



## *Classification of Monkeypox Disease Using Support Vector Machine (SVM)*

### **Klasifikasi Penyakit Cacar Monyet Menggunakan Support Vector Machine (SVM)**

Rohmatullah Sony Wijaya<sup>1\*</sup>, Arie Qur'ania<sup>2</sup>, Irma Anggraeni<sup>3</sup>

<sup>1,2,3</sup>Program Studi Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Pakuan

E-Mail: <sup>1</sup>rsonywijaya@gmail.com, <sup>2</sup>qurania@unpak.ac.id, <sup>3</sup>irmairhamna@unpak.ac.id

Received May 28th 2024; Revised Jun 23th 2024; Accepted Jul 15th 2024  
Corresponding Author: Rohmatullah Sony Wijaya

#### **Abstract**

Monkeypox disease became an epidemic in several non-endemic countries in July 2022. Therefore, appropriate prevention or treatment measures need to be taken early by identifying the disease using a classification method. Classification is done using the knowledge discovery in database (KDD) method with the support vector machine (SVM) algorithm that uses 4 kernels which is linear, RBF, sigmoid, and polynomial with parameter settings for each kernel. The SVM algorithm was chosen because it uses different kernels to explore different decision forms in the higher feature space to capture non-linear patterns. The best results were obtained by the polynomial kernel with an accuracy rate of 75% while the linear kernel was 70,5%, RBF 66%, and sigmoid 45%. The value of ROC AUC curve graph for the polynomial kernel is 0.81. This shows that the classification model created is good and can be developed into the next stage of research.

Keyword: Classification, Kernels, Monkeypox, ROC AUC Curve Graph, SVM

#### **Abstrak**

Penyakit cacar monyet menjadi sebuah wabah di beberapa negara non endemik pada bulan Juli tahun 2022. Oleh karena itu, tindakan pencegahan atau pengobatan yang tepat perlu dilakukan secara dini dengan cara melakukan identifikasi penyakit menggunakan suatu metode klasifikasi. Klasifikasi dilakukan menggunakan metode *knowledge discovery in database* (KDD) dengan algoritma *support vector machine* (SVM) yang menggunakan 4 kernel yaitu *linear*, *RBF*, *sigmoid*, dan *polynomial* dengan pengaturan parameternya pada masing masing kernel. Algoritma SVM dipilih karena penggunaan berbagai kernelnya memungkinkan eksplorasi bentuk-bentuk keputusan yang berbeda dalam ruang fitur yang lebih tinggi untuk menangkap pola pola yang tidak linear. Hasil terbaik didapatkan oleh kernel *polynomial* dengan tingkat akurasi sebesar 75%, sementara kernel *linear* sebesar 70,5%, *RBF* sebesar 66%, dan *sigmoid* sebesar 45%. Kemudian nilai grafik kurva *receiver operating characteristic area under control* (ROC AUC) untuk kernel *polynomial* sebesar 0.81. Hal tersebut menunjukkan bahwa model klasifikasi yang dibuat sudah baik dan dapat dikembangkan ke penelitian tahap selanjutnya.

Kata Kunci: Cacar Monyet, Grafik Kurva ROC AUC, Kernel, Klasifikasi, SVM

#### **1. PENDAHULUAN**

Cacar adalah suatu kondisi pada kulit manusia yang menyebabkan timbulnya jerawat pada seluruh tubuh dan wajah dengan berisi cairan pada setiap jerawatnya [1]. Terdapat beberapa jenis cacar seperti cacar air, cacar api, dan cacar monyet (*monkeypox*). *Monkeypox disease* atau penyakit cacar monyet merupakan salah satu jenis penyakit cacar yang terjadi pada bulan Juli hingga Oktober 2022 mewabah di beberapa negara non-endemik saat pandemi covid-19 belum berakhir. Pada pertengahan tahun 2022, *monkeypox* mewabah di Amerika Serikat sehingga *Center for Disease Control and Prevention* (CDC) dan *World Health Organization* (WHO) mendeklarasikan *monkeypox* sebagai darurat kesehatan global [2]. Individu yang terpapar *monkeypox* mengalami gejala demam, ruam, sakit punggung dan otot, kelelahan, dan pembekakan kelenjar getah bening [3]. *Monkeypox* sendiri dapat dengan mudah menular kepada manusia melalui kontak langsung antara individu dengan penderita, terkena hewan terinfeksi, atau dengan mengonsumsi daging hewan yang dimasak dengan cara tidak benar [4]. Oleh karena itu, tindakan yang sangat perlu dilakukan adalah pencegahan atau pengobatan

