

Klasifikasi Penyakit *Stroke* dengan Metode *Support Vector Machine* (SVM)

Siskawati Rahayu¹, Yuni Yamasari²

^{1,2}Jurusan Teknik Informatika, Universitas Negeri Surabaya

¹siskawati.17051204012@mhs.unesa.ac.id

²yuniyamasari@unesa.ac.id

Abstrak— *Stroke* adalah penyakit yang menyerang pada gangguan fungsi syaraf dan menyebabkan gangguan fungsi seperti gangguan penglihatan atau mata, bicara pelo atau cedal, mobilitas terbatas dan kelumpuhan pada wajah atau ekstremitas. Diseluruh dunia, *stroke* adalah penyakit ketiga yang menyebabkan gangguan syaraf. Untuk itu penelitian ini memfokuskan pada domain tersebut dengan tujuan untuk menemukan model terbaik. Ujicoba dilakukan dengan 14 skenario dan ukuran kinerja model akurasi, presisi dan recall. Selanjutnya, metode yang diterapkan adalah *Support Vector Machine* dengan kernel *linear*, *polynomial*, *sigmoid* dan *RBF*. Hasil klasifikasi memperlihatkan bahwa kernel *linear* menghasilkan nilai tinggi pada perbandingan 80:20, dengan nilai akurasi 75.73%, nilai presisi 74.07%, dan nilai recall 56.16%. Pada kernel *RBF* menghasilkan nilai tinggi pada perbandingan 90:10, dengan nilai akurasi 73.38%, nilai presisi 71.79% dan nilai recall 55.85%. Pada kernel *polynomial* menghasilkan nilai tinggi pada perbandingan 80:20, dengan nilai akurasi 78.86%, nilai presisi 73.98% dan nilai recall 56.75%. Kemudian pada kernel *sigmoid* menghasilkan nilai tinggi dengan perbandingan 10:90, yaitu dengan nilai akurasi 46.14%, nilai presisi 39.84% dan nilai recall 48.13%. Sehingga nilai tertinggi dari pengujian empat kernel dan dengan perbandingan 14 rasio, nilai tertinggi dihasilkan oleh kernel *polynomial* dengan nilai akurasi 78.86%, nilai presisi 73.98% dan nilai recall 56.75% pada perbandingan rasio 80:20.

Kata Kunci— *Stroke*, Kernel *Support Vector Machine*, Klasifikasi

I. PENDAHULUAN

Stroke adalah penyakit yang menyerang gangguan fungsi syaraf karena adanya sumbatan oleh gumpalan darah [1]. Salah satu penyakit ketiga didunia yang menyebabkan kecacatan yaitu *stroke* [3]. Beberapa faktor yang mempengaruhi terjadinya *stroke* antara lain faktor keturunan, ras, hipertensi, hiperkolesterolemia, obesitas, konsumsi alkohol, stres, pola makan yang buruk, umur, jenis kelamin, dan penggunaan pil KB. Namun, di antara faktor-faktor tersebut, hanya tekanan darah tinggi yang dapat mempengaruhi *stroke* secara signifikan dibandingkan banyak penyebab lainnya [1].

Pada tahun 2020 penelitian terkait hipertensi dilakukan oleh Putri Nadhirah Puspitasari dengan judul “Hubungan Hipertensi terhadap Kejadian *Stroke*”. Dalam penelitian tersebut dijelaskan bahwa dari beberapa faktor penyebab penyakit *stroke*, hipertensi adalah satu-satunya penyebab utama yang dapat mempengaruhi terjadinya *stroke* [1].

Kemudian pada tahun 2023, Nabilla Yolanda Paramitha dkk melakukan penelitian dengan judul “Klasifikasi Penyakit *Stroke* Menggunakan Metode *Naïve Bayes*”. Tujuannya yaitu untuk mengklasifikasi pasien *stroke* berdasarkan data latih dan

data survei. Hasil nilai akurasi dalam proses klasifikasi sebesar 80% di dapat dari proporsi data pelatihan dan data pemeriksaan dengan perbandingan 80:20 [2].

Klasifikasi adalah proses memasukkan suatu anggota ke dalam kelas tertentu, namun kelas tersebut ditentukan terlebih dahulu, kemudian anggota-anggota kelas tersebut ditempatkan dalam suatu kelas berdasarkan karakteristik data yang sama. [2]. Beberapa jenis algoritma berbeda yang dapat digunakan untuk klasifikasi, diantaranya DT, SVM, NN, K-NN dan NB. DT adalah *Decision Tree*, SVM adalah *Support Vector Machine*, NN adalah *Neural Network*, K-NN adalah *K-Nearest Network* dan NB adalah *Naïve Bayes*.

Penelitian terdahulu dilakukan oleh Nur Ghaniaviyanto Ramadhan pada tahun 2021 dengan judul “Klasifikasi Data Malaria Menggunakan Metode *Support Vector Machine* (SVM)”, pada penelitian ini membandingkan kedua metode tersebut. Dalam proses ini algoritma SVM mengungguli algoritma NB dengan nilai akurasi rata – rata 25%[11].

