

KLASIFIKASI PENYAKIT JANTUNG SESEORANG MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

Retsya Lapiza¹ Bahri A Sinaga² Ade E Saputri³ Velya R Putri⁴

Program Studi Sarjana Statistika, Universitas Negeri Padang

retsyalapiza00@gmail.com

ABSTRAK

Penyakit jantung atau kardiovaskular masih menjadi salah satu penyebab kematian tertinggi di dunia dan terus meningkat setiap tahunnya. Untuk itu perlu dilakukan upaya pencegahan sedini mungkin, misalnya dengan memprediksi seseorang terkena penyakit jantung dengan menggunakan pendekatan Machine Learning yaitu Support Vector Machine (SVM). SVM merupakan metode dalam analisis statistika yang digunakan untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. Dengan begitu, SVM dinilai cocok untuk memprediksi dan mendiagnosa penyakit jantung seseorang. Salah satu kelebihan dari Support Vector Machine adalah baik digunakan untuk data yang tidak terlalu besar dan nilai akurasi yang diperoleh baik. Pemodelan dilakukan dengan menggunakan beberapa fungsi kernel pada model linear, polynomial, radial dan sigmoid. Hasil penelitian ini menunjukkan bahwa akurasi prediksi SVM menggunakan polynomial kernel sebesar 78,69%, artinya metode dapat digunakan untuk memprediksi penyakit jantung dengan tingkat akurasi yang cukup tinggi.

Keyword: *Machine Learning, Support Vector Machine, Classification, Heart Disease*

PENDAHULUAN

Penyakit jantung atau penyakit kardiovaskuler merupakan gangguan dari jantung dan pembuluh darah. Penyakit jantung dapat menyebabkan komplikasi serius diantaranya nyeri dada (angina), irama jantung tidak normal (aritmia), gagal jantung hingga menyebabkan kematian. Penduduk di Dunia yang meninggal karena penyakit jantung terus meningkat setiap tahunnya. Hasil survey dari *sample registration system* pada tahun 2014 di Indonesia penyakit jantung berada di posisi tertinggi setelah penyakit *stroke*. Oleh karena itu perlu untuk melakukan pencegahan penyakit jantung sedini mungkin. Namun karena terkendala biaya, minim nya pengetahuan, dan fasilitas yang tidak memadai menyebabkan masyarakat umumnya tidak melakukan cek kesehatan.

Salah satu solusi untuk melakukan pencegahan penyakit jantung adalah dengan memprediksi seseorang terkena penyakit jantung dengan menggunakan pendekatan *machine learning* (ML). Dengan menggunakan pendekatan *machine learning* (ML) membantu masyarakat umum untuk sedikit mengetahui kesehatan jantung mereka melalui faktor-faktor yang diamati. Dengan demikian tujuan penelitian adalah untuk mengetahui dan memprediksi seseorang terkena penyakit jantung dan seseorang tidak terkena penyakit jantung.

Support Vector Machine

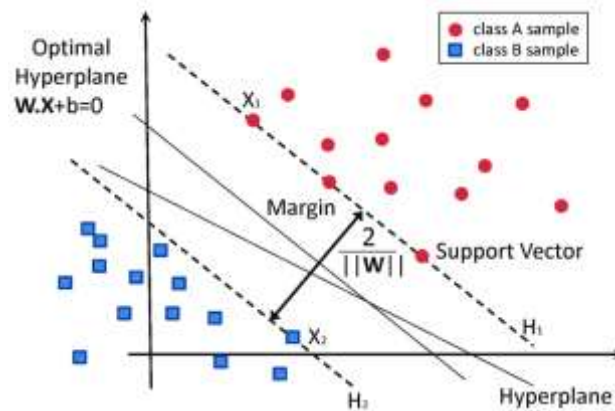
Analisis yang cocok digunakan pada penelitian ini adalah *Support Vector Machine* (SVM). *Support Vector Machine* merupakan metode dalam analisis statistika yang digunakan untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. (Santoso, 2007). Sehingga tujuan dari analisis dengan *Support Vector Machine* untuk mengklasifikasikan dan memprediksi seseorang terkena penyakit jantung atau tidak.

