

## KLASIFIKASI PENYAKIT HIPERTENSI MENGGUNAKAN METODE SVM GRID SEARCH DAN SVM GENETIC ALGORITHM (GA)

Fithroh Oktavi Awalullaili<sup>1\*</sup>, Dwi Ispriyanti<sup>2</sup>, Tatik Widiharih<sup>3</sup>

<sup>1,2,3</sup>Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

\*e-mail: [fithrohoktavi21@gmail.com](mailto:fithrohoktavi21@gmail.com)

DOI: 10.14710/j.gauss.11.4.488-498

### Article Info:

Received: 2022-10-20

Accepted: 2022-12-12

Available Online: 2023-02-25

### Keywords:

*Hypertention; Support Vector Machine; Grid Search; Genetic Algorithm*

**Abstract:** Hypertension is an abnormally high pressure that occurs inside the arteries. Hypertension increased by 8.3% from 2013 based on health research in 2018. Some of the factors that cause hypertension include gender, age, salt consumption, cigarette consumption, cholesterol levels and a family history of hypertension. The data in this study are data on normal and hypertensive patients at the Padangsari Health Center for the period of July – December 2021. This study will classify blood pressure with the aim of obtaining the results of the accuracy of the classification of the methods used. The method used in this study is a *support vector machine* (SVM). SVM is a well-known algorithm, producing optimal solutions to classification problems. SVM uses kernel functions for *separable nonlinear* data. The displacement kernels used in this study are linear and RBF. SVM has the disadvantage of determining the best parameters, to overcome these weaknesses developed the method of finding the best parameters. The search for the parameters of this study used *grid search* and *genetic algorithm* (GA). *Grid search* has the advantage of producing parameters that are close to the optimal value, while GA has the advantage of being easy to find global optimum values. This study will compare the classification results of the SVM *grid search* and SVM GA methods. The results of this study obtained the method that has the best accuracy, namely SVM *grid search* using a *radial base function* (RBF) kernel with an accuracy of 89.22%.

## 1. PENDAHULUAN

Hipertensi merupakan kondisi tekanan darah yang abnormal tinggi terjadi di dalam arteri. Faktor penyebab terjadinya hipertensi terdiri dari jenis kelamin, usia, konsumsi garam, kadar kolestrol, konsumsi rokok, dan riwayat hipertensi dalam keluarga. Klasifikasi tekanan darah pada umumnya dibagi menjadi tiga yaitu tekanan darah rendah, normal dan tekanan darah tinggi. Penelitian ini akan mengklasifikasikan kasus tekanan darah menjadi dua yaitu pasien dengan tekanan darah normal dan hipertensi. Metode statistika diperlukan untuk mendapatkan klasifikasi pasien dengan tekanan darah normal dan pasien hipertensi. *Support vector machine* (SVM) adalah salah satu metode klasifikasi yang dapat memisahkan data menjadi dua kelas yaitu kelas positif dan kelas negatif (Darmawan, 2015). Kelemahan dari metode SVM yaitu terdapat kesulitan dalam menentukan nilai parameter yang optimal, sehingga hal tersebut dapat mempengaruhi performa kinerja dari SVM (Yanaeng, *et al.*, 2014). Kesulitan mendapatkan parameter yang optimal dapat diatasi dengan mengoptimalkan nilai parameter. Metode yang digunakan dalam pencarian parameter optimal pada penelitian ini adalah *tuning hyperparameter* menggunakan *Grid Search* dan *Genetic algorithm* (GA).

*Tuning hyperparameter* menggunakan *grid search* akan menghasilkan parameter yang lebih mendekati nilai optimal dibandingkan dengan penggunaan *random search* (Agrawal, 2021) *Genetic algorithm* (GA) merupakan suatu algoritma yang biasanya digunakan untuk mencari solusi optimal untuk berbagai permasalahan yang sulit, misalnya

masalah optimasi (Haryanto dan Hidayatullah, 2016). Kelebihan GA yaitu memiliki ruang solusi yang luas, mudah ditemukannya nilai optimum global, mudah dimodifikasi untuk berbagai masalah, hanya menggunakan fungsi evaluasi dan mampu menangani ruang pencarian yang besar dan sulit dipahami dengan mudah (Sivanandam dan Deepa, 2008). Penelitian mengenai GA telah dilakukan oleh (Haryanto dan Hidayatullah, 2016), penelitian tersebut membandingkan SVM dan SVM dengan optimasi GA dan diperoleh hasil bahwa SVM GA mampu memberikan peningkatan akurasi sebesar 2,53% lebih baik dibandingkan dengan SVM, akurasi yang dihasilkan sebesar 89,06%. Pada penelitian ini penulis membandingkan metode SVM *grid search* dan SVM GA dengan membandingkan nilai akurasi dan AUC nya.

