

## Klasifikasi Penyakit Cacar Monyet Menggunakan Metode Support Vector Machine

Wendy Anugrah, Elin Haerani\*, Yusra, Lola Oktavia

Fakultas Sains dan Teknologi, Program Studi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim, Pekanbaru, Indonesia

Email: <sup>1</sup>11751102108 @students.uin-suska.ac.id, <sup>2\*</sup>elin.haerani@uin-suska.ac.id, <sup>3</sup>yusra@uin-suska.ac.id, <sup>4</sup>lola.oktavia@uin-suska.ac.id

Email Penulis Korespondensi: elin.haerani@uin-suska.ac.id

Submitted: 07/05/2024; Accepted: 18/05/2024; Published: 30/05/2024

**Abstrak**—Cacar monyet merupakan suatu penyakit *zoonosis* yang disebabkan oleh virus cacar monyet dan penyakit ini sangat berbahaya. Penyakit cacar monyet dapat dideteksi sebelumnya dengan menggunakan informasi yang ada pada data pasien dan menerapkan teknik machine learning. Pada penelitian ini bertujuan untuk mengklasifikasi penyakit cacar monyet menggunakan metode *Support Vector Machine* (SVM). Pengujian ini dilakukan dengan menggunakan *confusion matrix* dengan membandingkan rasio data latih dan data uji dengan perbandingan 70:30, 80:20, 90:10 dan menggunakan kernel RBF. Berdasarkan hasil pengujian didapatkan hasil rasio tertinggi pada 90:10 dengan nilai akurasi terbaik sebesar 65% dengan pengujian parameter SVM yaitu nilai  $C=10$  dan  $\gamma=1$ . Berdasarkan hasil pengujian yang telah dilakukan menggunakan metode *Support Vector Machine* mendapatkan nilai akurasi yang cukup baik.

**Kata Kunci:** Klasifikasi; Cacar Monyet; Support Vector Machine; Kernel Rbf; Confusion Matrix

**Abstract**—Monkey pox is a zoonotic disease caused by the monkey pox virus and this disease is very dangerous. Monkey pox can be detected in advance by using information contained in patient data and applying machine learning techniques. This study aims to classify monkey pox using the Support Vector Machine (SVM) method. This test is carried out using a confusion matrix by comparing the ratio of training data and test data with a ratio of 70:30, 80:20, 90:10 and using the RBF kernel. Based on the test results, the highest ratio results were obtained at 90:10 with the best accuracy value of 65% with SVM parameter testing, namely the value  $C=10$  and  $\gamma=1$ . Based on the results of tests carried out using the Support Vector Machine method, the accuracy values were quite good.

**Keywords:** Classification; Monkey Pox; Support Vector Machine; RBF Kernel; Confusion Matrix

### 1. PENDAHULUAN

Setelah kejadian pandemi Covid-19 pada tahun 2020 yang menggemparkan dan sekarang kita dihadapkan dengan penyakit yang juga berbahaya yaitu penyakit cacar monyet. Cacar Monyet merupakan suatu penyakit *zoonosis* yang disebabkan oleh virus cacar monyet, orthopoxvirus dan virus variola [1]. Cacar memiliki beberapa jenis namun pada penelitian ini berfokus kepada cacar monyet. Cacar monyet pertama kali dideteksi pada tahun 1958 saat dilakukan isolasi pada lesi *vesikuloid pustular* di antara monyet tawanan di Kopenhagen [2]. Gejala awal yang ditimbulkan pada penyakit cacar monyet yaitu demam. Demam sering disertai dengan sakit kepala berat, berkeringat, menggigil, sakit punggung, sesak nafas dan batuk. Setelah itu dalam waktu beberapa hari terjadinya demam akan mengeluarkan ruam. Ruam sering mulai dari wajah dan kemudian menjalar keseluruhan tubuh [3]. Timbulnya lesi pada kulit mengakibatkan nyeri pada lokasi lesinya, seperti lesi pada rongga mulut dan bagian-bagian lainnya pada kulit manusia. Adanya edema pada bagian tubuh akibat pembengkakan pada bagian tubuh tertentu karena adanya penumpukan cairan berlebih. Melakukan kontak seksual juga dapat mengakibatkan penyebaran infeksi [4].

Penyakit cacar monyet sebagian besar terjadi di Afrika bagian tengah dan barat pada hewan pengerat liar. Walaupun belum menjadi pandemic secara global, namun sejak Januari 2022, 3413 laboratorium dari 50 negara yang mengkonfirmasi adanya penyakit cacar monyet ini dan ada yang meninggal, sehingga penyakit cacar monyet ini menjadi perhatian khusus [5]. Hal yang mendasari meningkatnya kasus cacar monyet tidak lepas dari faktor kesenjangan sosial ekonomi yang melanda di negara bagian endemik. Hingga saat ini masih dilakukan investigasi lebih lanjut untuk meminimalisir penyebaran penyakit cacar monyet, salah satunya dengan cara mendeteksi secara dini.