Salah satu kelebihan dari *Support Vector Machine* adalah baik digunakan untuk data yang tidak terlalu besar dan nilai akurasi yang diperoleh baik (Sutanto, 2021). Pada awalnya *Support Vector Machine* digunakan pada data yang variabel terikatnya merupakan kategorik biner yaitu variabel yang nilainya terdiri dari dua kelas (James, Witten , Hastie , & Tibshirani, 2013). Namun kemudian

Support Vector Machine dikembangkan kembali untuk klasifikasi variabel multikelas (Santoso, 2007)

a. *Support Vector Machine Linier*

Support Vector Machine memiliki prinsip dasar yaitu kasus klasifikasi yang secara linier dapat dipisahkan, namun *Support Vector Machine* telah dikembangkan agar dapat bekerja pada problem non-linier dengan memasukkan konsep kernel pada ruang kerja berdimensi tinggi. Pada *Support Vector Machine* metode klasifikasi untuk menentukan hyperplane dari variabel terikat dua kelas maupun lebih dari dua kelas kategori. Hyperplane digunakan sebagai bidang pemisah pada metode *Support Vector Machine*. Hyperplane yang baik adalah hyperplane yang memiliki ukuran margin terbesar. Berikut gambar hyperplane sebagai pembatas kelas.



Definisi matematika untuk dari hyperplane sederhana sebagai berikut:

$$\beta_0 + \beta_1 X_1 + \beta_3 X_2 = 0 \quad (1)$$

(James, Witten , Hastie , & Tibshirani, 2013)

Dalam melakukan klasifikasi dua kelas yaitu $y_i \in \{-1,1\}$ pada p-dimensional (sebanyak p variabel bebas) dan data dinotasikan $x \in \mathbb{R}^p$, maka hyperplane p-dimensional didefinisikan sebagai berikut:

$$wx + b = 0 \quad (2)$$

Sehingga diperoleh pattern masing-masing kelas sebagai berikut

$$wx + b \geq +1 \text{ jika } y = +1 \quad (3)$$

$$wx + b \leq -1 \text{ jika } y = -1 \quad (4)$$

Hyperplane (fungsi pemisah) terbaik terletak ditengah-tengah dua set objek kelas yaitu hyperplane dengan nilai margin terbesar. Margin terbesar dihitung dengan selisih persamaan hyperplane kedua kelas:

$$\begin{aligned} (wx + b = +1) - (wx + b = -1) \\ w(x_1 - x_2) = 2 \end{aligned} \quad (5)$$

Untuk menemukan hyperplane terbaik adalah dengan memaksimalkan margin atau jarak kedua objek support vector dari kelas yang berbeda dengan metode *Quadratic Progamming* (QP) *Problem* sebagai berikut:

$$\min_w \tau(w) = \frac{\|w\|^2}{2} \quad (6)$$

Subject to $y_i(wx_i + b) \geq +1, i = 1, \dots, \lambda$

Problem ini diselesaikan dengan *Lagrange Multiplier*

$$L(w, b, a) = \frac{\|w\|^2}{2} - \sum_{i=1}^l a_i (y_i(x_i w_i + b) - 1) \quad (7)$$

a_i adalah *Lagrange multipliers*, yang bernilai nol atau positif ($a_i \geq 0$). Nilai optimal dari Persamaan (6) dapat dihitung dengan meminimalkan L terhadap $w \rightarrow$ dan b, dan memaksimalkan L

terhadap a_i . Dengan memperhatikan sifat bahwa pada titik optimal gradient $L = 0$, persamaan (7) dapat dimodifikasi sebagai maksimisasi problem yang hanya mengandung a_i , sebagaimana persamaan (8) berikut:

Maksimasi:

$$\sum_{i=1}^l a_i - \frac{1}{2} \sum_{j=1}^l a_i a_j y_i y_j x_i x_j \quad (8)$$