## 2. TINJAUAN PUSTAKA

Hipertensi secara umum merupakan kejadian tanpa gejala, tekanan abnormal tinggi di dalam arteri sehingga dapat menyebabkan meningkatnya risiko terhadap penyakit berbahaya seperti stroke, aneurisma, gagal jantung, serangan jantung, dan kerusakan ginjal (Ilyas, 2016). Tekanan darah dikatakan tinggi apabila tekanan sistolik mencapai 140 mmHg atau lebih, dan tekanan diastolik mencapai 90 mmHg pada pemeriksaan berulang (Soenarta, *et al.*, 2015). Faktor penyebab hipertensi terbagi menjadi dua yaitu faktor genetik dan faktor non genetik. Faktor genetik yang mempengaruhi hipertensi adalah riwayat hipertensi dalam keluarga dan faktor non genetik yang menyebabkan hipertensi diantaranya jenis kelamin, kebiasaan merokok, usia, pola asupan garam (Nuraini, 2015). Selain itu hipertensi juga disebabkan oleh kadar kolestrol (Nugroho dan Fahrurrozi, 2018). Langkah pertama yang dilakukan pada penelitian ini adalah *pre-processing* data.

*Pre-processing* merupakan tahapan awal yang akan mentransformasikan data lama menjadi data baru dengan format yang sesuai dan siap untuk diproses (Setyohadi, *et al.*, 2017). Tujuan *preprocessing* dalam data mining adalah mentransformasi data ke suatu format yang prosesnya lebih mudah dan efektif sesuai dengan kebutuhan (Meilina, 2015). Salah satu metode transformasi data yaitu *scaling* atau normalisasi dimana data akan diubah dalam *range* yang lebih kecil seperti [-1,1] atau [0,1]. Normalisasi yang digunakan dalam penelitian ini yaitu *min max normalization*, data akan di transformasi secara linier kedalam *range* [0,1] menggunakan persamaan (1) (Han, *et al.*, 2012)

$$x_{new} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

dengan  $x_{new}$  adalah data hasil transformasi,  $x$  adalah data awal,  $\min(x)$  adalah nilai minimum pada data, dan  $\max(x)$  adalah nilai maksimum pada data.

*Support vector machine* (SVM) merupakan salah satu algoritma yang dikenal baik dalam menghasilkan solusi yang optimal untuk permasalahan klasifikasi, diperkenalkan oleh Vapnik sebagai *machine learning method* berbasis kernel untuk masalah klasifikasi dan regresi (Aulia, *et al.*, 2021). Tujuan dari *support vector machine* (SVM) adalah untuk menemukan *hyperplane* yang paling baik pada ruang fitur berdimensi tinggi (Cristianini dan Shawe-Taylor, 2000). *Hyperplane* terbaik pada SVM adalah *hyperplane* yang berada di tengah – tengah antara dua set objek dari dua kelas berbeda, mencari *hyperplane* terbaik sama saja dengan memaksimalkan margin atau jarak antara dua objek dari dua kelas yang berbeda (Santosa, 2007). Klasifikasi SVM dapat dibagi menjadi dua yaitu SVM Linier dan SVM Non linier.

Data *training* input pada SVM berdimensi  $m$ , dengan  $m$  merupakan jumlah variabel bebas, variabel bebas  $x_i (i = 1, 2, \dots, n)$  dan  $n$  menyatakan jumlah data *training*. Data *training* terbagi menjadi dua kelas diberi label  $y_i = 1$  untuk kelas 1 dan  $y_i = -1$  untuk kelas

2. Apabila data dapat dipisahkan secara linier, maka garis pemisah/*hyperplane* adalah persamaan (2) (Abe, 2005)

$$D(x) = \mathbf{w}^T \mathbf{x} + b \quad (2)$$

dengan  $\mathbf{w}^T$  adalah vektor parameter bobot dan  $b$  adalah bias atau *error* yang akan dicari nilainya. Inti klasifikasi menggunakan SVM akan melakukan klasifikasi dengan pemisahan kelas yang sesuai dengan pertidaksamaan (3) (Abe, 2005)

$$\mathbf{w}^T \mathbf{x}_i + b \begin{cases} > 0 & \text{untuk } y_i = 1, \\ \leq 0 & \text{untuk } y_i = -1, \end{cases} \quad (3)$$

Data *training* karena dapat dipisahkan secara linier, artinya tidak ada data *training* yang sempurna tepat berada pada persamaan *hyperplane*  $\mathbf{w}^T \mathbf{x} + b = 0$ . Syarat pemisahan kelas pada persamaan (3) adalah persamaan (4) (Abe, 2005)

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, i = 1, 2, \dots, n \quad (4)$$