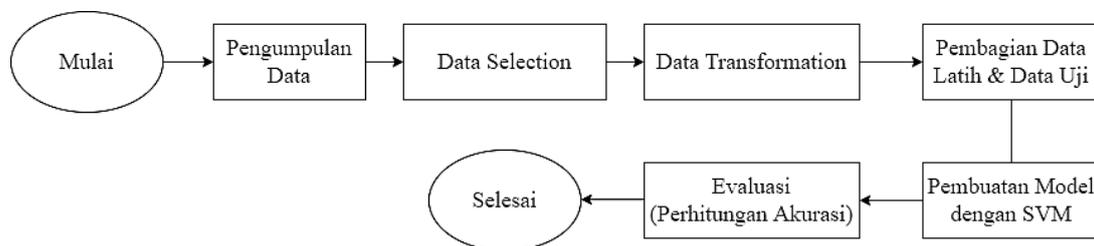
yang tepat dengan cara melakukan identifikasi penyakit menggunakan suatu metode klasifikasi. Salah satunya adalah dengan menggunakan algoritma Support Vector Machine (SVM).

Algoritma SVM merupakan salah satu jenis algoritma dalam teknik *machine learning*. Ketersediaan beberapa kernel membuat berbagai jenis dataset dapat diimplementasikan dengan metode SVM ini. Algoritma SVM juga menggunakan prinsip SRM atau *Structural Risk Minimization* sehingga performa klasifikasinya meminimalisasi kesalahan saat data dilatih [5]. Ketika melakukan prediksi klasifikasi potensi pada data, algoritma SVM dikenal dengan metodenya yang mempunyai hasil tinggi [6]. Terdapat penelitian terdahulu yang dilakukan menggunakan 4 algoritma machine learning untuk mengklasifikasikan penyakit *monkeypox*, yaitu SVM, Random Forest, Gradient Boost, dan KNN. Dari keempat *machine learning* yang digunakan, GB mendapatkan performa yang paling optimal dengan akurasi sebesar 70% [7]. Kemudian sebuah studi mengimplementasikan metode *Prophet* untuk prediksi yang lebih akurat terhadap epidemi *monkeypox*. Dari penelitian yang dilakukan selama 7 sampai 17 hari, metode prophet diimplementasikan dalam peramalannya dan didapatkan nilai MAPE di bawah 20% yang menunjukkan bahwa hasil ramalan cukup baik yaitu antara 0,018 dan 0,117 [8]. Selain itu, terdapat penelitian lain yang membandingkan algoritma klasifikasi yaitu *machine learning* dengan *deep learning*. Algoritma yang digunakan yaitu SVM dan *Deep Neural Network*. Dari penelitian tersebut, didapatkan hasil yang lebih unggul menggunakan *Deep Neural Network* dibandingkan algoritma *machine learning* dengan fungsi *sigmoid* sehingga akurasinya sebesar 70,08%, F1 sebesar 79,18%, presisi sebesar 68,59%, *recall* sebesar 62,65%, dan AUC sebesar 62,65% [9].

Berdasarkan permasalahan yang ada pada latar belakang dan penelitian terdahulu, maka akan dibuat penelitian dengan mengimplementasikan algoritma SVM untuk klasifikasi penyakit *monkeypox*. Penelitian ini mengutamakan eksplorasi penggunaan kernel yang ada pada algoritma SVM sehingga tidak hanya memberikan model klasifikasi yang lebih akurat, tetapi juga memperluas pemahaman tentang penggunaan kernel dari algoritma SVM. Diharapkan dengan penelitian ini dapat menjadi sebuah model baru dalam pengklasifikasian penyakit *monkeypox* yang memiliki nilai akurasi tinggi.

## 2. METODOLOGI PENELITIAN

Penelitian ini melakukan klasifikasi terhadap dataset penyakit *monkeypox* menggunakan algoritma SVM. Tahapan-tahapannya disajikan dalam bentuk diagram alur supaya penelitian yang dilakukan sesuai dengan urutan yang sudah ditentukan. Tahapan tersebut dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

### 2.1. Pengumpulan Data

Pengumpulan data dilakukan dengan mengambil dataset penyakit *monkeypox* dari Kaggle yang berisi sebanyak 1000 baris dan memiliki 11 atribut, yaitu *Patient\_ID*, *Systemic Illness*, *Rectal Pain*, *Sore Throat*, *Penile Oedema*, *Oral Lesions*, *Solitary Lesion*, *Swollen Tonsils*, *HIV Infection*, *Sexually Transmitted Infection*, dan *MonkeyPox* yang memiliki 2 kelas target (positif dan negatif). Perbandingan jumlah data kelas positif dan kelas negatif yaitu 1:1 atau sebanyak 500 baris untuk setiap kelasnya.