*Support Vector Machine* merupakan metode klasifikasi *supervised learning* dengan konsep dasar yang sebenarnya merupakan kombinasi yang baik dalam teori-teori komputasi yang ada puluhan tahun sebelumnya, seperti margin *hyperplane* yang diperkenalkan oleh Aronszajn pada tahun 1950 dan begitu juga dengan konsep-konsep pendukung yang lainnya. *Support Vector Machine* membutuhkan training set positif dan negatif. Training set positif dan negatif ini diperlukan *Support Vector Machine* untuk membuat keputusan terbaik dalam memisahkan data positif dengan data negatif di ruang n-dimensi, yang disebut dengan *hyperplane*. Secara sederhana konsep *Support Vector Machine* dapat diartikan sebagai proses mencari garis pemisah *hyperplane* dengan mengoptimalkan *hyperplane*, dan memaksimalkan margin antara dua kelas [6].

Metode *Support Vector Machine* merupakan metode perhitungan yang baik di dalam mendapatkan hasil klasifikasi dengan tingkat akurasi yang tinggi. Metode *Support Vector Machine* ini merupakan metode yang banyak digunakan untuk klasifikasi penyakit kulit [7]. Proses klasifikasi memiliki dua langkah pengerjaan, yaitu

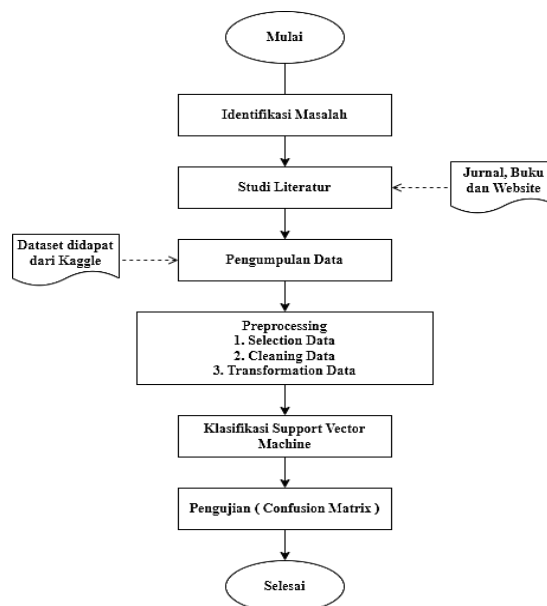
proses pengujian dan proses pelatihan. Proses pelatihan digunakan untuk membuat model suatu pengujian set [8].

Penggunaan metode *Support Vector Machine* sudah banyak diterapkan dalam permasalahan klasifikasi seperti penelitian sebelumnya yaitu tentang klasifikasi *pneumonia* menggunakan metode *Support Vector Machine* dan menggunakan dataset sebanyak 3140. Ada beberapa tahapan yang dilakukan yaitu *cropping*, *resizing*, *contrast stretching* dan *thresholding*. Akurasi terbaik yang diperoleh sebesar 62.66% [9]. Kemudian penelitian klasifikasi penyakit kanker serviks menggunakan algoritma *Support Vector Machine* (SVM). Pengujian pada penelitian ini menggunakan *python* dan menggunakan tools *orange*. Hasil yang diperoleh dari menggunakan tools *orange* dan membagi data 80:20 mendapatkan nilai akurasi sebesar 92,9% dan *python* sebesar 87% [10]. Selanjutnya penelitian tentang klasifikasi data malaria menggunakan metode *Support Vector Machine*. Penelitian ini juga membandingkan metode *Support Vector Machine* dengan *Naïve Bayes* dimana nilai akurasi SVM lebih unggul dibandingkan dengan *Naïve Bayes* dengan gap akurasi rata-rata 25%. Hasil akurasi yang diperoleh dalam penelitian ini sebesar 92.3% [11]. Selanjutnya penelitian tentang prediksi penyakit diabetes melitus menggunakan metode *Support Vector Machine* dan *Naïve Bayes*. Hasil penelitian yang diperoleh dari metode SVM memiliki nilai akurasi lebih baik dibandingkan dengan metode *Naïve Bayes*. Nilai akurasi yang diperoleh dari metode *Support Vector Machine* adalah 78,04% dan nilai akurasi dari metode *Naïve Bayes* adalah 76,98 [12].

Penelitian ini dilakukan atas dasar kekhawatiran karena adanya dampak negatif dari penyebaran penyakit cacar monyet yang terus meningkat. Sebagian pasien yang terjangkit penyakit cacar monyet tidak menyadari bahwa penyakit ini memiliki resiko yang sangat tinggi dan dapat mengakibatkan kematian. Dengan adanya resiko yang besar akan terjadi maka tujuan dari penelitian ini yaitu membuat model *machine learning* yang dapat mendeteksi penyakit cacar monyet. Dalam penelitian ini, tujuan utamanya untuk mengklasifikasi penyakit cacar monyet menggunakan metode *Support Vector Machine*. Penelitian ini menggunakan *python* dan *google colab* agar memudahkan pengembangan dan implementasi model *machine learning*. Dataset yang digunakan dalam penelitian ini diperoleh dari web *Kaggle*. Penelitian ini dapat memberikan kontribusi dalam mengklasifikasikan penyakit cacar monyet terhadap pasien dengan metode *Support Vector Machine*.