Pada tahun 2021 Siti Silvia Arifin et al melakukan penelitian dengan judul “Klasifikasi Penyakit Kanker Serviks Menggunakan Algoritma *Support Vector Machine* (SVM)”. Pada penelitian menggunakan perbandingan 80:20. Pada penelitian menggunakan tools orange sebagai data test dan python digunakan sebagai data train. Hasil pengujian dengan tools orange menghasilkan akurasi 92.9% dan dengan python menghasilkan nilai akurasi 87% [4].

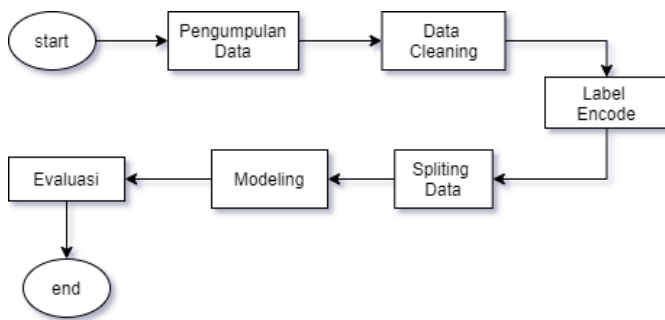
Pada tahun 2021, penelitian terkait klasifikasi juga dilakukan oleh Abu Wildan Mucholladin et al dengan judul “Klasifikasi Penyakit Diabetes Menggunakan Metode *Support Vector Machine* (SVM)”. Dalam penelitian ini dibagi menjadi dua model, yang pertama adalah model *benchmark* diimplementasikan dengan algoritma SMO (*Sequential Minimal Optimizatio*) dan kedua adalah model *scratch* yang diimplementasikan dengan algoritma *Sequential Learning*. Nilai akurasi dengan model *benchmark* memiliki nilai 0.87 dan model *scratch* 0.78 [10].

Ada berbagai algoritma atau teknik yang dapat digunakan untuk proses klasifikasi, menurut beberapa penelitian sebelumnya dan literature. Penelitian ini menyelidiki kernel *sigmoid*, *Gaussian Radial Basic Function* (RBF), *linear* dan *polynomial*. Beberapa literature sebelumnya menunjukkan bahwa metode SVM (*Support Vector Machine*) menghasilkan akurasi yang cukup baik. Oleh karena itu, judul penelitian ini adalah “Klasifikasi Penyakit *Stroke* dengan Metode *Support Vector Machine* (SVM)”.

II. METODOLOGI PENELITIAN

Metode SVM (*Support Vector Machine*) digunakan untuk klasifikasi dengan studi kasus penyakit *stroke*. Klasifikasi yang digunakan pada metode SVM (*Support Vector Machine*) adalah kernel sigmoid, linear, RBF, dan polynomial.

Tujuan dari penelitian ini yaitu untuk menentukan nilai akurasi, presisi dan recall pada data penyakit *stroke* dengan menggunakan metode SVM (*Support Vector Machine*) berdasarkan penerapan kernel linear, polynomial, sigmoid dan juga RBF. Ada beberapa alur dalam penelitian, alur tersebut digambarkan dalam Gbr 1.



Gbr 1 Flowchart Implementasi Penelitian

A. Pengumpulan data

Data dari penelitian ini berjenis data public, data tersebut diambil dari website kaggle. Dataset yang digunakan yaitu *Stroke Prediction Dataset* kemudian diunduh dalam bentuk csv lalu disimpan dengan nama dataset. Pada dataset stroke terdiri dari 5110 data dengan 12 atribut. Atribut tersebut diantaranya seperti tabel 1.

Tabel 1 Atribut Dataset

Atribut	Keterangan
<i>Id</i>	Identitas unik
<i>Gender</i>	Jenis kelamin
<i>Age</i>	Umur
<i>Hypertension</i>	0 = tidak hipertensi, 1 = ya hipertensi
<i>Avg_glucose_level</i>	Nilai rata-rata glukosa
<i>Residence_type</i>	Tempat tinggal pasien
<i>Work_type</i>	Pekerjaan pasien
<i>Ever_married</i>	Pernah menikah atau tidak
<i>Heart_disease</i>	0 tidak, 1 ya
<i>Bmi</i>	Indeks massa tubuh
<i>Smoking_status</i>	Kategori merokok pada pasien
<i>Stroke</i>	0 tidak, 1 ya

B. Data Cleaning

Data cleaning adalah proses yang digunakan untuk menghapus atau memperbaiki suatu data yang salah, rusak, formatnya salah, data duplikat atau tidak lengkap dari kumpulan data. Proses membersihkan data sangatlah penting selama proses pengolahan dan analisis data, karena kualitas hasil pengolahan data sangatlah bergantung pada data yang telah dimasukkan. Ada beberapa manfaat dari *cleaning* data antara lain :

1. Hasil analisis data lebih akurat karena data yang bersih dapat mengurangi resiko munculnya informasi yang salah akibat data yang kotor.
2. Mendorong pengambilan keputusan yang lebih baik.
3. Data yang bersih dapat mengurangi kemungkinan mengulang proses analisis data akibat dari data yang kotor atau eror, sehingga dapat menyebabkan lebih sedikit kesalahan.