### 2.2. Data Selection

Tahap data selection dilakukan untuk menyeleksi atribut yang digunakan pada proses analisis supaya proses klasifikasi dapat berjalan dengan baik. Proses seleksi data ini penting untuk dilakukan agar proses mining data dapat lebih efektif [10]. Pada tahap ini, atribut *Patient\_ID* tidak digunakan karena hanya berisi nomor unik dan tidak mempunyai pengaruh terhadap analisis data sehingga atribut yang digunakan pada proses analisis berjumlah 10.

### 2.3. Data Preprocessing

*Data Preprocessing* merupakan salah satu tahapan untuk mengatasi masalah yang dapat mempengaruhi hasil klasifikasi data [11]. Tahap data preprocessing dilakukan untuk mengecek apakah terdapat data yang kosong atau tidak pada data mentah sebelum dilakukan proses analisis. Pada penelitian ini, dataset penyakit *monkeypox* yang digunakan sudah baik dan tidak ada data yang hilang sehingga tidak dilakukan penanganan data yang hilang dan normalisasi karena data akan langsung dilakukan transformasi.

#### 2.4. Data Transformation

Transformasi data adalah langkah untuk mengubah data ke dalam bentuk format tertentu sehingga proses penggalian informasi dapat dilakukan [12]. Pada penelitian ini, dataset yang digunakan memiliki 2 tipe data, yaitu *integer* dan *boolean*. Kedua tipe data tersebut kemudian akan diubah ke *integer* untuk dilakukan proses *data mining*.

#### 2.5. Pembagian Data Latih dan Data Uji

Klasifikasi merupakan sebuah proses untuk mencari pola dalam dataset yang dapat memilah kelas dari suatu data dengan data yang lain untuk menentukan apakah objek pada data tersebut masuk ke dalam suatu kelas dengan melihat atribut dan ciri dari kelas yang sudah ditentukan [13]. Model klasifikasi yang dibuat berdasarkan proses dari pelatihan data menggunakan data latih. Sedangkan untuk pengujian dan evaluasi kualitas dari model klasifikasi yang dikembangkan menggunakan data uji [14]. Pada penelitian ini dilakukan pembagian data latih dan data uji yaitu sebesar 80:20.

#### 2.6. Support Vector Machine (SVM)

SVM merupakan salah satu sistem pembelajaran mesin dan memakai sebuah ruang dengan hipotesis yang terdiri dari beberapa fungsi linear pada suatu fitur dengan dimensi tinggi serta dilatih oleh algoritma pembelajaran yang berasal dari teori untuk optimasi. Secara sederhana, konsep SVM yaitu upaya untuk memisahkan dua kelas dengan mencari hyperplane terbaik pada suatu input space dengan memaksimalkan jarak antarkelas [15]. Ketika melakukan prediksi klasifikasi potensi pada data, algoritma SVM dikenal dengan metodenya yang mempunyai hasil tinggi [6]. Berikut adalah persamaan algoritma SVM.

$$y_i (x_i \cdot w + b) \geq 0 \quad (1)$$

Untuk klasifikasi nonlinear, SVM menggunakan *kernel trick* dimana efektivitasnya dipengaruhi oleh jenis fungsi kernel yang digunakan berdasarkan karakteristik dataset. Kernel trick bekerja dengan memetakan vektor input (*input space*) dari ruang dengan dimensi rendah ke ruang dengan dimensi tinggi (*feature space*). Formulasi fungsi kernel trick adalah sebagai berikut [16]:

$$K(x_i, x_j) = \phi_i(x_i) \cdot \phi_j(x_j) \quad (2)$$

Terdapat beberapa jenis *kernel trick* yang biasa digunakan, yaitu *polynomial*, RBF, dan *sigmoid*. Kernel *polynomial* umumnya digunakan untuk klasifikasi data dengan data latih yang digunakan sudah normal. Parameter yang umum digunakan yaitu *degree* (d), *koefisien* (coef), C, dan *gamma*. Persamaan kernel *polynomial* adalah sebagai berikut:

$$K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i \cdot \vec{x}_j + 1)^d \quad (3)$$

Kernel *radial basis function* (RBF) atau kernel *Gaussian* menggunakan parameter yang umumnya digunakan kernel RBF adalah *gamma* dan C. Kernel *Gaussian* lebih banyak digunakan karena kinerja pembelajarannya yang sangat baik [17]. Persamaan kernel *Gaussian* adalah sebagai berikut:

$$K(\vec{x}_i, \vec{x}_j) = \exp(-\gamma \|\vec{x}_i - \vec{x}_j\|^2) \quad (4)$$

Kernel *sigmoid* adalah salah satu kernel yang dikembangkan dari JST atau jaringan saraf tiruan. Parameter yang umumnya digunakan kernel *sigmoid* yaitu *gamma* dan C yang persamaannya adalah sebagai berikut:

$$K(\vec{x}_i, \vec{x}_j) = \tanh(\alpha \vec{x}_i \cdot \vec{x}_j + \beta) \quad (5)$$