Dengan *Constraint*:

$$\sum_{i=1}^l a_i y_i y_j = 0 \quad (9)$$

Dari hasil perhitungan ini diperoleh a_i yang kebanyakan bernilai positif. Data yang berkorelasi dengan a_i yang positif inilah yang disebut sebagai *support vector* (Vapnik, 1995). Penjelasan di atas berdasarkan asumsi bahwa kedua belah kelas dapat terpisah secara sempurna oleh *hyperplane* (*linear separable*). Akan tetapi, pada umumnya dua belah kelas pada *input space* tidak dapat terpisah secara sempurna (*non linear separable*). Hal ini menyebabkan constraint pada persamaan (8) tidak dapat terpenuhi, sehingga optimisasi tidak dapat dilakukan. Untuk mengatasi masalah ini, SVM dirumuskan ulang dengan memperkenalkan teknik *softmargin*. Dalam *softmargin*, Persamaan dimodifikasi dengan memasukkan *slackvariable* ξ ($\xi > 0$) sebagai berikut:

$$y_i(x_i w + b) \geq 1 - \xi_i \forall_i \quad (10)$$

$$\min_w \tau(w, \xi) = \frac{\|w\|^2}{2} + C \sum_{i=1}^l \xi_i \quad (11)$$

Parameter C dipilih untuk mengontrol *tradeoff* antara margin dan error klasifikasi ξ . Nilai C yang besar berarti akan memberikan *penalty* yang lebih besar terhadap error klasifikasi tersebut (Sembiring, 2007).

b. Support Vector Machine Non-Linear

Secara umum banyak dijumpai kasus klasifikasi dengan ketidaklinearan, cara mengatasinya adalah dengan metode kernel. Fungsi Kernel adalah suatu fungsi k yang mana untuk semua vector input x, z akan memenuhi kondisi

$$k(x, z) = \phi(x) T \phi(z) \quad (12)$$

dimana $\phi(\cdot)$ adalah fungsi pemetaan dari ruang input ke ruang fitur. Dengan kata lain, fungsi kernel adalah fungsi perkalian dalam (inner product) pada ruang fitur.

Berikut merupakan berbagai fungsi kernel yang terdapat pada SVM:

1. Fungsi linear, memiliki persamaan $K(x, y) = x \cdot y$
2. Fungsi polinomial of degree, memiliki persamaan $K(x, y) = (x \cdot y)^d$
3. Fungsi polinomial of degree 2, memiliki persamaan $K(x, y) = (x \cdot y + c)^d$
4. Fungsi Gaussian RBF ; memiliki persamaan $K(x, y) = \exp\left(\frac{-\|x-y\|^2}{2\sigma^2}\right)$

Keterangan :

$K(x, y)$	= fungsi kernel
x_i	= nilai data ke i
x_j	= nilai data ke j
c	= nilai konstanta c
σ	= sigma terkecil dari kernel RBF.

c. Testing Support Vector Machine

Secara umum ada dua proses yang dilakukan dalam klasifikasi, yaitu proses training dan proses testing. Pada proses training dibangun sebuah model atau fungsi dari training set yang tidak diketahui label-labelnya. Dan proses testing dilakukan untuk mengetahui keakuratan model atau fungsi yang akan dibangun pada proses training yang menggunakan testing set untuk memprediksi label-labelnya. (Munawarah, Soesanto, & Faisal, 2016).

Dalam proses testing digunakan fungsi keputusan. Fungsi keputusan dapat dihitung dengan rumus dibawah ini.

$$f(x) = \sum_{i=1}^m \alpha_i y_i K(x_i, x) + b \quad (13)$$

Dengan nilai *bias* (b),

$$b = -\frac{1}{2} [\sum_{i=1}^m \alpha_i y_i K(x_i, x^+) + \sum_{i=1}^m \alpha_i y_i K(x_i, x^-)] \quad (14)$$

Keterangan :

α_i	= alfa data ke i
y_i	= data nilai dari kelas ke i
m	= data jumlah dari SV
$K(x_i, x)$	= fungsi kernel
b	= nilai bias
σ	= sigma terkecil dari kernel RBF.
$K(x_i, x^+), K(x_i, x^-)$	= kernel nilai yang dipakai

(Pratama, Wihandika, & Ratnawati, 2018)

Evaluation

Tingkat akurasi pada model yang akan dihasilkan oleh proses peralihan dengan *Support Vector Machine* sangat bergantung terhadap fungsi kernel dan parameter yang digunakan (Parapat, Furqon , & Sutrisno, 2018). Dengan menggunakan confusion matriks dapat diketahui keakuratan yang dicapai dari model. Berikut adalah gambar dari confusion matriks.