## 2. METODOLOGI PENELITIAN

Tahapan yang dilakukan dalam penelitian ini dimulai dari melakukan identifikasi masalah, selanjutnya mencari referensi mengenai penelitian dengan melakukan studi literatur, kemudian melakukan pengumpulan data dan analisa data, selanjutnya melakukan proses preprocessing dimana didalamnya terdapat *Selection data*, *Cleaning data*, dan *Transformation data*, kemudian proses klasifikasi menggunakan metode *Support Vector Machine*, selanjutnya dilakukan pengujian dengan menggunakan confusion matrix dan yang terakhir memberikan kesimpulan. Dalam gambar 1 akan menjelaskan alur dari proses penelitian.



Gambar 1. Alur Penelitian

### 2.1 Identifikasi Masalah

Pada tahap identifikasi masalah ini dilakukan untuk mengetahui latar belakang masalah, rumusan masalah dan tujuan dari penelitian yang akan dibuat setelah semua itu didapatkan maka peneliti akan mudah untuk mendapatkan gambaran penelitian yang akan dilakukan.

## 2.2 Studi Literatur

Pada tahap ini dilakukan pencarian jurnal yang terkait dengan klasifikasi penyakit menggunakan metode Support Vector Machine yang mana nantinya akan dijadikan referensi dalam penelitian ini. Referensi yang digunakan bisa didapatkan dari jurnal, buku dan website.

## 2.3 Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data dan data yang telah dikumpulkan sesuai dengan kebutuhan penelitian lalu data tersebut akan diolah dan digunakan sebagai input dari *machine learning*.

## 2.4 Preprocessing

Pada tahap ini merupakan tahapan awal untuk mengubah data mentah yang dikumpulkan dari berbagai sumber menjadi bersih dan informasi yang lebih mudah untuk diakses yang akan digunakan untuk pemrosesan lebih lanjut. Preprocessing juga dapat dilakukan untuk mengeliminasi data yang tidak sesuai dan data akan diubah kedalam bentuk yang mudah dipahami oleh sistem. Ada beberapa tahapan yang dilakukan dalam preprocessing yaitu :

a. *Selection Data*

Pada proses *selection data* ini digunakan untuk menyeleksi data yang tidak digunakan dalam penelitian. Proses ini dilakukan agar dapat meminimalkan jumlah data yang digunakan dalam proses mining data.

b. *Cleaning Data*

Pada proses *cleaning data* ini digunakan untuk mengecek data yang kosong atau tidak relevan agar dapat memudahkan proses mining data.

c. *Transformation Data*

Pada proses *transformation data* ini digunakan untuk mentransformasikan data kedalam bentuk yang sesuai untuk proses mining.

## 2.5 Klasifikasi Support Vector Machine

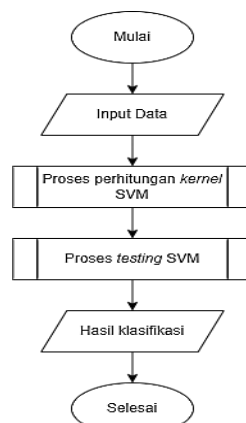
*Support Vector Machine* (SVM) merupakan teknik pembelajaran mesin yang digunakan untuk pengenalan pola. Algoritma SVM pertama kali ditemukan oleh Vladimir Vapnik pada tahun 1992 dan algoritma SVM termasuk dalam kategori pembelajaran terbimbing (*supervised learning*) berguna dalam masalah klasifikasi dan SVM melakukan klasifikasi data kedalam dua kelas yang berbeda dengan cara membentuk garis batas keputusan atau *hyperplane* [13].

Metode penelitian yang digunakan yaitu *Support Vector Machine* untuk klasifikasi atau yang disebut *Support Vector Classification* (SVC). Klasifikasi merupakan proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas yang tidak diketahui dari suatu objek. Dalam klasifikasi secara umum memiliki dua proses yaitu :

a. Proses training : proses ini digunakan untuk training set yang digunakan label-labelnya untuk membangun model atau fungsi.

b. Proses testing : proses ini digunakan untuk mengetahui keakuratan model atau fungsi yang akan dibangun pada proses training, maka digunakan data yang disebut dengan testing set untuk memprediksi label-labelnya.

Langkah pertama yang dilakukan dalam proses *Support Vector Machine* (SVM) yaitu menginputkan data cacar monyet untuk proses training dan testing. Langkah kedua yang dilakukan yaitu setelah data didapatkan selanjutnya menghitung *kernel* SVM, dimana *kernel* yang digunakan yaitu *kernel* RBF. Setelah nilai *kernel* didapatkan maka proses selanjutnya yaitu proses *testing* SVM dan langkah terakhir dari proses SVM yaitu evaluasi dari klasifikasinya. Dalam gambar 2 akan menjelaskan alur dari proses *Support Vector Machine*.



