

# PERAMALAN JUMLAH KASUS TUBERKULOSIS DI RUMAH SAKIT UMUM HAJI MEDAN DENGAN METODE *SUPPORT VECTOR REGRESSION- PARTICLE SWARM OPTIMIZATION*

Sumawiyah Hsb<sup>1</sup>, Ismail Husein<sup>2</sup>, Rina Widyasari<sup>3</sup>  
Universitas Islam Negeri Sumatera Utara<sup>1,2,3</sup>

Email: [sumawiyah58@gmail.com](mailto:sumawiyah58@gmail.com)<sup>1</sup>, [husein.ismail@uinsu.ac.id](mailto:husein.ismail@uinsu.ac.id)<sup>2</sup>, [rinawidyasari@uinsu.ac.id](mailto:rinawidyasari@uinsu.ac.id)<sup>3</sup>

**Corresponding Author:** Sumawiyah Hsb email: [sumawiyah58@gmail.com](mailto:sumawiyah58@gmail.com)

**Abstrak.** Tuberkulosis adalah penyakit menular yang merupakan penyebab utama kesehatan yang buruk dan salah satu penyebab utama kematian diseluruh dunia, pada tahun 2021, Sumatra Utara menempati urutan ke-6 sebagai Provinsi dengan kasus TB terbesar setelah Provinsi Jawa Barat, Jawa Tengah, Jawa timur, DKI Jakarta, dan Banten. Hal ini bisa disebabkan oleh lingkungan yang tidak sehat, meningkatnya kejadian gizi buruk, munculnya HIV/AIDS. Oleh karena itu penelitian ini bertujuan untuk membuat model ramalan dengan metode Support Vector Regression (SVR) dengan optimasi Particle Swarm Optimazion (PSO). Tahap awal yang dilakukan pada penelitian ini yaitu melakukan analisis data jumlah penderita tuberkulosis yang diawali dengan menghitung korelasi antar data dengan faktor penyebab tuberkulosis. Kemudian melakukan preprocessing untuk menginisialisasikan nilai awal data, memilih jumlah fitur dan normalisasi data. Setelah tahap analisis, maka dilakukan perhitungan regresi untuk membandingkan nilai peramalan dan nilai aktual menggunakan metode Support Vector Regression (SVR) dengan optimasi parameter Particle swarm optimization (PSO) sehingga didapatkan hasil peramalan yang baik. Hasil dari penelitian ini didapatkan hasil peramalan dengan nilai MAPE = 34.85.

**Kata Kunci:** Peramalan, Tuberkulosis, Support Vector Regression, Particle Swarm Optimization.

**Abstract.** Tuberculosis is an infectious disease that is the leading cause of poor health and one of the major causes of death around the world, in 2021 north Sumatra ran sixth as the province with the highest Tuberculosis rate after Provinsi Jawa Barat, Jawa Tengah, Jawa timur, DKI Jakarta, and Banten. This may result from an unhealthy environment, an increase in nutrition events, the appearance of HIV/AIDS. Hence, this study aims to create a forecast model by the method of regression support (SVR) with an optimist Particle Swarm Optimazion (PSO). The initial stage of the study involves analyzing the data of those with tuberculosis that begins by calculating the correlation between data with the underlying factors. Then do the preprocessing to initial data value, selecting the number of features and normalization of data. After the analysis stage, regression calculations are made to compare the value of browsing and actual value using the Support Vector Regression (SVR) method of Support Vector Regression (SVR) with the Particle Swarm Optimization (PSO) so that a good fortune-giving result is obtained. The results of this study were obtained from an analysis with a value of MAPE = 35.85.

**Keywords:** Prediction, Tuberkulosis, Support Vector Regression, Particle Swarm Optimization.

## A. Pendahuluan

Pada tahun 2021, World Health Organization (WHO) melaporkan bahwa penyakit tuberkulosis (TB) sebagai penyakit menular paling mematikan pada urutan ke-2 didunia setelah Covid-19, dan berada pada urutan ke-13 penyebab utama kematian di seluruh dunia. WHO melaporkan bahwa estimasi jumlah orang terdiagnosis TBC secara global sebanyak 10,6 juta kasus atau naik sekitar 600.000 kasus dari tahun 2020 yang diperkirakan 10 juta kasus TB. Indonesia sendiri berada pada posisi ke-2 dengan jumlah penderita TB terbanyak di dunia setelah India, diikuti oleh China, Filipina,



Pakistan, Nigeria, Bangladesh dan Republik Demokratik Kongo secara berurutan. Pada tahun 2020, Indonesia berada pada posisi ketiga dengan beban jumlah kasus terbanyak, sehingga tahun 2021 jelas tidak lebih baik. Kasus TB di Indonesia diperkirakan sebanyak 969.000 kasus TB (satu orang setiap 33 detik). Angka ini naik 17% dari tahun 2020, yaitu sebanyak 824.000 kasus. Insidensi kasus TB di Indonesia adalah 354 per 100.000 penduduk, yang artinya setiap 100.000 orang di Indonesia terdapat 354 orang di antaranya yang menderita TB (WHO, Global Tuberculosis Report, 2022).