Jarak terdekat antara data  $\mathbf{x}$  dengan *hyperplane* pada kelas 1 dan 2 masing – masing adalah  $\frac{1}{\|\mathbf{w}\|}$ , maka nilai margin diperoleh sebesar  $\frac{2}{\|\mathbf{w}\|}$ . *Hyperplane* terbaik dicari dengan rumus *Quadratic Programming* pada SVM untuk optimasi klasifikasi SVM linier seperti pada persamaan (5) (Santosa, 2007)

$$\min \frac{1}{2} \|\mathbf{w}\|^2 \quad (5)$$

Solusi dari permasalahan persamaan kuadrat dapat diselesaikan dengan fungsi *Lagrange Multipliers* adalah persamaan (6) (Abe, 2005).

$$L_p(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1] \quad (6)$$

dengan  $\boldsymbol{\alpha}_i = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$  dan  $\alpha_i$  adalah fungsi *Lagrange Multipliers* yang bernilai positif ( $\alpha_i \geq 0$ ). Permasalahan tersebut adalah permasalahan primal yang harus diubah kedalam bentuk persamaan *dual*.

$$\frac{\partial}{\partial \mathbf{w}} L_p(\mathbf{w}, b, \boldsymbol{\alpha}) = 0 \rightarrow \mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i \quad (7)$$

$$\frac{\partial}{\partial b} L_p(\mathbf{w}, b, \boldsymbol{\alpha}) = 0 \rightarrow \sum_{i=1}^n \alpha_i y_i = 0 \quad (8)$$

dengan menstutitusikan persamaan (6), (7) dan (8) akan diperoleh persamaan *dual* adalah

$$\text{Max } L_d(\boldsymbol{\alpha}) = \text{Max} \left( \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \alpha_i \alpha_j y_i y_j (\mathbf{x}_i^T \mathbf{x}_j) \right) \quad (9)$$

terhadap  $\alpha_i$  dengan fungsi batasan  $\sum_{i=1}^n \alpha_i y_i = 0, \alpha_i \geq 0$  untuk  $i = 1, 2, \dots, n$ .

Klasifikasi SVM *nonlinier* terjadi apabila data tidak dapat dipisahkan secara linier. Kasus data yang tidak linier, teknik *soft margin* dilakukan dengan memodifikasi persamaan (4) dengan memasukkan variabel *slack* positif ( $\xi_i \geq 0$ ), sehingga diperoleh persamaan (Abe, 2005):

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, i = 1, 2, \dots, n \quad (10)$$

*Hyperplane* yang optimal diperoleh dengan meminimumkan persamaan (11) (Abe, 2005)

$$Q(\mathbf{w}, b, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \quad (11)$$

yang memenuhi persamaan (16), dengan  $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$  dan C adalah parameter yang ditentukan untuk mengontrol *trade off* antara margin dengan kesalahan klasifikasi  $\xi$  (Abe, 2005). Persamaan primal yang terbentuk seperti pada persamaan (12).

$$L_p(\mathbf{w}, b, \xi, \alpha, \beta) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i (y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i) - \sum_{i=1}^n \beta_i \xi_i \quad (12)$$

Persamaan *dual* yang terbentuk seperti pada persamaan (13).

$$\text{Max } L_d = \text{Max} \left( \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \alpha_i \alpha_j y_i y_j (\mathbf{x}_i^T \mathbf{x}_j) \right) \quad (13)$$

*Hyperplane* optimal pada SVM ditentukan untuk memaksimalkan kemampuan generalisasi, tetapi apabila data *training* tidak dapat dipisahkan secara linier, pengklasifikasi memungkinkan tidak memiliki kemampuan generalisasi yang tinggi walaupun *hyperplane* ditentukan secara optimal, sehingga untuk meningkatkan keterpisahan secara linier ruang input asli dipetakan menjadi produk titik berdimensi tinggi yang disebut dengan ruang fitur (Abe, 2005). Jenis kernel yang digunakan dalam penelitian ini yaitu:

1) Kernel linier =  $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j)$

2) Kernel RBF =  $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$

dengan  $\mathbf{x}_i$  dan  $\mathbf{x}_j$  merupakan pasangan dua data *training* dan  $\gamma$  merupakan suatu parameter positif.

Hasil klasifikasi dari data x diperoleh dari persamaan (14):

$$f(\Phi(x)) = \text{sign} \left( \sum_{i=1}^{ns} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_j) + b \right) \quad (14)$$

$$f(\Phi(x)) = \begin{cases} 1, & \text{jika } \sum_{i=1}^{ns} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_j) + b > 0 \\ -1, & \text{jika } \sum_{i=1}^{ns} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_j) + b < 0 \end{cases}$$

dengan  $f(\Phi(x))$  adalah hasil klasifikasi dari data ke x,  $y_i$  adalah kelas data,  $\alpha_i$  adalah koefisien *lagrange*,  $K(\mathbf{x}_i, \mathbf{x}_j)$  adalah fungsi kernel data *testing* dan data *training*,  $b$  adalah bias dan  $ns$  adalah banyaknya *support vector*. Pencarian parameter pada penelitian ini menggunakan *grid search* dan *genetic algorithm*.