#### 2.7. Evaluasi

Evaluasi dilakukan setelah semua proses data mining dilakukan. Pola yang terbentuk maupun model prediksi dari hasil data mining akan diidentifikasi dan dievaluasi untuk menilai apakah hasilnya sesuai dengan hipotesa atau tidak sehingga menjadi sebuah *knowledge* (pengetahuan) menggunakan *confusion matrix* [18]. Dimana perhitungan nilai akurasi dan galat (error) diperlukan untuk menyatakan jumlah data yang benar diklasifikasi dan menyatakan ketidakbenaran model untuk mengklasifikasi data [19]. Berikut ini persamaan *confusion matrix*:

$$\text{Akurasi} = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \quad (6)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \times 100\% \quad (7)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (8)$$

$$f1 - \text{score} = \frac{2 \times (\text{presisi} \times \text{recall})}{(\text{presisi} + \text{recall})} \times 100\% \quad (9)$$

Kemudian dilakukan proses perhitungan menggunakan *k-fold cross validation* untuk mengevaluasi model klasifikasi yang sudah dibuat. Pola yang terbentuk ditampilkan dengan membuat visualisasi data agar hasilnya lebih mudah dipahami dan dapat dibuat keputusan setelahnya. Berikut persamaan *k-fold cross validation*:

$$\text{Akurasi} = \frac{\sum \text{data uji benar klasifikasi}}{\sum \text{total data uji}} \times 100\% \quad (10)$$

Kurva ROC atau *receiver operating characteristic* menjadi sebuah alat ukur yang digunakan untuk menilai kemampuan klasifikasi suatu sistem yang dibuat. Kurva ROC ini berbentuk sebuah grafik yang membandingkan *true positive rate* atau *sensitivity* dengan *false positive rate* atau *specificity* [20].

### 3. HASIL DAN DISKUSI

Pengumpulan data dilakukan dengan mengambil dataset penyakit monkeypox dari Kaggle yang berisi sebanyak 1000 baris dan memiliki 11 atribut dan memiliki 2 kelas target (positif dan negatif). Data yang digunakan dapat dilihat pada Tabel 1.

**Tabel 1.** Dataset Penyakit *Monkeypox*

Patient ID	Systemic Illness	Rectal Pain	Sore Throat	Penile Oedema	Oral Lesions	Solitary Lesiion	Swollen Tonsils	HIV Infection	Sexually Transmitted Infection	MonkeyPox
P1	Fever Swollen	True	False	True	True	False	False	True	False	Positive
P2	Lymph Nodes	True	True	True	False	True	False	True	False	Negative
P3	Fever Muscle	True	False	True	True	True	True	True	True	Positive
P4	Aches And Pain	False	True	False	True	False	True	False	False	Negative
P5	None	True	False	True	True	True	True	True	True	Positive
....	...	...	...	...	...	...	...	...	...	...
P996	None Swollen	True	False	True	True	False	False	True	False	Positive
P997	Lymph Nodes	True	True	True	False	True	False	True	False	Negative
P998	Fever	True	False	True	True	True	True	True	True	Positive
P999	None	False	True	False	True	False	True	False	False	Negative
P1000	None	True	False	True	True	True	True	True	True	Positive

Setelah dataset didapatkan, dilakukan seleksi atribut dimana kolom *Patient\_ID* tidak digunakan pada proses data mining karena atribut tersebut tidak berpengaruh terhadap proses analisis data mining dan tidak diperlukan sehingga atribut yang digunakan hanya 10.

Kemudian dilakukan tahap *data preprocessing* untuk mengecek apakah terdapat data yang kosong atau tidak. Pada penelitian ini dataset yang digunakan sudah baik dan tidak ada data yang hilang sehingga tidak dilakukan penanganan data yang hilang dan normalisasi data karena data akan langsung dilakukan transformasi. Transformasi data dilakukan terhadap seluruh atribut data termasuk mengubah tipe datanya. Data yang telah ditransformasi dapat dilihat pada Tabel 2.