**Gambar 2.** Alur Proses Support Vector Machine

## 2.6 Pengujian

Pada proses pengujian ini dilakukan untuk mencari nilai akurasi terbaik dengan menggunakan metode *Support Vector Machine* dalam mengklasifikasikan data ke dalam kelas yang telah ditentukan. Untuk mendapatkan tingkat keberhasilan dari pengklasifikasian metode SVM, maka pengujian yang dilakukan menggunakan metode *Confusion Matrix*. Ada empat istilah dalam proses klasifikasi pada *Confusion Matrix* yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Pada tahapan ini yang proses yang dilakukan yaitu menghitung nilai *Precision*, *Recall*, dan *Accuracy*. Pada tabel 1 dan persamaan (1), (2), dan (3) menjelaskan cara perhitungan *Confusion Matrix*.

Tabel 1. Confusion Matrix

		Prediksi	
		Positif	Negatif
Aktual	Positif	TP	FN
	Negatif	FP	TN

$$Precision = \frac{TP}{TP+FP} \times 100\%$$

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

*True Positive* (TP) adalah Jumlah data positif penyakit cacar monyet yang terklasifikasi dengan benar oleh sistem. *True Negative* (TN) adalah Jumlah data negatif penyakit cacar monyet yang terklasifikasi dengan benar oleh sistem. *False Positive* (FP) adalah: Jumlah data positif penyakit cacar monyet akan tetapi terklasifikasi salah oleh sistem. *False Negative* (FN) adalah Jumlah data negatif penyakit cacar monyet akan tetapi terklasifikasi salah oleh sistem.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Persiapan Data

Mempersiapkan data merupakan awal dari sebuah penelitian yang mana data tersebut nantinya akan diolah dan digunakan sebagai *input* dari *machine learning*. Data penyakit cacar monyet (Monkeypox) yang digunakan dalam penelitian ini diperoleh dari web *Kaggle* dengan nama *Monkey-Pox PATIENTS Dataset*. Dalam *dataset* ini, sampel yang dikumpulkan melalui penelitian yang diterbitkan yaitu gambaran klinis dan presentasi baru cacar monyet pada manusia di pusat kota London selama wabah tahun 2022.

Ada 9 data gejala yang ada pada penelitian ini yang terdapat pada tabel 2 berikut:

Tabel 2. Gejala Monkeypox

Variabel	Keterangan
X1	Systemic Illness
X2	Rectal Pain
X3	Sore Throat
X4	Penile Oedema
X5	Oral Lesion
X6	Solitary Lesion
X7	Swollen Tonsils
X8	HIV Infection
X9	Sexually Transmitted Infection

Data yang digunakan pada penelitian ini berjumlah 5000 *dataset* dan data sudah berlabel. *Dataset* tersebut memiliki 2 kelas dengan jumlah masing-masing yaitu 2510 yang teridentifikasi positif dan 2490 yang teridentifikasi negatif. Data yang digunakan dapat dilihat pada tabel 3 berikut:

Tabel 3. Data Penelitian

	Patient_ID	X1	X2	X3	X4	X5	X6	X7	X8	X9	Monkeypox
0	P1	Fever	True	False	True	True	False	False	True	False	Positive
1	P2	Fever	False	True	True	False	False	False	True	False	Positive
2	P3	None	True	False	False	False	True	True	True	False	Positive
3	P4	SLN	True	True	True	False	False	True	true	False	Positive
4	P6	Fever	False	true	False	False	False	False	True	False	Positifive

	Patient_ID	X1	X2	X3	X4	X5	X6	X7	X8	X9	Monkeypox
...	...	...	...	...	...	...	...	...	...	...	...
4995	P6792	MAP	False	True	False	False	False	True	True	False	Negative
4996	P6797	SLN	True	True	False	True	True	True	False	True	Negative
4997	P6798	MAP	True	True	True	True	False	False	True	False	Negative
4998	P6801	SLN	False	False	False	False	False	True	False	True	Negative
4999	P6807	fever	False	True	True	False	True	True	True	False	Negative

Pada tabel 3 memperlihatkan data penelitian yang akan diproses. Pada tabel *Systemic Illness* (X1) yang mana didalam kolomnya terdapat SLN atau *Swollen Lymph Nodes* dan MAP atau *Muscle Aches and Pain*.

### 3.2 Preprocessing

Data Preprocessing merupakan tahapan untuk menghilangkan permasalahan yang mengganggu saat pemrosesan data. Untuk membuat pemrosesan data berjalan dengan baik, maka diperlukan tahapan Preprocessing yaitu :

#### 3.2.1 Selection Data

Pada tahap ini melakukan proses penyeleksian data yang tidak digunakan dalam penelitian. Dimana proses yang dilakukan yaitu untuk meminimalkan jumlah data yang digunakan untuk proses mining namun tetap menjelaskan data aslinya. Dari 11 atribut yang tidak digunakan ada 2 atribut yaitu *Patiend\_ID* (Id Pasien) karena tidak akan mempengaruhi apapun dalam penelitian ini dan *HIV Infection* (Infeksi HIV) karena tidak adanya bukti ilmiah yang menunjukkan keterkaitan antara infeksi HIV dan penularan penyakit cacar monyet. Setelah terjadinya *selection data* maka ada 8 gejala yang akan menjadi variabel input dalam proses pembuatan sistem dapat dilihat pada tabel 4 diantaranya yaitu :