Pada tahun 2021, Sumatera Utara menempati urutan ke-6 sebagai Provinsi dengan kasus TB terbesar setelah Jawa Barat, Jawa Tengah, Jawa Timur, DKI Jakarta, Banten. Sumatera Utara menyumbang 22.169 kasus TB dari jumlah keseluruhan kasus TB di Indonesia. Data yang diperoleh dari Dinas Kesehatan Provinsi Sumatera Utara, penemuan kasus TB BTA positif tertinggi Sumatera Utara tahun 2020 yaitu Kota Meda, Deli Serdang, dan Simalungun. Sedangkan untuk penemuan kasus TB tahun 2021 di Kota Medan baru mencapai 10% (lebih kurang 1.000 kasus) dari target 18.000.

Hal ini disebabkan oleh lingkungan yang tidak sehat, semakin meningkatnya kejadian gizi buruk diberbagai negara berkembang, serta munculnya epidemik HIV/AIDS khususnya di Indonesia. Lebih cepatnya penyebaran TB juga mengakibatkan cukup tingginya jumlah individu latently-infected (individu-individu pengidap penyakit tetapi belum menularkan penyakit) dan jumlah individu actively-infected (individu-individu pengidap penyakit dan dapat menularkan penyakit) (Ramadhan et al, 2018).

Meninjau perubahan temporal dan melakukan peramalan terhadap angka penderita TB dimasa mendatang, memiliki peran penting dalam memberikan bahan pertimbangan dalam pengambilan keputusan bagi Dinas Kesehatan Provinsi Sumatera Utara sebagai upaya dalam menekan angka peningkatan penderita TB dan untuk mendukung target Wakil Presiden RI, yaitu Indonesia bebas TB ditahun 2025. Metode yang akan digunakan dalam penelitian ini adalah metode peramalan.

Peramalan biasanya diperlukan untuk memperkirakan apa yang akan terjadi dimasa yang akan datang, sehingga para pengambil keputusan dapat membuat perencanaan untuk masa yang akan datang (Montgomery et al, (2015) menyatakan peramalan (Forecasting) merupakan suatu prediksi untuk memperkirakan suatu nilai pada periode tertentu yang akan terjadi pada masa mendatang dengan menggunakan penjelasan secara matematik dan statistik.

Terdapat beberapa metode yang dapat digunakan untuk mendapatkan hasil peramalan. Salah satu contoh metodenya adalah Support Vector Regression (SVR), yang merupakan cabang dari Support Vector Machine (SVM). SVM muncul sebagai suatu teknik baru dan kuat dalam proses pembelajaran mesin (machine learning) dari data dan secara khusus dalam penyelesaian secara klasifikasi dan SVR sebagai penyelesaian permasalahan regresi.

SVR terbukti sebagai salah satu metode peramalan yang memberikan hasil yang jauh lebih baik dibandingkan dengan metode yang lainnya. Support Vector Regression (SVR) ini sendiri memiliki kelemahan yaitu ketidakmampuan SVR dalam menemukan kombinasi koefisien parameter yang mampu memberikan hasil yang lebih optimal, oleh karena itu diperlukan metode optimasi untuk memudahkan SVR menemukan kombinasi koefisien parameternya (Fadilah, et al, 2018).

Beberapa metode optimasi untuk mengoptimalkan hasil peramalan yaitu Genetic Algorithm (GA), Particle Swarm Optimization (PSO) dan yang lainnya. Metode optimasi yang digunakan dalam penelitian ini merupakan metode optimasi Particle Swarm Optimization (PSO). Optimasi PSO mampu memberikan hasil yang lebih baik dibandingkan dengan metode optimasi yang lain. Hal ini didukung oleh penelitian yang telah dilakukan oleh (Zeng, et al, 2009) mengenai peramalan jumlah kecelakaan, hasil pengujiannya mengindikasikan bahwa metode SVR-PSO mempunyai hasil yang lebih baik daripada menggunakan metode BP Neural Network. Selanjutnya yang dilakukan oleh (Lu & Geng, 2011) mengenai peramalan volume penjualan mobil dengan metode SVR-PSO menyatakan bahwa metode SVR-PSO lebih baik dari pada dengan metode SVR-GA dalam tingkat efisiensi dan akurasi prediksi. PSO merupakan sebuah algoritma dengan konsep sederhana tetapi sangat efektif dalam mengoptimasi permasalahan berbagai fungsi, selain itu algoritma PSO mudah dalam implementasi dan memiliki konvergensi yang cepat.