*Grid search* mengoptimasi parameter SVM ( $C, \gamma, \text{degree}, \text{etc}$ ) menggunakan teknik *cross validation* (CV) sebagai performansi matriks (Syarif, et al., 2016). Algoritma *grid search* bertujuan mengidentifikasi parameter optimal dalam data *training*, sehingga model tersebut mampu secara akurat memprediksi data *testing* (Purnama, 2020). *K-Fold cross validation* merupakan salah satu teknik untuk mengevaluasi kinerja sebuah model, dalam *k-fold cross validation* sebuah data (D) secara acak dibagi menjadi  $k$  *subsets* data (*folds*), yaitu

$D_1, D_2, \dots, D_k$  dengan ukuran yang sama (Han, *et al.*, 2012). Metode *K-Fold cross validation* membagi data *training* dan data *testing* sebanyak  $k$  bagian data dan memiliki fungsi supaya tidak ada *overlapping* pada data *testing* (Sasongko, 2016). *Genetic algorithm* (GA) merupakan salah satu metode *optimasi* yang pertama kali ditemukan oleh John Holland pada tahun 1975 (Sivanandam & Deepa, 2008). Konsep *Genetic algorithm* (GA) diselesaikan dengan proses evolusi dan menghasilkan solusi yang paling baik (Yanaeng, *et al.*, 2014). Dua operator utama dalam *genetic algorithm* menurut (Gorunescu, 2011) adalah *crossover* dan mutasi. *Crossover* merupakan proses menukar bagian dari solusi (kromosom) dengan bagian “parent” lain untuk menghasilkan jenis kromosom yang berbeda (*offspring* atau kromosom anak) yang mungkin menjadi solusi baru untuk menyelesaikan permasalahan dan Mutasi merupakan pergantian salah satu bagian solusi yang dipilih secara acak. Kontrol parameter diperlukan untuk mengendalikan kinerja GA dalam memecahkan masalah, ada dua parameter kontrol dalam genetic algoritma yaitu probabilitas *crossover* ( $P_c$ ) dan probabilitas mutasi ( $P_m$ ) (Windarto & Wijaya, 2012). Akurasi klasifikasi merupakan ketepatan klasifikasi yang menunjukkan performansi teknik klasifikasi secara keseluruhan, evaluasi klasifikasi terdiri dari akurasi, *sensitivity* (*recall*), dan *spesifisitas* (Han, *et al.*, 2012).

Tabel 1. *Confusion matrix*

Kondisi yang sebenarnya	Hasil Prediksi		
	Positif	Negatif	Total Baris
Positif	TP	FN	TP+FN
Negatif	FP	TN	FP+TN
Total Kolom	TP+FP	FN+TN	$N = TP+TN+FP+FN$

Akurasi dapat dihitung dengan rumus  $Akurasi = \frac{TN+TP}{TN+TP+FN+FP}$ , sensitivitas dihitung dengan  $Sensitivitas = \frac{TP}{TP+FN}$ , spesifisitas dihitung dengan rumus  $Spesifisitas = \frac{TN}{TN+FP}$  dan nilai AUC dihitung dengan rumus  $AUC = \frac{1}{2}(sensitivitas + spesifisitas)$ . Nilai AUC semakin mendekati satu hasilnya semakin baik.

### 3. METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data penyakit hipertensi yang diperoleh dari rekam medis Puskesmas Padangsari, Kecamatan Banyumanik Kota Semarang, Jawa Tengah. Jumlah data yang diambil yaitu 510 pasien penderita hipertensi yang terdiri dari pasien dengan tekanan darah normal dan pasien hipertensi. Variabel yang digunakan dalam penelitian ini terbagi menjadi dua variabel yaitu variabel dependen berupa status pasien hipertensi (normal dan hipertensi), sedangkan variabel independen terdapat enam variabel diantaranya jenis kelamin ( $x_1$ ), usia ( $x_2$ ), konsumsi garam berlebih ( $x_3$ ), kolesterol tinggi ( $x_4$ ), konsumsi rokok ( $x_5$ ), dan riwayat hipertensi dalam keluarga ( $x_6$ ). Langkah analisis pada penelitian ini adalah

- 1) Mempersiapkan data klasifikasi menggunakan data pasien hipertensi yang dibagi menjadi variabel dependen dan variabel independen.
- 2) Melakukan *preprocessing* pada data pasien hipertensi.
- 3) Menentukan fungsi kernel yang digunakan yaitu kernel linier dan kernel *radial basis function* (RBF).
- 4) Menentukan *range* dari tiap parameter, kernel linier dengan nilai parameter *cost* (C) dan kernel *Radial Basis Function* (RBF) dengan nilai parameter  $\gamma$  dan *cost* (C).
- 5) Membagi data menjadi data *training* dan *testing*

- 6) Melakukan pencarian parameter terbaik SVM dengan *tuning hyperparameter* menggunakan *grid search*.
- 7) Melakukan klasifikasi menggunakan metode SVM *grid search*.
- 8) Melakukan proses GA dan melakukan klasifikasi SVM GA.
- 9) Membandingkan nilai Akurasi dan nilai AUC dari SVM *grid search* dan SVM GA.
- 10) Mendapatkan hasil klasifikasi dari kedua metode yang paling baik.

