pada penelitian yang berjudul Klasifikasi Penyakit Stroke Menggunakan Support Vector Machine (SVM) dan Particle Swarm Optimization (PSO) setelah melakukan beberapa pengujian mendapatkan beberapa kesimpulan sebagai berikut:

Hasil dari preprocessing dataset stroke yang akan digunakan untuk klasifikasi yaitu penghapusan atribut id karena tidak digunakan dan tidak mempengaruhi hasil klasifikasi penyakit stroke, pada dataset stroke terdapat nilai hilang pada fitur atau atribut BMI, maka dilakukan pengisian nilai BMI yang hilang dengan rata-rata, dan dataset stroke yang akan digunakan untuk klasifikasi penyakit stroke ternyata tidak seimbang atau imbalance sehingga dilakukan SMOTE yang bertujuan untuk mengatasi data yang tidak seimbang.

Pada pengujian klasifikasi penyakit stroke menggunakan metode SVM dengan SMOTE mendapatkan nilai uji yaitu 85%, Recall 92%, dan F1 Score 88%.

Pada pengujian klasifikasi penyakit stroke menggunakan metode PSO-SVM dengan SMOTE mendapatkan nilai uji yaitu 94%, Recall 96%, dan F1 Score 95%.

B. Saran

Dalam penelitian yang telah dilakukan terdapat beberapa saran yang bisa digunakan untuk mengembangkan penelitian selanjutnya:

Melakukan perbandingan dengan berbagai metode dengan tujuan untuk mengetahui metode yang lebih baik dalam masalah klasifikasi penyakit stroke.

Melakukan lebih banyak pengujian agar mendapatkan akurasi yang lebih baik.

Dapat merealisasikan pembuatan aplikasi agar dapat digunakan banyak orang.

UCAPAN TERIMA KASIH

Pada kesempatan ini penulis ingin mengucapkan dan menyampaikan rasa terima kasih terhadap Tuhan Yang Maha Esa dan berbagai pihak yang telah memberikan bantuan dan dukungan yakni:

REFERENSI

- [1] Idsam, S. A. (2022). Hal-Hal Yang Ada Hubungan Dengan Stroke Iskemik Pada Penderita Di Beberapa Lokasi Di Wilayah Indonesia Periode Tahun 2010 Sampai Dengan Tahun 2020.
- [2] Bugis, H. (2022). Metode Naïve Bayes Untuk Memprediksi Penyakit Stroke. *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, 6(1), 8-14.
- [3] Thalib, A. H. S., & Dimara, H. (2022). Efektifitas Mirror Therapy Terhadap Peningkatan Kekuatan Otot Pada Pasien Post Stroke: Literature Review. *Imj (Indonesian Midwifery Journal)*, 5(1), 11-15
- [4] Rizki, A. M., & Nurlaili, A. L. (2020). Algoritme Particle Swarm Optimization (PSO) untuk Optimasi Perencanaan Produksi Agregat Multi-Site pada Industri Tekstil Rumahan. *Journal of Computer, Electronic, and Telecommunication (COMPLETE)*, 1(2).
- [5] Diah, A. A. (2021). Metoda Random Over-Under Sampling Dan Random Forest Untuk Klasifikasi Prediksi Stroke.
- [6] Maulana, A., Nugroho, A., & Romli, I. (2021). Optimalisasi Support Vector Machine Menggunakan Particle Swarm Optimization Untuk Mendiagnosa Penyakit Kanker Payudara. *Journal of Practical Computer Science*, 1(2), 1-11.
- [7] Sina, Ibnu. (2022) Metodologi Penelitian. Diedit oleh Putranto, Agung T. CV Widina Media Utama.