C. Label Encode

Label Encode adalah teknik yang digunakan untuk mengubah kolom kategoris menjadi kolom numerik, sehingga dapat disesuaikan dengan suatu model pembelajaran yang hanya mengambil data numerik. Pengkodean mengubah data kategori menjadi data numerik dengan memberikan kode unik pada setiap kelas data. *Label encode* merupakan salah satu *library python* dari *Scikit Learn*.

D. Splitting Data

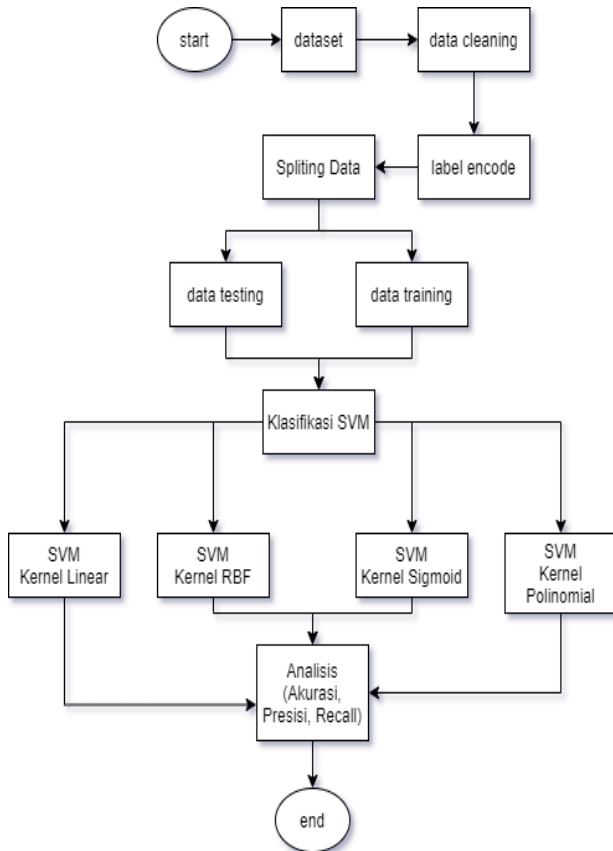
Splitting data yaitu proses membagi kumpulan data menjadi testing data dan *training* data. Dalam pembuatan model untuk klasifikasi algoritma yaitu dengan data *training*. Sedangkan untuk melakukan pengujian terhadap algoritma yang digunakan yaitu data *testing*. Kumpulan dari data percobaan (*train*) digunakan untuk membandingkan kinerja dari suatu model atau untuk memperkirakan parameter yang berbeda. Data *test* atau data uji digunakan setelah proses pelatihan atau *train* selesai. Pada proses *splitting data* menggunakan 14 perbandingan rasio. Angka perbandingan pada *splitting data* ditunjukkan pada tabel 2.

Tabel 2 Perbandingan *Splitting Data*

Training	Testing
95	5
80	20
70	30
60	40
50	50
40	60
30	70
20	80
10	90
90	10
85	15
75	25
65	35
55	45

E. Modeling

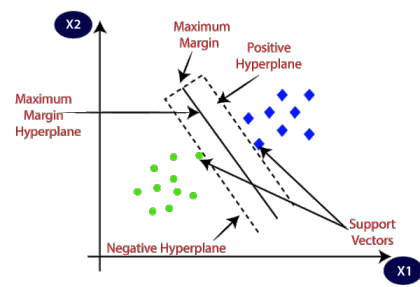
Dalam penelitian, algoritma SVM (*Support Vector Machine*) digunakan dalam proses klasifikasi, klasifikasi tersebut dengan menerapkan dari kernel linear, polinomial, RBF dan sigmoid. Dalam klasifikasi tersebut menentukan nilai akurasi, presisi dan recall. Pada Gbr 2 merupakan alur dalam proses klasifikasi.



Gbr 2 Alur Klasifikasi SVM (Support Vector Machine)

F. Support Vector Machine

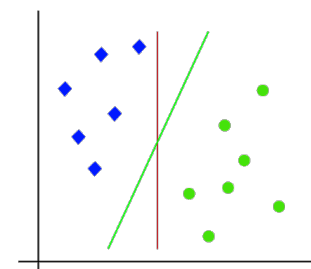
Metode *Support Vector Machine* (SVM) didasarkan pada prinsip minimisasi risiko structural (SRM). Tujuan SVM adalah untuk menemukan *hyperplane* yang baik pada ruang input untuk memisahkan dua kelas. *Hyperplane* ditemukan untuk mengukur margin dari kedua kelas dengan mencari titik maksimalnya. Menurut Burges proses klasifikasi dibedakan menjadi dua jenis berdasarkan kernel yang digunakan, yaitu SVM-Linear dan SVM Non-Linear. Gbr 3 adalah gambar dari SVM (*Support Vector Machine*).



Gbr 3 Support Vector Machine

1. SVM Linear

Pada proses klasifikasi menggunakan SVM *Linear*, data dapat dipisahkan dengan *hyperplane* menggunakan SVM (*Support Vector Machine*) dengan pembagian data yang jelas. Fungsi kernel *linear* : $K(x, x_i) = x \cdot x_i$. Gbr 4 merupakan gambar dari kernel *linear*.