#### 4. PEMBAHASAN

*Pre-processing* data dilakukan untuk menghasilkan bentuk data yang siap di proses. Pada penelitian ini dilakukan transformasi pada variabel usia. Transformasi data dilakukan menggunakan *min-max normalization*, nilai pada variabel yang ditransformasi akan berada pada *range* [0,1]. Kernel yang di gunakan untuk klasifikasi SVM linier adalah kernel linier dengan nilai parameter C yang di gunakan yaitu pada *range* ( $2^{-1} - 2^3$ ), ( $2^3 - 2^7$ ) dan ( $2^7 - 2^{11}$ ). Kernel yang digunakan untuk klasifikasi SVM *non* linier adalah kernel RBF. Parameter pada kernel RBF terdiri dari C dan  $\gamma$ . Nilai parameter C dan  $\gamma$  yang digunakan merupakan kombinasi dari *range* C ( $2^{-1} - 2^3$ ), ( $2^3 - 2^7$ ) dan ( $2^7 - 2^{11}$ ) dan parameter  $\gamma$  yaitu ( $2^{-9} - 2^{-3}$ ) dan ( $2^{-3} - 2^3$ ). Data *training* dan data *testing* di bagi menjadi dua dengan perbandingan 80% dan 20%. Data *training* dihasilkan dari pengambilan random pada *software R* sebanyak 408 dan data *testing* sebanyak 102. Selanjutnya dilakukan pencarian *range* parameter terbaik menggunakan *grid search*. Nilai *range* tiap parameter C diujikan menggunakan bantuan *software R* untuk mencari *range* parameter terbaik dan mendapatkan rata – rata nilai akurasi dari tiap *range* tersebut. Hasil rata – rata akurasi dari *range* parameter C pada kernel Linier seperti pada Tabel 2.

Tabel 2. Hasil rata - rata akurasi SVM kernel linier

<b>Range Parameter C</b>	<b>Rata – rata akurasi (%)</b>
$2^{-1} - 2^3$	88,94
$2^3 - 2^7$	89,07
<b><math>2^7 - 2^{11}</math></b>	<b>89,26</b>

Berdasarkan Tabel 2. dapat diketahui bahwa *range* nilai parameter SVM optimal ketika nilai C berada pada *range*  $2^7 - 2^{11}$  dengan nilai rata – rata akurasi paling tinggi. *Range* parameter tersebut selanjutnya akan digunakan dalam pencarian parameter menggunakan GA.

Pada SVM *non* linier, nilai *range* tiap parameter C dan  $\gamma$  di ujikan menggunakan bantuan *software R* untuk mendapatkan rata – rata nilai akurasi dari tiap *range* tersebut. Hasil rata – rata akurasi dari kombinasi parameter C dan parameter  $\gamma$  pada kernel RBF seperti pada Tabel 3.

Tabel 3. Hasil rata - rata akurasi SVM dengan Kernel RBF

<b>Range Parameter</b>		<b>Rata – rata akurasi (%)</b>
<b>C</b>	<b><math>\gamma</math></b>	
$2^{-1} - 2^3$	$2^{-9} - 2^{-3}$	89,02
$2^3 - 2^7$	$2^{-9} - 2^{-3}$	89,14
$2^7 - 2^{11}$	$2^{-9} - 2^{-3}$	89,31
<b><math>2^{-1} - 2^3</math></b>	<b><math>2^{-3} - 2^3</math></b>	<b>89,46</b>
$2^3 - 2^7$	$2^{-3} - 2^3$	89,31
$2^7 - 2^{11}$	$2^{-3} - 2^3$	89,29

Berdasarkan Tabel 3. dapat diketahui bahwa *range* nilai parameter SVM optimal ketika parameter  $C = 2^{-1} - 2^3$  dan  $\gamma = 2^{-3} - 2^3$ . Rata - rata akurasi yang diperoleh pada *range* tersebut merupakan *range* dengan nilai akurasi paling tinggi.

Pada Klasifikasi SVM linier, klasifikasi SVM *Grid search* kernel linier dilakukan menggunakan parameter terbaik yang diperoleh dari nilai *range* yang paling baik. C yang di dapatkan yaitu  $C = 512$ . Hasil dari *confusion matrix* adalah seperti pada Tabel 4.