**Tabel 2.** Hasil Transformasi Data

Patient ID	Systemic Illness	Rectal Pain	Sore Throat	Penile Oedema	Oral Lesions	Solitary Lesiion	Swollen Tonsils	HIV Infection	Sexually Transmitted Infection	Monkey Pox
P1	2	1	0	1	1	0	0	1	0	1
P2	2	1	1	1	0	1	0	1	0	-1

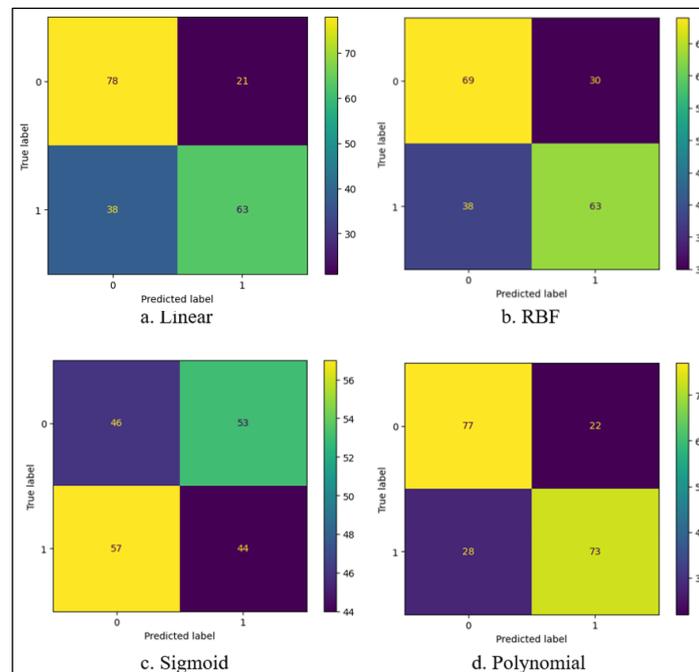
Patient ID	Systemic Illness	Rectal Pain	Sore Throat	Penile Oedema	Oral Lesions	Solitary Lesiion	Swollen Tonsils	HIV Infection	Sexually Transmitted Infection	Monkey Pox
P3	1	1	0	1	1	1	1	1	1	1
P4	3	0	1	0	1	0	1	0	0	-1
P5	0	1	0	1	1	1	1	1	1	1
....	...	...	...	...	...	...	...	...	...	...
P996	0	1	0	1	1	0	0	1	0	1
P997	2	1	1	1	0	1	0	1	0	-1
P998	1	1	0	1	1	1	1	1	1	1
P999	0	0	1	0	1	0	1	0	0	-1
P1000	0	1	0	1	1	1	1	1	1	1

Setelah dilakukan transformasi, tahap selanjutnya dataset dibagi menjadi data latih dan data uji dengan perbandingan 80:20 dan mengimplementasikan beberapa kernel yang ada pada SVM dengan parameter tertentu sehingga dihasilkan berbagai nilai akurasi yang berbeda. Kernel *linear* menggunakan parameter  $C = 1$ , kernel *polynomial* menggunakan parameter  $d = 2$ ,  $C = 1$ ,  $\gamma = 'scale'$ , dan  $\text{coef} = 1$ , kernel RBF menggunakan parameter  $\gamma = 1$ , dan  $C = 10$ , dan kernel *sigmoid* menggunakan parameter  $C = 1$ ,  $\gamma = 'scale'$ , dan  $\text{coef} = 0.1$ . Berikut ini perbandingan nilai akurasi menggunakan *confusion matrix* terhadap setiap kernel pada Tabel 3.

**Tabel 3.** Perbandingan Nilai Akurasi Setiap Kernel

No	Jenis Kernel	Akurasi
1	<i>Linear</i>	70,5%
2	<i>Polynomial</i>	75%
3	RBF	66%
4	<i>Sigmoid</i>	45%

Dapat dilihat dari Tabel 3 di atas perbandingan antara beberapa kernel yang digunakan dalam proses klasifikasinya memiliki perbedaan yang tidak terlalu jauh. Nilai akurasi tertinggi didapatkan oleh kernel *polynomial* sebesar 75% dan akurasi terendah didapatkan oleh kernel *sigmoid* sebesar 45%. Pengujian akurasi tersebut menggunakan *confusion matrix* yang hasilnya bisa dilihat pada Gambar 2.



**Gambar 2.** Confusion Matrix

Berikut ini detail yang lebih rinci mengenai kinerja model klasifikasi SVM dengan kernel *linear* pada Tabel 4.

**Tabel 4.** Hasil Kernel Linear

Kelas	Akurasi (train)	Akurasi (test)	Presisi	Recall	F1-Score
Negatif	67.875	70.5	0.67	0.73	0.73
Positif	67.875	70.5	0.75	0.68	0.68

Berikut ini detail yang lebih rinci mengenai kinerja model klasifikasi SVM dengan kernel RBF pada Tabel 5.

**Tabel 5.** Hasil Kernel RBF

Kelas	Akurasi (train)	Akurasi (test)	Presisi	Recall	F1-Score
Negatif	92.875	66.0	0.64	0.70	0.67
Positif	92.875	66.0	0.68	0.60	0.65

Berikut ini detail yang lebih rinci mengenai kinerja model klasifikasi SVM dengan kernel *sigmoid* pada Tabel 6.