Gbr 4 SVM Kernel Linear

2. SVM Non-Linear

SVM Non-Linear menggunakan fungsi kernel untuk memisahkan data non-linear menjadi data linear. Fungsi kernel, juga dikenal sebagai trik kernel. Trik kernel digunakan untuk mengelompokkan data dari dimensi kecil menjadi dimensi besar. Ada beberapa fungsi kernel non-linear yang dapat digunakan dalam proses klasifikasi, antara lain :

a. Kernel polynomial

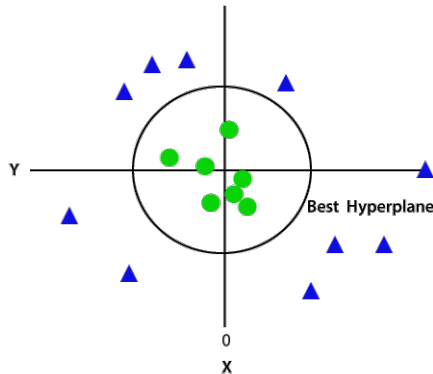
Kernel *polynomial* digunakan untuk merubah suatu data masukan ke dalam suatu dimensi yang memiliki ruang lebih tinggi. Kernel polinomial berfungsi untuk menemukan *hyperplane* yang baik untuk memisahkan data ruang yang telah diubah.

b. Kernel Gaussian RBF

Kernel RBF adalah kernel yang digunakan untuk mengklasifikasi data yang tidak dapat dipisah secara *linear*. Ruang fitur RBF memiliki jumlah dimensi yang tak terbatas yang ditentukan oleh parameternya. Oleh karena itu, ketika diulangi, dapat menghasilkan solusi *linear* yang unik yang menjadikannya sebagai proses klasifikasi terbaik.

c. Kernel Sigmoid

Kernel *sigmoid* merupakan pengembangan dari JST (Jaringan Syaraf Tiruan). Kernel ini diusulkan secara teoritis untuk *Support Vector Machine* karena berasal dari jaringan syaraf. *Kernel sigmoid* umumnya bermasalah karena sulit untuk mendapatkan parameter dengan nilai positif.



Gbr 5 SVM Non-Linear

G. Evaluasi

Pada tahap evaluasi ada akurasi, presisi dan recall. Recall dan presisi digunakan untuk mengevaluasi kinerja sistem. Kecocokan antara elemen data yang diambil dengan informasi yang diperlukan dikenal sebagai presisi.

$$\text{Rumus Presisi} = \frac{TP}{TP + FP} \quad (1)$$

Tingkat keberhasilan sistem dalam menemukan kembali data disebut recall.

$$\text{Rumus Recall} = \frac{TP}{TP + FN} \quad (2)$$

Akurasi didefinisikan sebagai seberapa jauh nilai yang didapatkan dan nilai yang sebenarnya berbeda.

$$\text{Rumus Akurasi} = \frac{TP + FN}{P + N} \quad (3)$$

III. HASIL DAN PEMBAHASAN

Hasil dan pembahasan ada dua poin yang akan dibahas, untuk yang pertama terkait implementasi dan yang kedua terkait hasil uji coba. Pada tahap ini merupakan hasil implementasi dari model yang telah ditentukan dalam proses klasifikasi.

A. Implementasi

Pada tahap ini membahas terkait tahapan dari implementasi, seperti pengumpulan data, data *cleaning*, *label encode*, *splitting data*, modelling dan evaluasi.

1. Pengumpulan Data

Dalam proses pengumpulan data, jenis data yang digunakan bersifat public. Dataset tersebut diambil dari website www.kaggle.com. Dataset ini terdiri diinput terlebih dahulu *Jupiter notebook* dengan menggunakan library *pandas*.

2. Data *cleaning*

Pada tahap ini data dicek terlebih dahulu ada yang missing atau tidak untuk datanya. Pada data yang digunakan terdapat kolom BMI yang memiliki data kosong dengan jumlah 201. Data yang kosong tersebut kemudian diisi dengan nilai mean atau median pad kolom BMI.

3. *Label encode*

Pada penelitian ini ada beberapa kolom tabel yang harus dirubah ke dalam numerik, kolom tabel tersebut diantaranya yaitu kolom *gender*, kolom *work_type*, kolom *ever_married*, kolom *Residence_type* dan kolom *smoking_status*.

4. *Splitting data*

Pada *splitting data* menggunakan nilai X dari kolom *gender*, *age*, *hypertension*, *avg_glucose_level* dan *bmi*. Sedangkan untuk y yaitu kolom *stroke*. Pada tahap ini menggunakan perbandingan rasio 80:20 dengan jumlah data train 4088 dan data test 1022.

5. *Modelling*

Dalam tahap modeling, menerapkan empat kernel dari metode *Support Vector Machine* dalam proses klasifikasinya. Empat kernel tersebut diantaranya kernel linear, polynomial, RBF dan sigmoid.