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Nilai false negatif dan false positif merupakan nilai yang perlu diperhatikan karena dari nilai tersebut dapat dihitung nilai recall, recision, dan akurasi. False negatif (F.N) menunjukkan suatu kondisi negatif namun terdeteksi tidak negatif, sedangkan false positif (F.P) adalah keadaan dimana suatu data kelas positif terdeteksi sebagai data kelas negatif. Dapat diartikan bahwa keadaan model salah memprediksi atau mengelompokkan data. (Azhari, Situmorang, & Rosnelly, 2021)

Kemampuan model untuk mengklasifikasi diukur dengan tiga nilai berikut:

- 1) precision

$$precision = \left(\frac{TP}{TP+FP} \right) \quad (15)$$

- 2) recall

$$Recall = \left(\frac{TP}{TP+FN} \right) \quad (16)$$

3) akurasi

$$accuracy = \left(\frac{TP+TN}{TP+TN+FP+FN} \right) \quad (17)$$

(Azhari, Situmorang, & Rosnelly, 2021)

N-fold Cross Validation

N-fold cross validation atau *k-fold cross validation* merupakan salah satu metode yang digunakan untuk mengetahui rata-rata keberhasilan dari suatu sistem dengan cara melakukan perulangan dengan mengacak atribut masukan sehingga sistem tersebut teruji untuk beberapa atribut input yang acak *n-fold cross validation* diawali dengan membagi data sejumlah *n-fold* yang diinginkan (Pandi,2012).

Dalam iterasi *ke – i* partisi akan menjadi data testing dan sisanya akan menjadi data training. Untuk penggunaan jumlah fold terbaik untuk uji validitas, dianjurkan menggunakan 10-fold cross validation dalam model (Prasetya,2006).

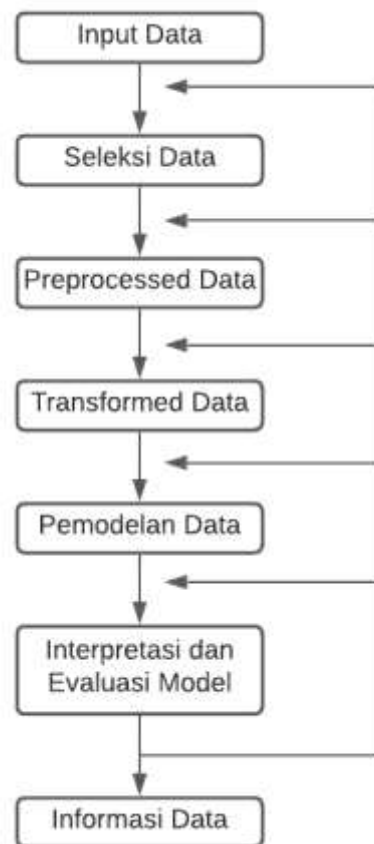
METODOLOGI

Dataset Penyakit Jantung berukuran 303×14 yang diperoleh dari database Cleveland (<https://archive.ics.uci.edu/ml/datasets.php>). Berikut variabel yang digunakan untuk memprediksi seseorang terkena penyakit jantung.

Tabel 1. Identifikasi Variabel Data Penyakit Jantung

No	Variabel	Keterangan
1	Umur	29 - 77 Tahun
2	Jenis Kelamin	0 : wanita, 1 : pria
3	Tipe Sakit Dada	1 : typical angina, 2 : atypical angina, 3 : non-anginal pain, 4: asymptomatic
4	Tekanan Darah	mmHg
5	Kolesterol	mg/dl
6	Kadar Gula Darah	1 : lebih dari 120 mg/dl, 2 : kurang dari 120 mg/dl
7	Grafik Elektrokardio	0 : normal, 1 : gelombang ST-T tidak normal dengan ST-T > 0.05, 2 : hipertrofi ventrikel kiri/pembesaran bilik kiri jantung
8	Detak Jantung Maksimal	
9	Keluhan Jantung Saat Olahraga	1 : ada, 2 : tidak ada
10	Oldpeak	1 : menanjak, 2 : mendatar, 3 : menurun
11	Slope	0 s.d 3
12	Jumlah Pembuluh Darah Utama	3 : normal, 6 : cacat permanen, 7 : cacat sementara
13	Thal	3 : normal, 6 : cacat permanen, 7 : cacat sementara
14	Diagnosa	0 : normal, 1 : penyakit jantung

Penelitian ini dilakukan untuk mengetahui performa terbaik dari *Support Vector Machine* dengan beberapa fungsi kernel.. Adapun tahapan melakukan pengolahan data prediksi penyakit jantung seseorang adalah sebagai berikut.