Tabel 4. *Confusion matrix* SVM *Grid Search* kernel linier

Aktual	Prediksi	
	Hipertensi	Normal
Hipertensi	40	5
Normal	7	50

Berdasarkan Tabel 4. dapat diketahui bahwa pasien yang menderita hipertensi diprediksi hipertensi berjumlah 40 orang, pasien normal diprediksi normal berjumlah 50 orang, pasien hipertensi di prediksi normal berjumlah 5 orang dan pasien normal di prediksi hipertensi berjumlah 7 orang. Hasil perhitungan nilai akurasi dan nilai AUC dari tabel *confusion matrix* dari SVM *Grid Search* dengan kernel linier seperti pada Tabel 5.

Tabel 5. Hasil Perhitungan *Confusion Matrix* SVM *Grid Search* dengan kernel linier

Ukuran Performa Klasifikasi	Nilai (%)
Akurasi	88,24
AUC	88,30

Berdasarkan Tabel 5. akurasi yang diperoleh adalah 88,24%, artinya dengan menggunakan parameter  $C=512$  model SVM *Grid Search* dengan kernel linier dapat mengklasifikasikan 88,24% pengamatan dengan benar. Nilai AUC yang diperoleh adalah 89,42% artinya bahwa model SVM dengan kernel linier dapat mengklasifikasikan data dengan sangat baik. Klasifikasi SVM dengan kernel RBF dilakukan menggunakan parameter terbaik yang diperoleh dari nilai *range* yang paling baik hasil pencarian *tuning hyperparameter* menggunakan *grid search*. C dan  $\gamma$  yang di dapatkan yaitu  $C = 0,5$  dan  $\gamma = 1$ . Hasil dari *confusion matrix* seperti pada Tabel 6.

Tabel 6. *Confusion matrix* SVM *Grid Search* kernel RBF

Aktual	Prediksi	
	Hipertensi	Normal
Hipertensi	41	4
Normal	7	50

Berdasarkan Tabel 6. didapatkan hasil bahwa pasien yang menderita hipertensi diprediksi hipertensi berjumlah 41 orang, pasien normal diprediksi normal berjumlah 50 orang, pasien hipertensi di prediksi normal berjumlah 4 orang dan pasien normal di prediksi hipertensi berjumlah 7 orang. Hasil perhitungan nilai akurasi dan AUC dari tabel *confusion matrix* SVM dengan kernel RBF seperti pada Tabel 7.

Tabel 7. Hasil Perhitungan *Confusion Matrix* SVM *Grid Search* dengan kernel RBF

Ukuran Performa	Nilai
-----------------	-------

Klasifikasi	(%)
Akurasi	89,22
AUC	89,42

Berdasarkan Tabel 7. akurasi yang diperoleh adalah 89,22%, artinya dengan menggunakan parameter  $C = 0,5$  dan  $\gamma = 1$  model SVM kernel RBF dapat mengklasifikasikan 89,22% pemangatan dengan benar. Nilai AUC sebesar 89,42% dapat menunjukkan bahwa, model SVM kernel RBF dapat mengklasifikasikan data hipertensi dengan sangat baik. Selanjutnya dilakukan proses pencarian parameter menggunakan GA. *Range* parameter yang digunakan pada proses GA kernel linier adalah *range* parameter terbaik yang dihasilkan pada *tuning hyperparameter* menggunakan *grid search*. Hasil 10 kali percobaan menggunakan GA kernel linier seperti pada Tabel 8.

Tabel 8. Hasil 10 percobaan dengan SVM GA kernel linier

Percobaan ke-	Parameter C	Akurasi (%)
1	<b>919,3715</b>	<b>89,71</b>
2	1131,1234	89,46
3	990,7888	89,46
4	1033,5954	89,46
5	<b>848,7002</b>	<b>89,71</b>
6	<b>846,2192</b>	<b>89,71</b>
7	<b>968,6003</b>	<b>89,71</b>
8	1676,9884	89,46
9	1042,8135	89,46
10	879.7409	89,46

Berdasarkan Tabel 8. didapatkan hasil bahwa dari 10 percobaan yang dilakukan pada data dengan GA, akurasi tertinggi adalah 89,71%. Dari percobaan tersebut, maka didapatkan parameter C memiliki empat nilai parameter yang sama karena memiliki akurasi tertinggi yang sama yaitu 919,3715; 848,7002; 846,2192; dan 968,6003. Selanjutnya dilakukan pembentukan model SVM GA kernel linier dengan salah satu nilai C karena tidak ada satu nilai parameter C yang paling baik, dan C yang digunakan adalah 846,2192.

*Range* parameter yang digunakan pada proses GA kernel RBF adalah *range* parameter terbaik yang dihasilkan pada *tuning hyperparameter* menggunakan *grid search*. Hasil 10 kali percobaan menggunakan GA kernel RBF seperti pada Tabel 9.