6. Evaluasi

Pada tahap evaluasi yaitu membandingkan nilai akurasi, presisi, dan recall dari keempat kernel yang telah ditentukan dalam modelling. Pada evaluasi ini, pada kernel *linear* menghasilkan nilai akurasi 75.73%, nilai presisi 74.07% dan nilai recall 56.16%. Pada gbr 6 merupakan hasil klasifikasi dengan kernel *linear*.

Kernel: linear				
SVM Accuracy Score -> 75.73385518590999				
SVM Precision Score -> 74.07598714416896				
SVM Recall Score -> 56.16911764705882				
	precision	recall	f1-score	support
0	0.9800	0.7593	0.8556	968
1	0.1434	0.7222	0.2393	54
accuracy			0.7573	1022
macro avg	0.5617	0.7408	0.5475	1022
weighted avg	0.9358	0.7573	0.8231	1022

Gbr 6 Hasil klasifikasi kernel *linear*

Hasil klasifikasi menggunakan kernel *polynomial* ditunjukkan pada gbr 7. Pada gbr 7 kernel *polynomial* menghasilkan nilai akurasi 78.86%, nilai presisi 73.98% dan nilai recall 56.75%.

```

Kernel: poly
SVM Accuracy Score -> 78.86497064579257
SVM Precision Score -> 73.98033363942456
SVM Recall Score -> 56.75755811446068
precision    recall    f1-score   support
0           0.9784    0.7944    0.8769     968
1           0.1568    0.6852    0.2552     54

accuracy
macro avg   0.5676    0.7398    0.5660    1022
weighted avg 0.9350    0.7886    0.8440    1022
    
```

Gbr 7 Hasil klasifikasi kernel *polynomial*

Hasil klasifikasi dari kernel RBF ditunjukkan pada gbr 8. Pada klasifikasi dengan kernel RBF menghasilkan nilai akurasi 71.81%, nilai presisi 75.5% dan nilai recall 55.93%.

```

Kernel: rbf
SVM Accuracy Score -> 71.81996086105676
SVM Precision Score -> 75.50696357514539
SVM Recall Score -> 55.93527421652421
precision    recall    f1-score   support
0           0.9843    0.7138    0.8275     968
1           0.1344    0.7963    0.2299     54

accuracy
macro avg   0.5594    0.7551    0.5287    1022
weighted avg 0.9394    0.7182    0.7960    1022
    
```

Gbr 8 Hasil klasifikasi kernel RBF

Pada gbr 9 merupakan hasil klasifikasi kernel sigmoid. Pada kernel sigmoid menghasilkan nilai akurasi 36.69%, nilai presisi 35.98% dan nilai recall 47.02%.

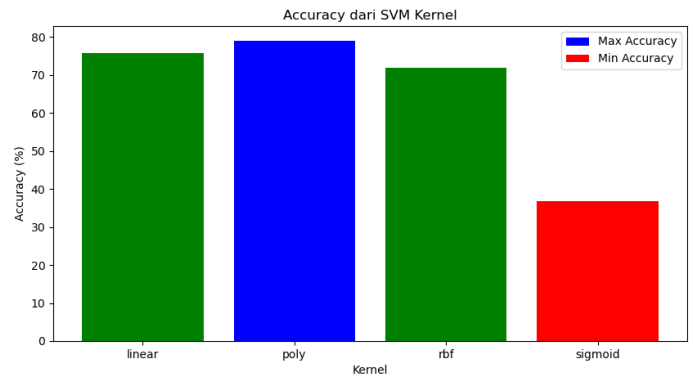
```

Kernel: sigmoid
SVM Accuracy Score -> 36.69275929549902
SVM Precision Score -> 35.98102234465871
SVM Recall Score -> 47.02984342638041
precision    recall    f1-score   support
0           0.9105    0.3678    0.5239     968
1           0.0301    0.3519    0.0555     54

accuracy
macro avg   0.4703    0.3598    0.2897    1022
weighted avg 0.8640    0.3669    0.4992    1022
    
```

Gbr 9 Hasil klasifikasi kernel sigmoid

Pada gbr 10 merupakan grafik dari linear, RBF, polynomial dan sigmoid. Pada grafik terlihat bahwa kernel polynomial menghasilkan nilai akurasi tertinggi dan kernel sigmoid menghasilkan nilai akurasi terendah dari ketiga kernel lainnya.



Gbr 10 Grafik akurasi 80:20

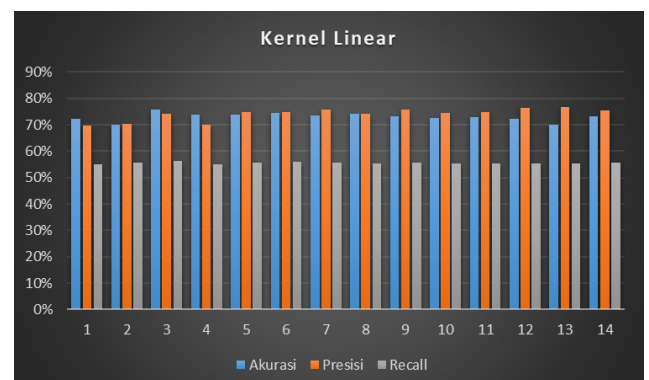
B. Hasil Uji Coba

Pada hasil uji coba membahas terkait hasil klasifikasi dari empat kernel yang telah ditentukan dengan perbandingan 14 rasio. Pada tabel 3 merupakan hasil perbandingan dengan menggunakan kernel *linear*. Pada kernel *linear* nilai akurasi tertinggi pada perbandingan 80:20 dengan nilai akurasi 75.73%, nilai presisi 74.07% dan nilai recall 56.16%.