**Tabel 4.** Variabel Data Input Gejala Monkeypox

Variabel	Keterangan
X1	Systemic Illness
X2	Rectal Pain
X3	Sore Throat
X4	Penile Oedema
X5	Oral Lesions
X6	Solitary Lesion
X7	Swollen Tensils
X8	Sexually Transmitted Infection

Jadi selanjutnya dalam penelitian ini akan menggunakan 9 atribut. Hasil dari *selection data* dapat dilihat pada tabel 5 berikut:

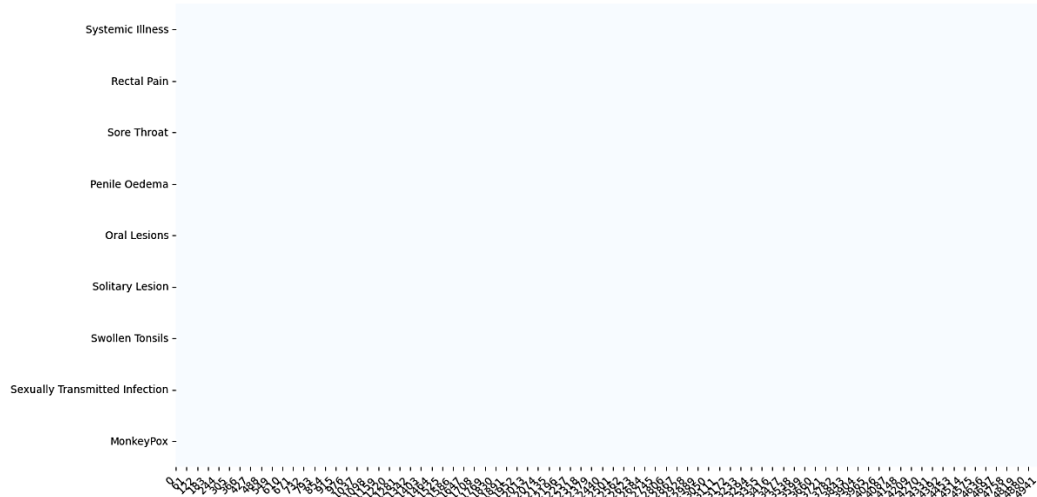
**Tabel 5.** Gejala Monkeypox

	X1	X2	X3	X4	X5	X6	X7	X8	MonkeyPox
0	Fever	True	False	True	True	False	False	False	Positive
1	Fever	False	True	True	False	False	False	False	Positive
2	None	True	False	False	False	True	True	False	Positive
3	SLN	True	True	True	False	False	True	False	Positive
4	Fever	False	True	False	False	False	False	False	Positive

Pada tabel 5 diatas memperlihatkan hasil dari *selection data*. Pada tabel *Systemic Illness* (X1) yang mana pada kolomnya terdapat SLN atau *Swollen Lymph Nodes*.

#### 3.2.2 Cleaning Data

Pada tahap ini melakukan pembersihan data yang kiranya data tersebut ada yang tidak lengkap atau tidak relevan maka akan di bersihkan. Dengan adanya proses ini maka data yang digunakan akan terhindar dari kesalahpahaman saat melakukan analisa terhadap data tersebut [14]. Dalam proses *cleaning data* untuk mengecek terjadinya *missing value*. Namun pada proses ini tidak adanya terjadi data yang harus di hapus atau di buang. Hasil dari *cleaning data* dapat dilihat pada gambar 3 berikut :



**Gambar 3.** Hasil Cleaning Data

Terlihat pada gambar 3 diatas memperlihatkan hasil dari *cleaning data*. Dimana jika grafiknya bersih dan tidak ada garis maka datanya bersih atau *cleaning*. Berarti data yang digunakan telah sesuai sehingga tidak ada data yang kosong dan di buang.

**3.2.3 Transformation Data**

Pada tahap *transformasi data* ini digunakan untuk menyamakan seluruh data. Karena dengan *transformation data* kita dapat merubah struktur data, format data maupun nilai data. Agar dataset yang digunakan sesuai dengan data mining atau algoritma yang dirancang [15]. Bentuk perubahan dari proses tranformasi data akan ditampilkan pada tabel 6 berikut :

**Tabel 6.** Hasil Transformation Data

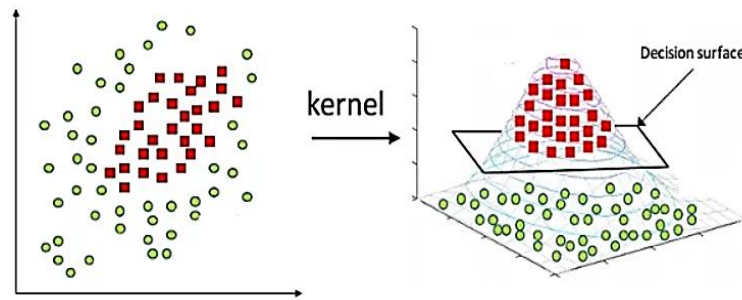
	X1	X2	X3	X4	X5	X6	X7	X8	MonkeyPox
0	0	1	0	1	1	0	0	0	1
1	0	0	1	1	0	0	0	0	1
2	2	1	0	0	0	1	1	0	1
3	3	1	1	1	0	0	1	0	1
4	0	0	1	0	0	0	0	0	1
...	...	...	...	...	...	...	...	...	...
4995	1	0	1	0	0	0	1	0	0
4996	3	1	1	0	1	1	1	1	0
4997	1	1	1	1	1	0	0	0	0
4998	3	0	0	0	0	0	1	1	0
4999	0	0	1	1	0	1	1	0	0

Pada tabel 6 memperlihatkan hasil perubahan dari *transformation data*. Setelah dilakukan *transformation data* maka proses selanjutnya dapat memudahkan dalam melakukan klasifikasi pada *machine learning*.