PEMBAHASAN

Eksplorasi Data

a. Statistik Deskriptif

```

> summary(data)
   age          sex          cp          trestbps          chol
Min.   :29.00   Min.   :0.0000   Min.   :0.000   Min.   : 94.0   Min.   :126.0
1st Qu.:47.50   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:120.0   1st Qu.:211.0
Median :55.00   Median :1.0000   Median :1.000   Median :130.0   Median :240.0
Mean   :54.37   Mean   :0.6832   Mean   :0.967   Mean   :131.6   Mean   :246.3
3rd Qu.:61.00   3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:140.0   3rd Qu.:274.5
Max.   :77.00   Max.   :1.0000   Max.   :3.000   Max.   :200.0   Max.   :564.0

   fbs          restecg          thalach          exang          oldpeak
Min.   :0.0000   Min.   :0.0000   Min.   : 71.0   Min.   :0.0000   Length:303
1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:133.5   1st Qu.:0.0000   Class :character
Median :0.0000   Median :1.0000   Median :153.0   Median :0.0000   Mode  :character
Mean   :0.1485   Mean   :0.5281   Mean   :149.6   Mean   :0.3267
3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:166.0   3rd Qu.:1.0000
Max.   :1.0000   Max.   :2.0000   Max.   :202.0   Max.   :1.0000

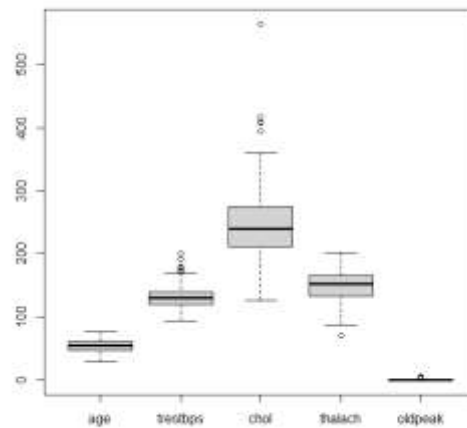
   slope          ca          thal          target
Min.   :0.000   Min.   :0.0000   Min.   :0.000   Min.   :0.0000
1st Qu.:1.000   1st Qu.:0.0000   1st Qu.:2.000   1st Qu.:0.0000
Median :1.000   Median :0.0000   Median :2.000   Median :1.0000
Mean   :1.399   Mean   :0.7294   Mean   :2.314   Mean   :0.5446
3rd Qu.:2.000   3rd Qu.:1.0000   3rd Qu.:3.000   3rd Qu.:1.0000
Max.   :2.000   Max.   :4.0000   Max.   :3.000   Max.   :1.0000
  
```

Kesimpulan: untuk data range data numerik normal dan satuan setiap amatan pada variabel yang sama adalah sama.

- b. Missing Value
> anyNA(data)
[1] FALSE

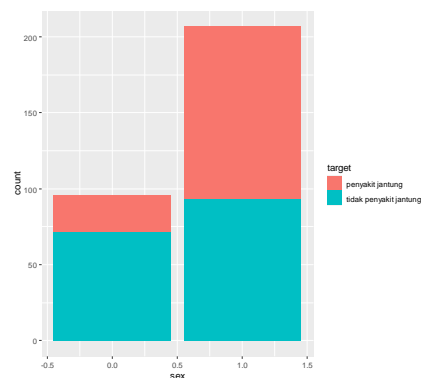
Kesimpulan: tidak ada missing value pada gugus data penyakit jantung.

- c. Outliers



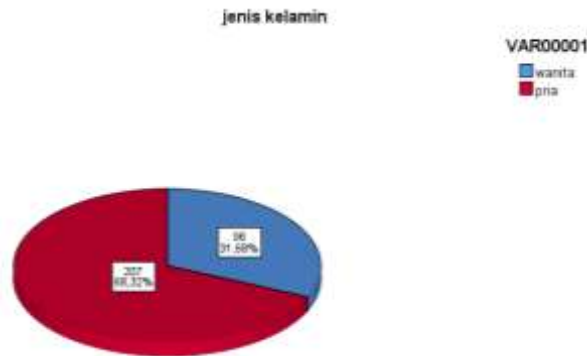
Kesimpulan: terdapat beberapa amatan yang terdeteksi sebagai outlier pada gugus data penyakit jantung diantaranya variabel trestbps, chol, dan peak.