Oleh karena itu penelitian ini bertujuan untuk menentukan hasil ramalan dengan metode Support Vector Regression (SVR) dengan optimasi Particle Swarm Optimazion (PSO). Hasil ini dapat digunakan untuk menentukan kebijakan ke depannya serta untuk peramalan. Tahapan analisis sebagai berikut: Preprocessing Data, Explorasi Data, Normalisasi Data, Membagi Data Training dan Testing, Membuat Model SVR, Evaluasi Model SVR, Membuat Model SVR dengan PSO, Evaluasi Model SVR PSO, Kesimpulan.

Dari penelitian diatas tujuan yang ingin dicapai dari penelitian ini adalah mengetahui hasil ramalan jumlah kasus penyakit tuberkulosis paru di RSUD Haji Medan dengan menggunakan Support Vector Regression-Particle Swarm Optimization.

## B. Metode Penelitian

Jenis penelitian yang digunakan adalah penelitian kuantitatif dan penelitian kualitatif, penelitian ini dilakukan dengan menggunakan metode kuantitatif. Penelitian kuantitatif yaitu suatu penelitian data yang di dapatkan dalam bentuk angka dan dapat di proses menggunakan rumus matematika. Penelitian kuantitatif memiliki tujuan untuk membantu menemukan hubungan antara variabel yang ada dalam sebuah populasi. Penelitian ini dilaksanakan di RSUD Haji Medan yang beralamat di Jl. Rumah Sakit H. No. 47, Medan Estate, Kec. Percut Sei Tuan, Kabupaten Deli Serdang, Sumatera Utara.

Variabel adalah suatu tanda pengenal atau sifat atau nilai dari orang, objek, atau kegiatan yang mempunyai variasi tertentu yang ditetapkan oleh peneliti untuk dipelajari dan kemudian ditarik kesimpulannya (Sugiyono, 2011). Variabel yang digunakan dalam penelitian ini adalah variabel bebas dan variabel tak bebas. Variabel bebas (independen) adalah variabel yang mempengaruhi atau yang menjadi sebab terjadinya perubahan dan timbulnya variabel terikat. Variabel terikat (dependen) adalah variabel yang dipengaruhi atau menjadi akibat karna adanya variabel bebas. Pada penelitian ini jumlah kasus atau penderita TB Paru merupakan variabel terikat. Variabel yang dipengaruhi oleh jenis kelamin, usia, penderita HIV, kebiasaan merokok serta status gizi/buruk. Jenis kelamin, usia, penderita HIV, kebiasaan merokok dan status gizi/buruk merupakan variabel bebas karena dapat mempengaruhi penyebab TB. Jenis kelamin, usia, kebiasaan merokok, penderita HIV dan status gizi/buruk merupakan faktor penyebab TB.

**Tabel 1 Variabel Penelitian**

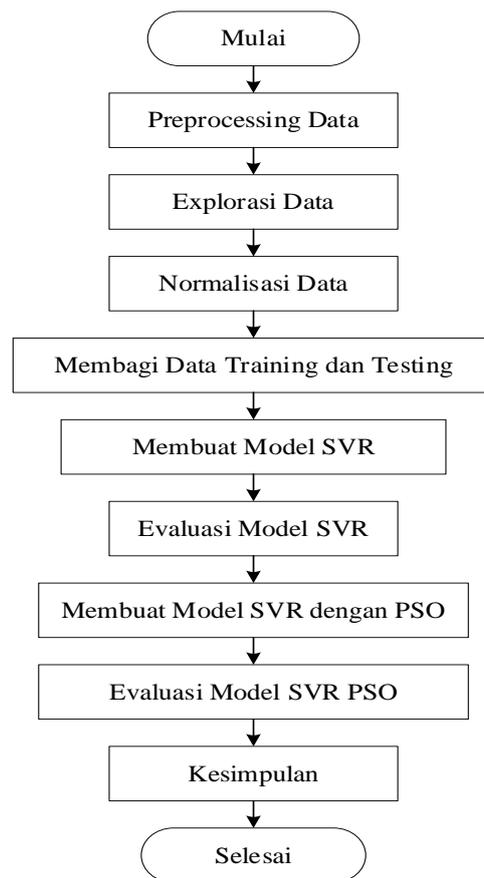
Variabel	Keterangan
Y	Jumlah Kasus TBC
X1	Jenis Kelamin Laki-Laki per Perempuan
X2	Usia
X3	Jumlah Penderita HIV
X4	Jumlah Perokok
X5	Jumlah Penderita Gizi Buruk