**3.3 Support Vector Machine**

Support vector machine merupakan metode klasifikasi yang dapat mengidentifikasi titik baru dengan akurasi yang tinggi [16]. SVM masuk ke dalam *supervised learning classification* dan *support vector regression*. Algoritma SVM mengklasifikasi data dengan mencari *hyperplane* terbaik yang memisahkan semua data dari satu kelas atau beberapa kelas [17]. SVM juga memiliki dasar dalam teori pembelajaran statistik dan dianggap dengan klasifikasi yang kuat [18]. Model algoritma SVM merupakan salah satu algoritma dari klasifikasi, yang bekerja untuk memisahkan dua kelompok data dalam mencari garis (*hyperplane*) terbaik. *Hyperplane* dengan pemisah terbaik dapat ditemukan dengan mengukur *margin* dari *hyperplane* dan mencari titik maksimum. *Margin* merupakan jarak antara *hyperplane* tersebut dengan data terdekat di antara dua kelas yang berbeda, data yang paling dekat disebut dengan *support vector* [19].

Berikut merupakan metode yang digunakan dalam mengklasifikasi data yang tidak dapat di klasifikasikan secara linear adalah dengan cara mentransformasikan data kedalam ruang fitur berdimensi tinggi sehingga dapat dipisah secara linear pada ruang fitur (*feature space*). *Feature space* dalam prosesnya biasanya memiliki dimensi yang lebih tinggi dari vector input (*input space*). Dalam gambar 4 menjelaskan proses perubahan dari yang tidak menggunakan *kernel* hingga menggunakan *kernel*.



Gambar 4. Data Non Linear

Pada gambar 4 [20], memperlihatkan data SVM yang tidak seimbang mengakibatkan proses klasifikasi menjadi tidak baik. Maka dilakukan proses menggunakan *kernel trick* sehingga proses klasifikasi dapat dilakukan dengan baik walaupun data dalam bentuk yang banyak.

Berikut ini adalah beberapa fungsi kernel yang umum digunakan yaitu :

- a. Kernel Linear  
 $K(x_i, x) = x_i^T x$
- b. Polynomial  
 $K(x_i, x) = (y \cdot x_i^T x + r)^p, y > 0$
- c. Radial Basis Function  
 $K(x_i, x) = \exp(-\gamma |x_i - x|^2), \gamma > 0$
- d. Sigmoid Kernel  
 $K(x_i, x) = \tanh(\gamma x_i^T x + r)$

Penggunaan fungsi *kernel* merupakan cara untuk mendapatkan hasil dari klasifikasi terbaik. Adapun *kernel* yang digunakan pada penelitian ini yaitu kernel RBF (*Radial Basic Function*). Pemilihan untuk menggunakan kernel RBF ini dilakukan atas dasar penelitian yang dilakukan oleh [21] pada penelitiannya yang memperlihatkan bahwa kernel RBF lebih baik dibandingkan dengan kernel polynomial dan kernel sigmoid. Kernel RBF dapat memproyeksikan data ke dalam ruang fitur yang memiliki dimensi tak hingga, memungkinkan SVM untuk menangani masalah yang tidak linear dengan baik [22].

### 3.4 Pengujian Data

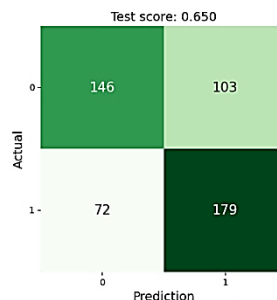
Dalam proses pengujian ini melakukan pengoptimalan parameter (parameter tuning) yang berguna untuk meningkatkan akurasi dari model *machine learning*. Dalam pengoptimalan *hyperparameter* pada model SVM memerlukan GridSearchCV. Formula yang digunakan untuk set parameter tuning tertera pada formula (1) berikut.

Kernel = {RBF}

C = {0.1, 1, 10}

Gamma = {0.1, 0.01, 1}

Dalam proses penerapan algoritma SVM langkah yang harus dilakukan yaitu membagi dataset menjadi 2 yaitu data training dan data testing. Data dibagi 2 dengan bobot nilai sebesar 70:30, 80:20 dan 90:10. Misalnya rasio yang digunakan 90:10 maka 90% data yang akan digunakan sebagai data training dan 10% data yang akan digunakan sebagai data testing dan begitulah seterusnya. Setelah pembagian data dilakukan, maka data sudah siap untuk diproses menggunakan algoritma *Support Vector Machine* sehingga mendapatkan hasil klasifikasi. Pada tahap pengujian yang dilakukan menggunakan *confusion matrix* dengan menggunakan library pendukung dari ScikitLearn dan rasio yang digunakan 90:10. Hasil dari visualisasi *confusion matrix* dapat dilihat pada gambar 5 berikut ini.