Visualisasi Data



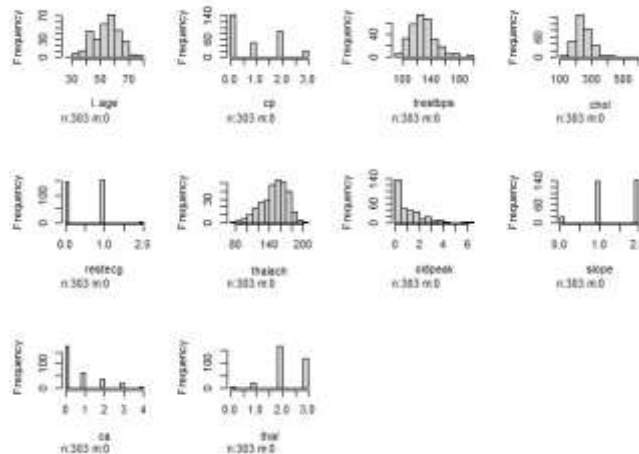
Kesimpulan:

Berdasarkan gambar di atas pasien yang di amati lebih banyak berjenis kelamin pria dibandingkan wanita. Lebih dari setengah pasien pria menderita penyakit jantung sedangkan pasien wanita lebih banyak tidak menderita penyakit jantung.



Kesimpulan:

Berdasarkan gambar di atas lebih dari setengah amatan berjenis kelamin pria dengan persentase 63,32% sedangkan sisanya 31,68% pasien berjenis kelamin wanita. Hal ini menunjukkan pasien dari gugus data di dominasi oleh pria.



Kesimpulan:

Berdasarkan gambar di atas umur pasien berdistribusi normal. Variabel tekanan darah pasien cenderung berada dikisaran 100 s.d. 140 mmHg. Sedangkan detak jantung maksimal pasien cenderung berada dikisaran 140 s.d. 200 detakan.

Preprocessing dan Seleksi Variabel

a. Split Data Train dan Test

Gugus data dipisah untuk memisahkan 80% data training dan 20% data test. Sebanyak 80% data digunakan untuk melakukan pemodelan kemudian sebanyak 20% data digunakan untuk mengetahui performa algoritma yang sudah dilatih dari data training.

b. Seleksi Variabel

Pada penelitian ini tidak dilakukan seleksi ataupun ekstraksi variabel karena jumlah variabel tidak tergolong banyak.

Pemodelan

Melakukan pemodelan pada data berdimensi tinggi dengan mencoba beberapa fungsi kernel. Parameter digunakan pada masing-masing fungsi kernel dipilih secara *trial and error*. Berikut adalah pemodelan dengan default parameter di R pada model linier, polynomial, radial, dan sigmoid.

1) Model Linier

```
> svm1 <- svm(target~., data = training_set, kernel = "linear")
> prediksi1 <- predict(svm1, training_set)
> confusionMatrix(prediksi1, training_set$target)
Confusion Matrix and Statistics
```

Prediction \ Reference	penyakit jantung	tidak penyakit jantung
penyakit jantung	91	11
tidak penyakit jantung	19	121

```

Accuracy : 0.876
95% CI : (0.8275, 0.9148)
No Information Rate : 0.5455
P-value [Acc > NIR] : <2e-16

Kappa : 0.7485

McNemar's Test P-value : 0.2012

Sensitivity : 0.8271
Specificity : 0.9167
Pos Pred Value : 0.8922
Neg Pred Value : 0.8643
Prevalence : 0.4545
Detection Rate : 0.3760
Detection Prevalence : 0.4215
Balanced Accuracy : 0.8720

'Positive' Class : penyakit jantung

```

Dengan menggunakan model linier diperoleh nilai akurasi sebesar 87,6%

2) Model Polynomial

```
> svm2 <- svm(target~., data = training_set, kernel = "polynomial")
> prediksi2 <- predict(svm2, training_set)
> confusionMatrix(prediksi2, training_set$target)
Confusion Matrix and Statistics
```