**Tabel 6.** Hasil Kernel Sigmoid

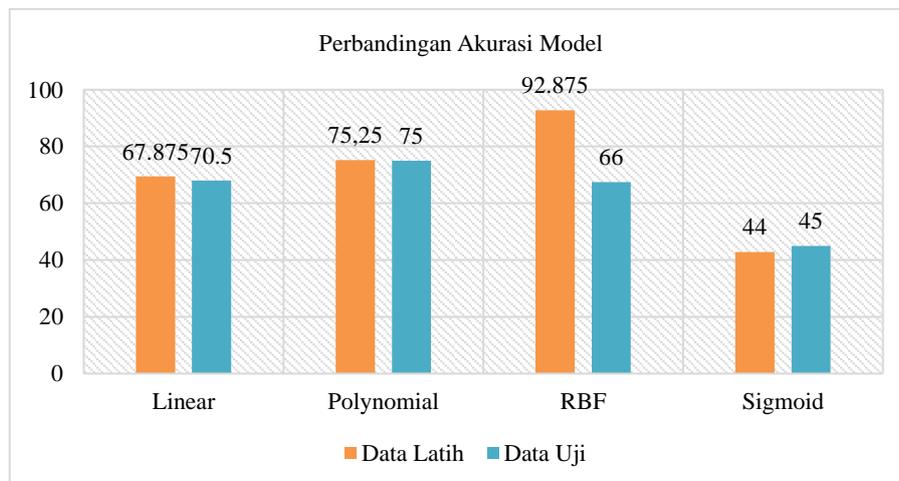
Kelas	Akurasi (train)	Akurasi (test)	Presisi	Recall	F1-Score
Negatif	44.0	45.0	0.66	0.80	0.72
Positif	44.0	45.0	0.75	0.60	0.67

Berikut ini detail yang lebih rinci mengenai kinerja model klasifikasi SVM dengan kernel *polynomial* pada Tabel 7.

**Tabel 7.** Hasil Kernel Polynomial

Kelas	Akurasi (train)	Akurasi (test)	Presisi	Recall	F1-Score
Negatif	75.25	75.0	0.73	0.78	0.75
Positif	75.25	75.0	0.77	0.72	0.74

Dari pengukuran performa yang dilakukan pada setiap kernel menggunakan confusion matrix, didapatkan performa terbaik dengan kernel polynomial dengan akurasi data latih sebesar 75,25% dan akurasi data uji sebesar 75%. Meskipun hasil kernel RBF menghasilkan akurasi data latih yang sangat tinggi yaitu sebesar 92,875%, tetapi hasil akurasi data ujinya hanya sebesar 66%. Hal ini menandakan bahwa batas keputusannya lebih kompleks dan model menjadi terlalu fit sehingga model kurang menggeneralisasi dengan baik ke data baru. Berikut ini grafik perbandingan hasil akurasi data latih dan data uji pada Gambar 3.

**Gambar 3.** Grafik Perbedaan Akurasi Setiap Kernel

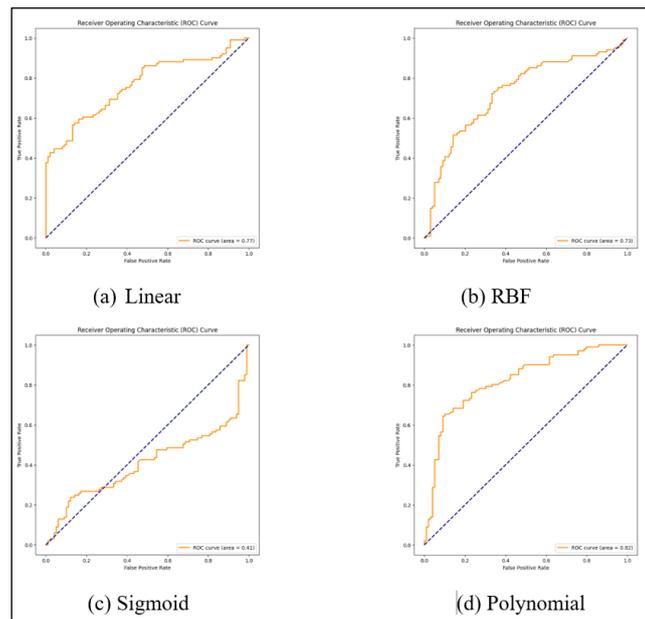
Kemudian dilakukan uji validasi terhadap semua kernel menggunakan *k-fold cross validation* dan didapatkan hasil pada Tabel 8.