Gambar 5. Confusion Matrix

Berdasarkan gambar 5 diatas ditemukan evaluasi sebagai berikut :

- True Positive* (TP) : 179 jumlah data positif pada penyakit cacar monyet yang terklasifikasi dengan benar pada sistem.
- True Negative* (TN) : 146 jumlah data negatif pada penyakit cacar monyet yang terklasifikasi dengan benar oleh sistem.
- False Positive* (FP) : 103 jumlah data positif pada penyakit cacar monyet tetapi terklasifikasi salah oleh sistem.
- False Negative* (FN) : 72 jumlah data negatif pada penyakit cacar monyet tetapi terklasifikasi salah oleh sistem.

Hasil pengujian yang dilakukan menggunakan parameter tuning dan *confusion matrix* untuk mengukur nilai *precision*, *recall*, *f1-score* dan akurasi. Pengujian dilakukan dengan membandingkan rasio latih dan rasio uji dengan perbandingan 70:30, 80:20, dan 90:10. Pada tabel 7 memperlihatkan hasil pengujian sistem.

**Tabel 7.** Hasil Pengujian

Rasio	Kelas	C	Gamma	Precision	Recall	F1-Score	Akurasi
70:30	Pos	10	0.1	63%	57%	60%	61.8%
	Neg			61%	67%	64%	
80:20	Pos	10	1	64%	70%	67%	64.8%
	Neg			66%	60%	63%	
90:10	Pos	10	1	67%	59%	63%	65%
	Neg			63%	71%	67%	

Berdasarkan pada tabel 7 memperlihatkan hasil akurasi tertinggi ada pada rasio 90:10 dengan nilai akurasi sebesar 65% dengan menghasilkan model terbaik yang didapatkan dari parameter tuning yaitu  $C=10$  dan  $\gamma$  (*gamma*)= 1.

## 4. KESIMPULAN

Pada penelitian ini berhasil menerapkan metode *Support Vector Machine* dalam melakukan klasifikasi pada penyakit cacar monyet. Perancangan sistem ini dilakukan menggunakan *google colab*. Dalam pembangunan model ini menggunakan parameter kernel *Radial Basis Function* (RBF) dengan melakukan eksplorasi dari parameter  $C$  (*cost*) dan  $\gamma$  (*gamma*). Dari hasil eksplorasi tersebut menghasilkan model terbaik dengan nilai parameter  $C=10$  dan  $\gamma$  (*gamma*)= 1. Berdasarkan hasil dari proses training dalam Pembangunan model, dilakukan pengukuran terhadap performa kinerja model dengan menggunakan metode *confusion matrix* dengan rasio 90:10 menghasilkan performa 67% precision, 59% recall dan 63% F1-Score pada kelas positif dan menghasilkan performa 63% precision, 71% recall dan 67% F1-Score pada kelas negatif. Nilai akurasi yang didapatkan dalam pengklasifikasian metode SVM sebesar 65%. Berdasarkan penelitian yang telah dilakukan adapun saran yang dapat dilakukan oleh peneliti selanjutnya yaitu menambahkan jumlah data dan melakukan penelitian dengan data yang seimbang.

## REFERENCES

- [1] K. Suarayasa and O. M. Kristoper, "Mekanisme Penyebaran Cacar Monyet dan Faktor-Faktor Yang Mempengaruhinya," *SEHATMAS (Jurnal Ilm. Kesehat. Masyarakat)*, vol. 2, no. 1, pp. 28–34, 2023, doi: 10.55123/sehatmas.v2i1.980.
- [2] E. E. Hutapea, I. Musfiroh, P. Studi, P. Apoteker, F. Farmasi, and U. Padjadjaran, "Farmaka Farmaka," *Farmaka*, vol. 18, pp. 53–59, 2021.
- [3] L. Qelina and R. Graharti, "Lia Qelina dan Risti Graharti I Human Mongkeypox Virus: Respon kesiapan Dunia Akan Wabah Infeksi Virus Monkeypox Human Monkeypox Virus : Respon Kesiapan Darurat Dunia," vol. 9, pp. 483–489, 2019.
- [4] L. Budiarto, A. A. Sabila, and H. C. Putri, "Infeksi Cacar Monyet (Monkeypox)," *J. Med. Utama*, vol. 4, no. 02 Januari, pp. 3224–3236, 2023.
- [5] D. D. Luthfiani, P. Sianturi, Ali Kusnanto, and H. Sumarno, "Pengaruh Laju Penularan Penyakit Dan Rata-Rata Kontak Individu Pada Model Ko-Infeksi Hiv/Aids Dan Cacar Monyet (Monkeypox)," *J. Math. Its Appl.*, vol. 18, no. 1, pp. 29–39, 2022, doi: 10.29244/milang.18.1.29-39.
- [6] I. Cholissodin, Sutrisno, A. A. Soebroto, U. Hasanah, and Y. I. Febiola, "Welcome to Machine Learning," *AI, Mach. Learn. Deep Learn. (Teori Implementasi)*, pp. 162–163, 2020, [Online]. Available: <http://bit.ly/3piOnnU>
- [7] R. A. Rizal, I. S. Girsang, and S. A. Prasetyo, "Klasifikasi Wajah Menggunakan Support Vector Machine (SVM)," *REMIK (Riset dan E-Jurnal Manaj. Inform. Komputer)*, vol. 3, no. 2, p. 1, 2019, doi: 10.33395/remik.v3i2.10080.
- [8] D. Septhya et al., "Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 15–19, 2023, doi: 10.57152/malcom.v3i1.591.
- [9] R. A. Wati, H. Irsyad, M. Ezar, and A. Rivan, "Klasifikasi Pneumonia Menggunakan Metode Support Vector Machine," *J. Algoritm.*, vol. 1, no. 1, pp. 21–32, 2020.