Tabel 3 Hasil kernel *linear*

Rasio	Akurasi	Presisi	Recall
90 : 10	72.21 %	69.61%	55.18%
95 : 5	69.92%	70.23%	55.68%
80 : 20	75.73%	74.07%	56.16%
85 : 15	73.92%	70.11%	55.07%
70 : 30	73.77%	74.98%	55.73%
75 : 25	74.49%	74.91%	55.95%
60 : 40	73.67%	75.8%	55.68%
65 : 35	74.06%	74.21%	55.23%
55 : 45	73.08%	75.84%	55.59%
50 : 50	72.48%	74.65%	55.32%
40 : 60	72.96%	74.74%	55.38%
30 : 70	72.23%	76.27%	55.47%
20 : 80	70.15%	76.70%	55.42%
10 : 90	73.27%	75.48%	55.63%

Pada gbr 11 adalah grafik dari kernel linear. Pada grafik tersebut menggunakan 14 perbandingan rasio.



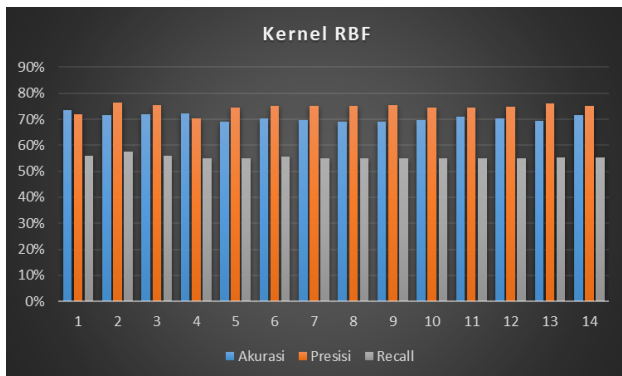
Gbr 11 Grafik kernel linear

Pada kernel RBF nilai akurasi tertinggi diperoleh dari perbandingan rasio 90:10, dengan nilai akurasi 73.38%, nilai presisi 71.79% dan nilai recall 55.85%. Hasil tabel perbandingan pada klasifikasi kernel RBF ditunjukkan pada tabel 4.

Tabel 4 Hasil klasifikasi kernel RBF

Rasio	Akurasi	Presisi	Recall
90 : 10	73.38%	71.79%	55.85%
95 : 5	71.48%	76.53%	57.46%
80 : 20	71.81%	75.5%	55.93%
85 : 15	72.35%	70.44%	54.96%
70 : 30	69.08%	74.38%	55.12%
75 : 25	70.42%	74.95%	55.48%
60 : 40	69.56%	75.21%	55.12%
65 : 35	69.14%	75.12%	54.91%
55 : 45	69.17%	75.53%	55.14%
50 : 50	69.66%	74.34%	54.99%
40 : 60	70.97%	74.34%	55.10%
30 : 70	70.31%	74.69%	54.98%
20 : 80	69.34%	76.03%	55.23%
10 : 90	71.62%	75.04%	55.36%

Pada gbr 12 merupakan grafik dari hasil percobaan kernel RBF. Pada grafik tersebut menggunakan 14 perbandingan dengan evaluasi menggunakan nilai, akurasi, nilai presisi dan nilai recall.



Gbr 12 Grafik kernel RBF

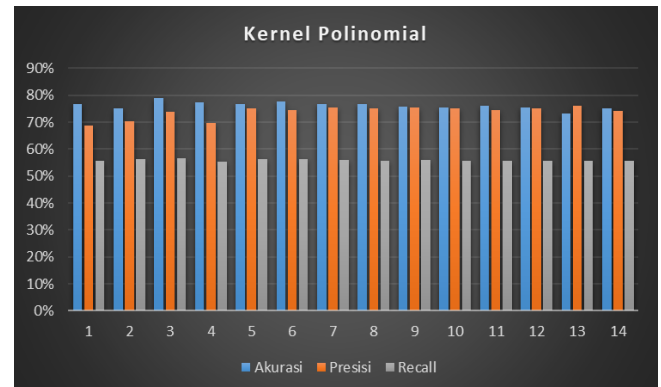
Hasil klasifikasi pada tabel 5 menunjukkan bahwa kernel *polynomial* menghasilkan nilai akurasi tertinggi pada perbandingan 80:20 dengan nilai akurasi 78.86%, nilai presisi 73.98% dan nilai recall 56.75%.