**Tabel 8.** Perbandingan Nilai K-Fold Setiap Kernel

No	Jenis Kernel	Rata-Rata Nilai K-Fold
1	Linear	67%
2	Polynomial	70,5%
3	RBF	68%
4	Sigmoid	43%

Tabel 8 di atas memperlihatkan perbandingan nilai uji validasi dari setiap kernel. Nilai k yang digunakan adalah 10 dan didapatkan rata-rata nilai k-fold tertinggi oleh kernel *polynomial* sebesar 70,5% sedangkan rata-rata nilai k-fold terendah oleh kernel *sigmoid* sebesar 43%.

Model yang dibuat juga dinilai menggunakan grafik kurva ROC pada setiap kernel dan menunjukkan AUC (Area Under Control) untuk mengevaluasi modelnya. Grafik kurva ROC untuk setiap kernel bisa dilihat pada Gambar 4.



**Gambar 4.** Grafik Kurva ROC Setiap Kernel SVM

Berdasarkan gambar 4 di atas dapat diketahui untuk interpretasi nilai AUC pada masing masing kernel yaitu, *linear* sebesar 0.77, RBF sebesar 0.73, *sigmoid* sebesar 0.43, dan *polynomial* sebesar 0.81. Grafik ROC AUC memberikan nilai dari 0 sampai dengan 1 dan dapat diketahui bahwa kernel *polynomial* memiliki nilai terbesar dari kernel lainnya yaitu sebesar 0.81.

Kelebihan utama dari algoritma SVM terlihat dari kernel *polynomial* yang secara khusus mencapai tingkat akurasi lebih tinggi dibandingkan dengan kernel lainnya. Hal ini menunjukkan adaptabilitas dari algoritma SVM dalam mengatasi masalah data yang kompleks. Meskipun penggunaan algoritma SVM untuk klasifikasi penyakit monkeypox mendapatkan tingkat akurasi sebesar 75% dengan kernel *polynomial*, tetapi perlu dilakukan pengaturan parameter yang tepat untuk setiap kernel yang digunakan dan disesuaikan dengan karakteristik dari dataset serta tujuan klasifikasi agar mendapatkan akurasi yang tinggi.

**4. KESIMPULAN**

Berdasarkan penelitian yang telah dilakukan, didapatkan akurasi data latih dan data uji menggunakan kernel *linear* sebesar 67.875% dan 70.5%, kernel RBF sebesar 92,875% dan 66%, kernel *sigmoid* sebesar 44% dan 45%, dan kernel *polynomial* sebesar 75.25% dan 75%. Hal ini menunjukkan bahwa kernel *polynomial* lebih efektif dibandingkan kernel lainnya dengan menggunakan parameter  $d = 2$ ,  $C = 1$ ,  $\gamma = 'scale'$ , dan  $\text{coef} = 1$  yang mendapat nilai akurasi sebesar 75%. Meskipun kernel RBF mendapatkan nilai akurasi data latih yang lebih tinggi dari kernel *polynomial* yaitu sebesar 92.875%, tetapi nilai akurasi data ujinya hanya 66% yang membuat model menjadi terlalu fit (*overfitting*) dan kurang menggeneralisir ke data baru sehingga kernel RBF tidak disarankan penggunaannya pada penelitian ini. Hasil validasi model klasifikasi tertinggi dengan kernel *polynomial* sebesar 70.5% menggunakan *k-fold cross validation* dengan nilai  $k = 10$ . Nilai grafik kurva ROC AUC pada kernel *polynomial* didapatkan sebesar 0.81 dimana lebih besar dari ketiga kernel lainnya yaitu *linear* sebesar 0.77, RBF sebesar 0.73, dan *sigmoid* sebesar 0.43. Setelah dilakukan perbandingan dengan penelitian terdahulu, dapat disimpulkan bahwa algoritma SVM mampu mengklasifikasikan dataset penyakit

monkeypox tersebut dengan cukup efektif menggunakan kernel *polynomial*. Dengan demikian, penelitian ini sudah menjawab permasalahan utama yaitu kebutuhan akan model yang efektif untuk mengklasifikasikan penyakit *monkeypox* meskipun hasilnya tidak terlalu memuaskan. Oleh karena itu, disarankan untuk penelitian berikutnya dapat menggunakan metode algoritma klasifikasi lain yang dapat meningkatkan kinerja model serta akurasi. Selain itu juga diharapkan dapat menggunakan dataset yang lebih besar dan melakukan uji coba secara berulang guna memperbaiki proses pelatihan model dan menghasilkan akurasi yang tinggi.