Prediction \ Reference	penyakit jantung	tidak penyakit jantung
penyakit jantung	97	4
tidak penyakit jantung	13	128

```

Accuracy : 0.9298
95% CI : (0.8899, 0.9585)
No Information Rate : 0.5455
P-value [Acc > NIR] : < 2e-16

Kappa : 0.8574

McNemar's Test P-value : 0.05235

Sensitivity : 0.8818
Specificity : 0.9697
Pos Pred Value : 0.9604
Neg Pred Value : 0.9078
Prevalence : 0.4545
Detection Rate : 0.4008
Detection Prevalence : 0.4174
Balanced Accuracy : 0.9258

'Positive' Class : penyakit jantung

```

Dengan menggunakan model polynomial dan parameter default dari R yaitu $\gamma = 1/14$; $\text{coef} = 0$; dan $\text{degree} = 3$ diperoleh hasil akurasi sebesar 92,98%

3) Model Radial

```
> confusionMatrix(prediksi3, training_set$target)
Confusion Matrix and Statistics
```

Prediction \ Reference	penyakit jantung	tidak penyakit jantung
penyakit jantung	84	9
tidak penyakit jantung	26	123

```

Accuracy : 0.8554
95% CI : (0.8046, 0.8971)
No Information Rate : 0.5455
P-value [Acc > NIR] : < 2.2e-16

Kappa : 0.7045

McNemar's Test P-value : 0.006841

Sensitivity : 0.7636
Specificity : 0.9318
Pos Pred Value : 0.9032
Neg Pred Value : 0.8255
Prevalence : 0.4545
Detection Rate : 0.3471
Detection Prevalence : 0.3843
Balanced Accuracy : 0.8477

'Positive' Class : penyakit jantung

```

Dengan menggunakan model polynomial dan parameter default dari R yaitu $\gamma = 1/14$ diperoleh hasil akurasi sebesar 85,54%

4) Model Sigmoid

```
> confusionMatrix(prediksi4, training_set$target)
Confusion Matrix and Statistics

              Reference
Prediction    penyakit jantung tidak penyakit jantung
penyakit jantung      83          9
tidak penyakit jantung 27         123

      Accuracy : 0.8512
    95% CI : (0.8, 0.8936)
  No Information Rate : 0.5455
    P-Value [Acc > NIR] : < 2.2e-16

      Kappa : 0.6959

  Mcnemar's Test P-Value : 0.004607

    Sensitivity : 0.7545
    Specificity : 0.9318
   Pos Pred Value : 0.9022
   Neg Pred Value : 0.8200
     Prevalence : 0.4545
   Detection Rate : 0.3430
   Detection Prevalence : 0.3802
   Balanced Accuracy : 0.8432

'Positive' Class : penyakit jantung
```

Dengan menggunakan model polynomial dan parameter default dari R yaitu $\gamma = 1/14$; $\text{coef} = 0$ diperoleh hasil akurasi sebesar 85,12%

Berdasarkan pada hasil akurasi beberapa model *Support Vector Machine* di atas, model polynomial memiliki tingkat akurasi tertinggi yaitu sebesar 92,98%. Oleh karena itu, fungsi kernel yang akan digunakan pada gugus data penyakit jantung adalah *Support Vector Machine* model polynomial. Selanjutnya menentukan nilai parameter yang paling optimum dari *Support Vector Machine* model polynomial secara *trial and error* . Penentuan parameter ini disebut dengan tuning parameter dan dengan memanfaatkan metode Cross Validation.

Evaluasi dan Interpretasi

a. Tuning Parameter dengan Cross Validation

```
> m = tune.svm(x = data[, -14], y = as.numeric(data$target),
+             tunecontrol = tune.control(cross = 5), cost = 1,
+             gamma = seq(0.4, by = 0.05))
> m$best.parameters
      gamma cost
2 0.05      1
> m$best.performance
[1] 0.1275155

> summary(m)
Parameter tuning of 'svm':
- sampling method: 5-fold cross validation

Best parameters:
  gamma cost
0.05      1

- best performance: 0.1275155

- Detailed performance results:
  gamma cost  error  dispersion
1 0.00      1 0.4173633 0.039626070
2 0.05      1 0.1275155 0.024326127
3 0.10      1 0.1348642 0.025362450
```