Tabel 5 Hasil klasifikasi kernel polinomial

No	Rasio	Akurasi	Presisi	Recall
1.	90 : 10	76.71%	68.87%	55.64%
2.	95 : 5	75%	70.21%	56.36%
3.	80 : 20	78.86%	73.98%	56.75%
4.	85 : 15	77.31%	69.60%	55.46%
5.	70 : 30	76.58%	75.20%	56.20%
6.	75 : 25	77.77%	74.46%	56.41%
7.	60 : 40	76.71%	75.53%	56.06%
8.	65 : 35	76.85%	75.10%	55.81%
9.	55 : 45	75.65%	75.45%	55.86%

10.	50 : 50	75.49%	75.06%	55.80%
11.	40 : 60	76.02%	74.40%	55.73%
12.	30 : 70	75.45%	75.09%	55.65%
13.	20 : 80	73.11%	76.04%	55.61%
14.	10 : 90	74.99%	74.03%	55.57%

Pada gbr 13 merupakan grafik dari kernel *polynomial*. Grafik tersebut adalah hasil uji coba dari 14 perbandingan rasio.



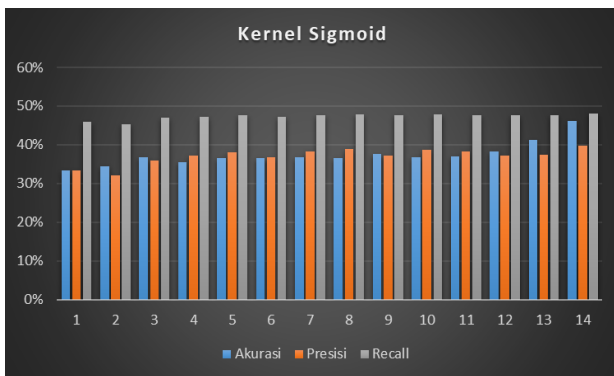
Gbr 13 Grafik kernel polinomial

Pada tabel 6 merupakan hasil perbandingan dari kernel sigmoid dengan menggunakan 14 perbandingan rasio. Pada kernel sigmoid menghasilkan nilai akurasi paling rendah dibandingkan ketiga kernel lainnya. Pada kernel sigmoid nilai akurasi terbaik dihasilkan pada perbandingan rasio 30:70, namun hasil tersebut masih dibawah 50%.

Tabel 6 Hasil klasifikasi kernel sigmoid

No	Rasio	Akurasi	Presisi	Recall
1.	90 : 10	33.46%	33.40%	45.99%
2.	95 : 5	34.37%	32.06%	45.23%
3.	80 : 20	36.69%	35.98%	47.02%
4.	85 : 15	35.59%	37.21%	47.21%
5.	70 : 30	36.66%	37.99%	47.59%
6.	75 : 25	36.54%	36.72%	47.27%
7.	60 : 40	36.69%	38.36%	47.75%
8.	65 : 35	36.61%	38.92%	47.93%
9.	55 : 45	37.73%	37.24%	47.57%
10.	50 : 50	36.75%	38.76%	47.82%
11.	40 : 60	37.08%	38.33%	47.75%
12.	30 : 70	38.32%	37.28%	47.64%
13.	20 : 80	41.36%	37.46%	47.69%
14.	10 : 90	46.14%	39.84%	48.13%

Pada gbr 14 merupakan hasil klasifikasi menggunakan kernel sigmoid. Pada kernel sigmoid menggunakan 14 perbandingan rasio dalam proses klasifikasinya.



Gbr 14 Klasifikasi kernel sigmoid

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

1. Penerapan metode *Support Vector Machine* (SVM) dalam proses klasifikasi penyakit *stroke* dengan menggunakan empat kernel dalam proses klasifikasinya. Kernel tersebut yaitu linear, RBF, polynomial dan sigmoid.
2. Dalam penelitian ini menggunakan empat kernel dalam proses evaluasi. Evaluasi tersebut yaitu menentukan dari beberapa nilai akurasi, presisi, dan recall pada proses klasifikasi. Pada klasifikasi dengan kernel *linear* menghasilkan nilai tinggi pada perbandingan 80:20, dengan nilai akurasi 75.73%, nilai presisi 74.07% dan nilai recall 56.16%. Pada kernel RBF menghasilkan nilai akurasi tinggi pada perbandingan 90:10 dengan nilai akurasi 73.38%, nilai presisi 71.79% dan nilai recall 55.85%. Pada kernel *polynomial* menghasilkan nilai akurasi tinggi pada perbandingan 80:20 dengan nilai akurasi 78.86%, nilai presisi 73.98% dan nilai recall 56.75%. Kemudian pada kernel sigmoid menghasilkan nilai akurasi tinggi pada perbandingan 10:90, namun nilai akurasi yang dihasilkan masih dibawah 50%. Dari pengujian keempat kernel dengan menggunakan 14 rasio perbandingan, maka nilai akurasi tertinggi dihasilkan oleh kernel *polynomial* dengan nilai akurasi 78.86%, nilai presisi 73.98% dan nilai recall 56.75% pada perbandingan rasio 80:20.

B. Saran

Saran pada penelitian selanjutnya yaitu melakukan penambahan dengan beberapa fitur tertentu yang bisa menghasilkan nilai akurasi lebih tinggi. Selain itu, perbandingan algoritma klasifikasi yang berbeda untuk menghasilkan hasil yang lebih variatif juga dapat dikembangkan, dan model algoritma mana yang akan memberikan nilai yang paling baik dalam proses klasifikasi penyakit stroke dapat ditemukan.