## REFERENSI

- [1] Y. Nawaz, I. Nazar, R. Z. Sattar, B. Saeed, and M. K. Ullah, "A Review on Smallpox," *J. Biomed. Sci.*, vol. 11, no. 6, p. 65, 2022, doi: 10.36648/2254-609X.11.6.65.
- [2] T. Mcandrew *et al.*, "Human judgment forecasts of human monkeypox transmission and burden in non-endemic countries," 2022.
- [3] A. Wang, D. Li, W. Shen, and X. Zhang, "Monkeypox Cases Prediction with Machine Learning," vol. 39, pp. 246–257, 2023.
- [4] E. Petersen *et al.*, "Monkeypox - Enhancing public health preparedness for an emerging lethal human zoonotic epidemic threat in the wake of the smallpox post-eradication era.," *J. Autoimmun.*, pp. 78–84, 2019, [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0896841122000634>
- [5] Teguh Arifiyanto, S. Sunaryo, and Lady Silk Moonlight, "Penggunaan Metode Support Vector Machine (Svm) Pada Teknologi Mobil Masa Depan Menggunakan Sidik Jari," *J. Tek. Inform. dan Teknol. Inf.*, vol. 2, no. 2, pp. 94–97, 2022, doi: 10.55606/jutiti.v2i2.363.
- [6] H. S. W. Hovi, A. Id Hadiana, and F. Rakhmat Umbara, "Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM)," *Informatics Digit. Expert*, vol. 4, no. 1, pp. 40–45, 2022, doi: 10.36423/index.v4i1.895.
- [7] F. Aldi, I. Nozomi, R. B. Sentosa, and A. Junaidi, "Machine Learning to Identify Monkey Pox Disease," *Sinkron*, vol. 8, no. 3, pp. 1335–1347, 2023, doi: 10.33395/sinkron.v8i3.12524.
- [8] Y. Huang, L. Mu, and W. Wang, "Monkeypox: epidemiology, pathogenesis, treatment and prevention," *Signal Transduct. Target. Ther.*, vol. 7, no. 1, pp. 1–22, 2022, doi: 10.1038/s41392-022-01215-4.
- [9] C. I. Agustyaningrum, R. Dahlia, and O. Pahlevi, "Comparison of Conventional Machine Learning and Deep Neural Network Algorithms in the Prediction of Monkey-Pox," *J. Ris. Inform.*, vol. 5, no. 3, pp. 523–532, 2023, doi: 10.34288/jri.v5i2.522.
- [10] R. S. Nugraha, A. D. Rachmanto, and Z. Munawar, "Analisis Keakuratan Nilai Bitcoin Menggunakan Algoritma Regresi Linier Pada Data Mining," *J. Teknol. Inf. dan Komun.*, vol. 12, no. 2, pp. 2–10, 2023, doi: 10.56244/fiki.v12i2.668.
- [11] M. D. Purbolaksono, M. Irvan Tantowi, A. Imam Hidayat, and A. Adiwijaya, "Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 393–399, 2021, doi: 10.29207/resti.v5i2.3008.
- [12] I. K. J. Arta, G. Indrawan, and G. R. Dantes, "Data Mining Rekomendasi Calon Mahasiswa Berprestasi Di Stmik Denpasar Menggunakan Metode Technique for Others Reference By Similarity To Ideal Solution," *JST (Jurnal Sains dan Teknol.)*, vol. 5, no. 2, pp. 11–21, 2019, doi: 10.23887/jstundiksha.v5i2.8549.
- [13] I. Romli and A. T. Zy, "Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 4, no. 2, pp. 694–702, 2020.
- [14] Galih, "Data Mining di Bidang Pendidikan untuk Analisa Prediksi Kinerja Mahasiswa dengan Komparasi 2 Model Klasifikasi pada STMIK Jabar," *J. Teknol. Sist. Inf. dan Apl.*, vol. 2, no. 1, p. 23, 2019, doi: 10.32493/jtsi.v2i1.2643.
- [15] A. M. Puspitasari, D. E. Ratnawati, and A. W. Widodo, "Klasifikasi Penyakit Gigi Dan Mulut Menggunakan Metode Support Vector Machine," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 2, pp. 802–810, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [16] A. Kowalczyk, *Support Vector Machine Succinctly*, First. USA, 2017.
- [17] J. Yang *et al.*, "Parameter selection of Gaussian kernel SVM based on local density of training set," *Inverse Probl. Sci. Eng.*, vol. 29, no. 4, pp. 536–548, 2021, doi: 10.1080/17415977.2020.1797716.
- [18] M. A. Muslim *et al.*, "Data Mining Algoritma C4.5," 2019.
- [19] B. P. Pratiwi, A. S. Handayani, and S. Sarjana, "Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix," *J. Inform. Upgris*, vol. 6, no. 2, pp. 66–75, 2021, doi: 10.26877/jiu.v6i2.6552.
- [20] L. Qadrini, A. Sepperwali, and A. Aina, "Decision Tree Dan Adaboost Pada Klasifikasi Penerima Program Bantuan Sosial," *J. Inov. Penelit.*, vol. 2, no. 7, pp. 1959–1966, 2021.