- [10] S. S. Arifin, A. M. Siregar, A. Ratna, and T. Al Mudzakir, "Klasifikasi Penyakit Kanker Serviks Menggunakan Algoritma Support Vector Machine ( SVM)," *4th Conf. Innov. Appl. Sci. Technol. (CIASTECH 2021)*, no. 4, pp. 521–528, 2021.
- [11] N. G. Ramadhan and A. Khoirunnisa, "Klasifikasi Data Malaria Menggunakan Metode Support Vector Machine," *J. Media Inform. Budidarma*, vol. 5, no. 4, p. 1580, 2021, doi: 10.30865/mib.v5i4.3347.
- [12] N. Maulidah, R. Supriyadi, D. Y. Utami, F. N. Hasan, A. Fauzi, and A. Christian, "Prediksi Penyakit Diabetes Melitus Menggunakan Metode Support Vector Machine dan Naive Bayes," *Indones. J. Softw. Eng.*, vol. 7, no. 1, pp. 63–68, 2021, doi: 10.31294/ijse.v7i1.10279.
- [13] Z. Alhaq, A. Mustopa, S. Mulyatun, and J. D. Santoso, "Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter," *J. Inf. Syst. Manag.*, vol. 3, no. 2, pp. 44–49, 2021, doi: 10.24076/joism.2021v3i2.558.
- [14] Suripto, R. N. Rahmanita, and A. S. Kirana, "Teknik pre-processing dan classification dalam data science," *Binus University*, 2022. <https://mie.binus.ac.id/2022/08/26/teknik-pre-processing-dan-classification-dalam-data-science/>
- [15] R. Dharma, "Data Preprocessing: Pengertian, Manfaat, dan Tahapan Kerjanya," *Accurate*, 2022. <https://accurate.id/teknologi/data-preprocessing/>
- [16] Y. Amrozi, D. Yuliati, A. Susilo, N. Novianto, and R. Ramadhan, "Klasifikasi Jenis Buah Pisang Berdasarkan Citra Warna dengan Metode SVM," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 3, pp. 394–399, 2022, doi: 10.32736/sisfokom.v11i3.1502.
- [17] R. Yohannes and M. E. Al Rivian, "Klasifikasi Jenis Kanker Kulit Menggunakan CNN-SVM," *J. Algoritm.*, vol. 2, no. 2, pp. 133–144, 2022, doi: 10.35957/algoritme.v2i2.2363.
- [18] A. Zalvadila, L. Syafie, and H. Darwis, "Klasifikasi Penyakit Tanaman Bawang Merah Menggunakan Metode SVM dan CNN," *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, pp. 255–260, 2023.
- [19] I. M. D. P. Asana and N. P. D. T. Yanti, "Sistem Klasifikasi Pengajuan Kredit Dengan Metode Support Vector Machine (SVM) I Made Dwi Putra Asana," *J. Sist. Cerdas*, vol. 6, no. 2, pp. 123–133, 2023.
- [20] Suraj Yadav, "What is Kernel Trick in SVM? Interview questions related to Kernel Trick," *medium.com*, 2023. [https://medium.com/@Suraj\\_Yadav/what-is-kernel-trick-in-svm-interview-questions-related-to-kernel-trick-97674401c48d](https://medium.com/@Suraj_Yadav/what-is-kernel-trick-in-svm-interview-questions-related-to-kernel-trick-97674401c48d) (accessed May 06, 2024).
- [21] M. Sahbuddin and S. Agustian, "JITE ( Journal of Informatics and Telecommunication Engineering ) Support Vector Machine Method with Word2vec for Covid-19 Vaccine Sentiment Classification on Twitter," vol. 6, no. July, pp. 288–297, 2022.
- [22] D. E. Safitri and A. S. Fitriani, "Implementasi Metode Klasifikasi Dengan Algoritma Support Vector Machine Kernel Gaussian Rbf Untuk Prediksi Partisipasi Pemilu Terhadap Demografi Kota Surabaya," *Indones. J. Bus. Intell.*, vol. 5, no. 1, p. 36, 2022, doi: 10.21927/ijubi.v5i1.2259.