Berdasarkan output di atas parameter paling optimum diantaranya $\text{degree} = 2$; $\gamma = 0,05$; $\text{cost} = 1$ dengan nilai akurasi sebesar 92.98%.

b. Model Akhir

Setelah parameter optimal didapatkan, selanjutnya melakukan pemodelan *Support Vector Machine* dengan fungsi kernel Polynomial dengan parameter $\text{degree} = 2$; $\gamma = 0,05$; $\text{cost} = 1$ dan diperoleh nilai akurasi sebesar 92.98% seperti pada hasil output berikut.

```

> ##kernel polynomial setelah tacc 92,98%
> svm_poly <- svm(target~., data = training_set, kernel = "polynomial",
+               ranges = list(degree = 2, cost = 1, gamma=0.05))
> prediksi_poly <- predict(svm_poly, training_set)
> confusionMatrix(prediksi_poly, training_set$target)
Confusion Matrix and Statistics

              Reference
Prediction      penyakit jantung tidak penyakit jantung
penyakit jantung          97              4
tidak penyakit jantung    13             128

      Accuracy : 0.9298
      95% CI   : (0.8899, 0.9585)
    No Information Rate : 0.5455
      P-value [Acc > NIR] : < 2e-16

      kappa : 0.8574

McNemar's Test P-value : 0.05235

      Sensitivity : 0.8818
      Specificity : 0.9697
      Pos Pred value : 0.9604
      Neg Pred value : 0.9078
      Prevalence : 0.4545
      Detection Rate : 0.4008
      Detection Prevalence : 0.4174
      Balanced Accuracy : 0.9258

'Positive' class : penyakit jantung

```

c. Testing Model

Testing model dilakukan untuk mengukur kebaikan model saat dicobakan pada data test. Pengukuran ini dilakukan sebagai simulasi penggunaan model di dunia nyata.

```

> ###Testing model
> test_pred <- predict(svm_poly, newdata = test_set)
> ##Akurasi 78,69%
> confusionMatrix(table(test_pred, test_set$target))
Confusion Matrix and Statistics

test_pred      penyakit jantung tidak penyakit jantung
penyakit jantung          19              9
tidak penyakit jantung     9             29

      Accuracy : 0.7869
      95% CI   : (0.6652, 0.8814)
    No Information Rate : 0.541
      P-value [Acc > NIR] : 5.878e-05

      kappa : 0.565

McNemar's Test P-value : 0.2073

      Sensitivity : 0.6786
      Specificity : 0.8788
      Pos Pred value : 0.8261
      Neg Pred value : 0.7632
      Prevalence : 0.4590
      Detection Rate : 0.3115
      Detection Prevalence : 0.3770
      Balanced Accuracy : 0.7787

'Positive' class : penyakit jantung

```

Berdasarkan output di atas, nilai akurasi yang diperoleh dari model *Support Vector Machine* adalah 78,69% . Oleh karena itu model SVM polynomial dengan parameter degree = 2; gamma = 0,05; cost = 1

KESIMPULAN

Berdasarkan hasil penelitian yang telah dipaparkan sebelumnya, pengklasifikasian menggunakan metode *Support Vector Machine* berbagai fungsi kernel terbaik adalah polynomial dengan parameter gamma = 0,05 ; coef = 1; degree = 2; dan cost = 1. Hasil klasifikasi dengan *Support Vector Machine* dan parameter tersebut menunjukkan akurasi sebesar 78,69%. Hal ini menunjukkan model *Support Vector Machine* yang dibangun tergolong baik.

DAFTAR PUSTAKA

- Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes. *Jurnal Media Informatika Budidarma*, 640-651.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. USA: Springer Science + Business New York.
- Munawarah, R., Soesanto, O., & Faisal, M. (2016). Penerapan Metode Support Vector Machine Pada Diagnosa Hepatitis. *Kumpulan Jurnal Ilmu Komputer*, 4(1), 103-113.
- Parapat, I. M., Furqon, M. T., & Sutrisno. (2018). Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3163-3169.
- Pratama, A., Wihandika, R. C., & Ratnawati, D. E. (2018). Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(4), 1704-1708.
- Santoso, B. (2007). *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Sutanto, T. (2021). *(Supervised Learning) Classification Models II*. Retrieved from Taudata: <https://tau-data.id/slcm-02/>