Tabel 9. Hasil akurasi dengan SVM GA kernel RBF

Percobaan ke-	Parameter		Akurasi (%)
	C	$\gamma$	
1	4,446401	4,971193	89,46
2	5,067163	5,288036	89,46
3	4,227403	4,480895	89,71
4	4,490806	4,020896	89,46
5	4,756970	5,102139	89,71
6	3.854353	3,940951	89,71
7	5,009153	4,464202	89,71
8	4,070082	4,817321	89,71
9	4,512878	3,393532	89,71
10	4,393968	6,772513	89,71

Berdasarkan Tabel 9. dapat diketahui bahwa dari 10 percobaan yang dilakukan pada data *training*, akurasi tertinggi 89,71%. Dari percobaan tersebut, parameter optimal tidak

didapatkan karena tidak ada satu kombinasi  $C$  dan  $\gamma$  yang memiliki akurasi paling baik. Klasifikasi SVM GA kernel linier dilakukan menggunakan salah satu parameter yang diperoleh dari pencarian parameter menggunakan GA.  $C$  yang di dapatkan yaitu  $C = 846,2192$ . Hasil dari *confusion matrix* seperti pada Tabel 10.

Tabel 10. *Confusion Matrix* SVM dengan kernel linier

Aktual	Prediksi	
	Hipertensi	Normal
Hipertensi	40	5
Normal	7	50

Berdasarkan Tabel 10. dapat diperoleh kesimpulan bahwa pasien yang menderita hipertensi diprediksi hipertensi berjumlah 40 orang, pasien normal diprediksi normal berjumlah 50 orang, pasien hipertensi di prediksi normal berjumlah 5 orang dan pasien normal di prediksi hipertensi berjumlah 7 orang. Hasil perhitungan nilai akurasi dan nilai AUC dari tabel *confusion matrix* dari SVM GA dengan kernel linier seperti pada Tabel 11.

Tabel 11. Hasil Perhitungan *Confusion Matrix* SVM dengan kernel linier

Ukuran Performa Klasifikasi	Nilai (%)
Akurasi	88,24
AUC	88,30

Berdasarkan Tabel 11. akurasi yang diperleh adalah 88,24%, artinya dengan menggunakan parameter  $C = 846,2192$  model SVM GA dengan kernel linier dapat mengklasifikasikan 88,24% pemangatan dengan benar. Nilai AUC yang diperoleh adalah 88,30% artinya bahwa model SVM dengan kernel linier dapat mengklasifikasikan data dengan sangat baik.

Klasifikasi SVM GA RBF dilakukan menggunakan salah satu nilai parameter yang diperoleh dari pencarian parameter menggunakan GA.  $C$  yang digunakan yaitu  $C = 4,393968$  dan  $\gamma = 6,77251$ . Hasil dari *confusion matrix* seperti pada Tabel 12.

Tabel 12. *Confusion Matrix* SVM GA kernel RBF

Aktual	Prediksi	
	Hipertensi	Normal
Hipertensi	40	5
Normal	7	50

Berdasarkan Tabel 12. didapatkan hasil bahwa pasien yang menderita hipertensi diprediksi hipertensi berjumlah 40 orang, pasien normal diprediksi normal berjumlah 50 orang, pasien hipertensi di prediksi normal berjumlah 5 orang dan pasien normal di prediksi hipertensi berjumlah 7 orang. Hasil perhitungan nilai akurasi dan AUC dari tabel *confusion matrix* dari SVM GA dengan kernel RBF seperti pada Tabel 13.

Tabel 13. Hasil Perhitungan *Confusion Matrix* SVM GA dengan kernel RBF

Ukuran Performa Klasifikasi	Nilai (%)
Akurasi	88,24
AUC	88,30

Berdasarkan Tabel 13. akurasi yang diperoleh adalah 88,24%, artinya dengan menggunakan parameter  $C = 4,393968$  dan  $\gamma = 6,77251$  model SVM kernel RBF dapat mengklasifikasikan 88,24% pemangatan dengan benar. Nilai AUC sebesar 88,30% dapat menunjukkan bahwa, model SVM kernel RBF dapat mengklasifikasikan data hipertensi dengan sangat baik. Selanjutnya akan dibandingkan nilai akurasi dan nilai AUC dari kedua metode tersebut. Hasil dari perbandingan nilai akurasi dan nilai AUC dari kedua metode menggunakan kernel linier dan RBF adalah sebagai berikut.

Tabel 14. Perbandingan hasil klasifikasi SVM *Grid Search* dan SVM GA

Confusion matrix	Metode			
	SVM		SVM GA	
	Linier	RBF	Linier	RBF
Akurasi	88,24	<b>89,22</b>	88,24	88,24
AUC	88,30	<b>89,42</b>	88,30	88,30

Berdasarkan Tabel 14. didapatkan hasil bahwa akurasi terbaik dihasilkan oleh SVM kernel RBF dengan nilai akurasi sebesar 89,22% dan berdasarkan Tabel 14. model SVM kernel RBF memiliki nilai AUC yang paling tinggi yang artinya model tersebut adalah model yang paling baik dalam mengklasifikasikan data hipertensi.