Berikut merupakan analisis data yang digunakan dalam penelitian ini yaitu dengan langkah-langkah sebagai berikut:

1. Mengumpulkan informasi mengenai jumlah data Tuberkulosis di RSUD Haji Medan
2. Menganalisis data-data yang telah diolah tersebut menggunakan metode Support Vector Regression (SVR) menggunakan Particle Swarm Optimization (PSO) untuk meramalkan jumlah kasus penyakit tuberkulosis.
3. Melakukan proses optimasi mulai dari preprocessing data, explorasi data, normalisasi data, membagi data training dan testing, membuat model SVR, evaluasi model SVR, membuat model SVR dengan PSO, evaluasi model SVR PSO, kesimpulan
4. Output yang diperoleh berupa hasil peramalan dan MAPE.

Berikut merupakan gambar diagram alir penelitian:





Gambar 1 Diagram Alir Penelitian

### C. Hasil dan Pembahasan Penelitian

Optimasi parameter dalam metode *Support Vector Regression* (SVR) menggunakan *Particle Swarm Optimazion* (PSO) memiliki tujuan untuk dapat menghasilkan nilai parameter paling baik atau paling optimal dengan nilai *error* (MAPE) paling kecil untuk meramalkan penderita TB di RSUD Haji Medan.

#### 1. Preprocessing Data

Sebelum data diolah menggunakan metode SVR-PSO, data di proses terlebih dahulu sehingga dapat maksimal saat digunakan. Data yang digunakan pada penelitian ini adalah penderita Tuberkulosis di RSUD Haji Medan tahun 2019 sampai dengan tahun 2022.

Penderita Tuberkulosis sepanjang tahun 2019-2022 sudah tergolong tinggi, yaitu mencapai 1.530 orang. Pasien dengan jenis kelamin laki-laki mencapai 792 orang, sedangkan pasien dengan jenis perempuan mencapai 738 orang, hal ini menunjukkan bahwa pasien Tuberkulosis berjenis kelamin laki-laki lebih dominan dibandingkan perempuan. Dari semua penderita Tuberkulosis, ada 26 orang yang juga menjadi penderita HIV, serta pada semua kasus Tuberkulosis, terdapat 82 orang yang memiliki kebiasaan merokok, dan ada 48 orang yang memiliki status gizi buruk.

#### 2. Tahap Analisis

Pada tahapan analisis adalah korelasi antara data kebiasaan merokok, pengidap HIV, jenis kelamin, usia, penderita gizi buruk terhadap penderita Tuberkulosis. Setiap parameter akan dihitung korelasinya yang akan bertujuan untuk mendapatkan nilai variabel dependen terhadap terhadap variabel independen.



a. Korelasi Antar Data

Korelasi antar data variabel yang berpengaruh terhadap penyakit Tuberkulosis dibagi menjadi beberapa bagian. Tiap faktor akan dihitung menggunakan persamaan untuk mendapatkan hasil dari nilai korelasi. Berikut merupakan hasil perhitungan korelasi data.

1) Variabel Jenis Kelamin

(a) Laki-laki

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{N\{\sum X^2 - (\sum X)^2\}\{N(\sum Y^2 - (\sum Y)^2)\}}}$$

$$r_{xy} = \frac{157896 - 1211760}{\sqrt{(858312) - (3002406)}}$$

$$r_{xy} = 0,22$$

(b) Perempuan

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{N\{\sum X^2 - (\sum X)^2\}\{N(\sum Y^2 - (\sum Y)^2)\}}}$$

$$r_{xy} = \frac{1424976 - 1129140}{\sqrt{(704382) - (3002406)}}$$

$$r_{xy} = 0,20$$

Berdasarkan hasil korelasi menunjukkan bahwa nilai korelasi 0,00-0,24 artinya tidak terdapat korelasi antara faktor kejadian Tuberkulosis dengan jenis kelamin penderita Tuberkulosis di RS Haji Medan.