UCAPAN TERIMAKASIH

Terimakasih penulis ucapkan kepada Allah SWT, karena berkat rahmat dan ridho-Nya sehingga penulis dapat menyelesaikan penelitian ini. Tidak lupa juga penulis ucapkan kepada kedua orang tua yang selalu mendukung maupun mendoakan penulis, sehingga penulis bisa bertahan sampai tahap ini. Penulis ucapkan terimakasih juga kepada pembimbing yang telah sabar dan selalu memberikan bimbingan yang terbaiknya. Terimakasih juga kepada teman-teman yang telah bersedia membantu dan menemani penulis dalam penyelesaian penelitian ini.

REFERENSI

- [1] Puspitasari, P. N. (2020). Hubungan Hipertensi Terhadap Kejadian Stroke. *Jurnal Ilmiah Kesehatan Sandi Husada*, 12(2), 922–926.
- [2] Paramitha, Y. N., Nuryaman, A., Faisol, A., Setiawan, E., & Nurvazly, D. E. (2023). Klasifikasi Penyakit Stroke Menggunakan Metode Naïve Bayes. *Jurnal Siger Matematika*, 04(01), 11–16.
- [3] Oktarina, Y., Nurhusna, Kamariyah, & Mulyani, S. (2020). Edukasi Kesehatan Penyakit Stroke Pada Lansia. *Medic*, 3(2), 106–109.
- [4] Arifin, S. S., Siregar, A. M., Ratna, A., & Mudzakir, T. Al. (2021). Klasifikasi Penyakit Kanker Serviks Menggunakan Algoritma Support Vector Machine (SVM). *The 4th Conference on Innovation and Application of Science and Technology (CLASTECH 2021)*, 4, 521–528.
- [5] Fitriyah, N., Warsito, B., & Maruddani, D. A. I. (2020). Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (Svm). *Jurnal Gaussian*, 9(3), 376–390. <https://doi.org/10.14710/j.gauss.v9i3.28932>
- [6] Fremmuzar, P., & Baita, A. (2023). Uji Kernel SVM dalam Analisis Sentimen Terhadap Layanan Telkomsel di Media Sosial Twitter SVM Kernel Test in Sentiment Analysis of Telkomsel Services on Twitter Social Media. 12(148). <https://doi.org/10.34010/komputika.v12i2.9460>
- [7] Frenica, A., Lindawati, L., Lindawati, L., Soim, S., & Soim, S. (2023). Implementasi Algoritma Support Vector Machine (SVM) untuk Deteksi Banjir. *INOVTEK Polbeng - Seri Informatika*, 8(2), 291. <https://doi.org/10.35314/isi.v8i2.3443>
- [8] Khafidhun, M., Muslim, A., & Yamasari, Y. (2021). Klasifikasi Opini Pengguna Twitter Terhadap Sekolah Daring dengan Metode Naive Bayes dan Support Vector Machine. 03, 171–179.
- [9] Muallfah, D., Fadila, W., & Firdaus, R. (2022). Teknik SMOTE untuk Mengatasi Imbalance Data pada Deteksi Penyakit Stroke Menggunakan Algoritma Random Forest. *Jurnal CoSciTech (Computer Science and Information Technology)*, 3(2), 107–113. <https://doi.org/10.37859/coscitech.v3i2.3912> “PDCA12-70 data sheet,” Opto Speed SA, Mezzovico, Switzerland.
- [10] Mucholladin, A. W., Bachtiar, F. A., & Furqon, M. T. (2021). Klasifikasi Penyakit Diabetes menggunakan Metode Support Vector Machine. 5(2).
- [11] Ramadhan, N. G., & Khoirunnisa, A. (2021). Klasifikasi Data Malaria Menggunakan Metode Support Vector Machine. *Jurnal Media Informatika Budidarma*, 5(4), 1580.
- [12] Firdaus, R., Muallfah, D., & Hasanah, J. S. (2023). Klasifikasi Multi-Class Penyakit Jantung Dengan SMOTE dan Pearson’s Correlation menggunakan MLP. *Jurnal CoSciTech (Computer Science and Information Technology)*, 4(1), 262–271.
- [13] Prasetyo, R. T., & Ripandi, E. (2019). Optimasi Klasifikasi Jenis Hutan Menggunakan Deep Learning Berbasis Optimize Selection. *Jurnal Informatika*, 6(1), 100–106. <https://doi.org/10.31311/ji.v6i1.5176>
- [14] Raysyah, S. R., Veri Arinal, & Dadang Iskandar Mulyana. (2021). Klasifikasi Tingkat Kematangan Buah Kopi Berdasarkan Deteksi Warna Menggunakan Metode Knn Dan Pca. *JSiI (Jurnal Sistem Informasi)*, 8(2), 88–95.
- [15] Wiguna, I. K. A. G., & Asana, I. M. D. P. (2021). Implementasi Zoning dan Fitur Arah Sebagai Ekstraksi Fitur. *Jurnal Rekayasa Sistem Informasi*, 4(1), 85–92.