## V. KESIMPULAN

Berdasarkan analisis dan pembahasan yang diperoleh, dapat disimpulkan bahwa nilai akurasi pada keempat model memiliki hasil yang baik. Keempat model dapat mengklasifikasikan penyakit hipertensi dengan sangat baik. Berdasarkan keempat metode tersebut hasil terbaik dalam mengklasifikasikan penyakit hipertensi adalah SVM *grid search* menggunakan kernel RBF, karena nilai akurasi yang diperoleh memiliki hasil yang paling tinggi yaitu 89,22%. Nilai AUC yang dihasilkan metode SVM *grid search* menggunakan kernel RBF juga menunjukkan hasil yang paling baik yaitu 89,42, yang artinya metode tersebut paling baik dalam mengklasifikasikan penyakit hipertensi.

## DAFTAR PUSTAKA

- Abe, S., 2005. *Support Vector Machines for Pattern Classification*. Verlag London: Springer.
- Agrawal, T., 2021. *Hyperparameter Optimization in Machine Learning*. India: Apress.
- Aulia, T. M. P., Arifin, N. dan Mayasari, R., 2021. *Perbandingan Kernel Support Vector Machine (SVM) dalam Penerapan Analisis Sentimen Vaksinasi Covid-19*. *SINTECH*, 4(2), pp. 139-145.
- Cristianini, N. dan Shawe-Taylor, J., 2000. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. United Kingdom: Cambridge University Pers.
- Darmawan, A., 2015. *Penerapan Model Support Vector Machine Text Mining Pada Komentar Review Smartphone Android vs Blackberry dengan Teknik Optimasi Genetic Algorithm*. *Factor Exacta*, VIII(2), pp. 100-115.
- Gorunescu, F., 2011. *Data Mining Concepts, Models and Techniques*. Verlag Berlin Heidelberg: Springer.
- Han, J., Kamber, M. dan Pei, J., 2012. *Data Mining Concepts and Techniques*. USA: Elsevier.
- Haryanto, Y. & Hidayatullah, R. S., 2016. *Komparasi Penerapan Algoritma Support Vector Machine dan SVM Optimasi Genetic Algorithms dalam Penentuan Penerimaan Dana*

- KJP pada SD Negeri 02 Meruya Utara Jakarta Barat*. Semarang, Seminar Nasional Ilmu Komputer (SNIK).
- Ilyas, S., 2016. *Tekanan Darah Tinggi (Hipertensi)*. Available at: <http://p2ptm.kemkes.go.id/uploads/2016/10/Tekanan-Darah-Tinggi-Hipertensi.pdf> [Diakses 11 Januari 2022].
- Meilina, P., 2015. *Penerapan Data Mining dengan metode Klasifikasi Menggunakan Decison Tree dan Regresi*. Jurnal Teknologi, 7(1), pp. 12-20.
- Nugroho, P. S. dan Fahrurrozi, D. S., 2018. Faktor Obesitas dan Kolestrol Terhadap Hipertensi di Indonesia (Indonesian Family Life Survey V). Ghidza : Jurnal Gizi dan Kesehatan, 2(2), pp. 44-48.
- Nuraini, B., 2015. *Risk Factors of Hypertension*. J Majority, IV(5), pp. 10-19.
- Purnama, D. I., 2020. *Peramalan Jumlah Penumpang Datang Melalui Transportasi Udara di Sulawesi Tengah Menggunakan Support Vector (SVR)*. Jurnal Ilmiah Matematika dan Terapan, 17(1), pp. 109-117.
- Santosa, B., 2007. *Data Mining: Teknik Pemanfaatan Data untuk Keputusan Bisnis*. Yogyakarta: Graha Ilmu.
- Setyohadi, D. B., Kristiawan, F. A. & Ernawati, 2017. *Perbaikan Performasi Klasifikasi dengan Preprocessing Iterative Partitioning Filter Algorithm*. TELEMATIKA, 14(1), pp. 12-20.
- Sivanandam, S. dan Deepa, S., 2008. *Introduction to Genetic Algorithms*. Verlag Berlin Heidelberg: Springer.
- Soenarta, A. A. et al., 2015. *Pedoman Tatalaksana Hipertensi Pada Penyakit Kardiovaskular*. Jakarta: Perki.
- Syarif, I., Prugel-Bennett, A. dan Wills, G., 2016. *SVM Parameter Optimization Using Grid Search and Genetic Algorithm to Improve Classification Performance*. TELKOMNIKA, 14(4), pp. 1502-1509.
- Windarto & Wijaya, B. S., 2012. *Aplikasi Penyusunan Jadwal dengan Algoritma Genetik pada Sekolah Menengah Kejuruan Budi Mulia Tangerang*. Jakarta, Prosiding Seminar Nasional Multidisiplin Ilmu.
- Yanaeng, S., Saelee, S. dan Samai, W., 2014. *Automatic Medical Case Study Essay Scoring by Support Vector Machine and Genetic Algorithms*. International Journal of Information and Education Technology, IV(2), pp. 132-137.