2) Variabel Usia

$$r_{xy} = \frac{N \sum XY - (\sum X) - (\sum Y)}{\sqrt{N\{\sum X^2 - (\sum X)^2\}\{N(\sum Y^2 - (\sum Y)^2)\}}}$$

$$r_{xy} = \frac{3477456 - 3369978}{\sqrt{(499113) - (3002406)}}$$

$$r_{xy} = 0,02$$

Berdasarkan hasil korelasi menunjukkan bahwa nilai korelasi 0,00-0,24 artinya tidak terdapat korelasi antara faktor kejadian Tuberkulosis dengan usia penderita Tuberkulosis di RS Haji Medan.

3) Variabel Status Penderita HIV

$$r_{xy} = \frac{N \sum XY - (\sum X) - (\sum Y)}{\sqrt{N\{\sum X^2 - (\sum X)^2\}\{N(\sum Y^2 - (\sum Y)^2)\}}}$$

$$r_{xy} = \frac{45600 - 39780}{\sqrt{(1798) - (3002406)}}$$

$$r_{xy} = 0,79$$

Berdasarkan hasil korelasi menunjukkan bahwa nilai korelasi 0,75-1 artinya terdapat korelasi yang sangat kuat antara faktor kejadian Tuberkulosis dengan penderita HIV di RS Haji Medan.

4) Variabel Kebiasaan Merokok



$$r_{xy} = \frac{N \sum XY - (\sum X) - (\sum Y)}{\sqrt{N\{\sum X^2 - (\sum X^2)\}\{N(\sum Y^2 - (\sum Y^2))\}}}$$

$$r_{xy} = \frac{166368 - 12546}{\sqrt{(11342) - (3002406)}}$$

$$r_{xy} = 0,22$$

Berdasarkan hasil korelasi menunjukkan bahwa nilai korelasi 0,00-0,24 artinya tidak terdapat korelasi antara faktor kejadian Tuberkulosis dengan penderita kebiasaan merokok di RS Haji Medan.

5) Variabel Status Penderita Gizi Buruk

$$r_{xy} = \frac{N \sum XY - (\sum X) - (\sum Y)}{\sqrt{N\{\sum X^2 - (\sum X^2)\}\{N(\sum Y^2 - (\sum Y^2))\}}}$$

$$r_{xy} = \frac{84864 - 7344}{\sqrt{(4368) - (3002406)}}$$

$$r_{xy} = 0,99$$

Berdasarkan hasil korelasi menunjukkan bahwa nilai korelasi 0,75-1 artinya terdapat korelasi yang sangat kuat antara faktor kejadian Tuberkulosis dengan penderita gizi buruk di RS Haji Medan.

### 3. Normalisasi Data

Pada tahap ini dilakukannya normalisasi agar data training dan data testing berada pada rentang yang sama sehingga kompleksitas perhitungan berkurang dan untuk mengurangi tingkat *error* dalam komputasi. Hal ini dikarenakan SVR akan bekerja lebih efektif dengan data yang memiliki rentang skala yang kecil. Pada penelitian ini, nilai  $x_{min} = 4$ , dan  $x_{max} = 70$ . Perhitungan normalisasi data training dan data testing seperti berikut :

$$x'_1 = \frac{(x_1 - x_{min})}{(x_{max} - x_{min})}$$

$$x'_1 = \frac{(43 - 4)}{(70 - 4)}$$

$$x'_1 = 0,59$$

$$x'_2 = \frac{(x_2 - x_{min})}{(x_{max} - x_{min})}$$

$$x'_2 = \frac{(40 - 4)}{(70 - 4)}$$

$$x'_2 = 0,66$$

$$x'_3 = \frac{(x_3 - x_{min})}{(x_{max} - x_{min})}$$

$$x'_3 = \frac{(45 - 4)}{(70 - 4)}$$

$$x'_3 = 0,62$$

$$x'_4 = \frac{(x_4 - x_{min})}{(x_{max} - x_{min})}$$

$$x'_4 = \frac{(56 - 4)}{(70 - 4)}$$

$$x'_4 = 0,78$$



$$x'_5 = \frac{(x_5 - x_{min})}{(x_{max} - x_{min})}$$

$$x'_5 = \frac{(70 - 4)}{(70 - 4)}$$

$$x'_5 = 1,0$$

$$x'_6 = \frac{(x_6 - x_{min})}{(x_{max} - x_{min})}$$

$$x'_6 = \frac{(43 - 4)}{(70 - 4)}$$

$$x'_6 = 0,59$$

$$x'_7 = \frac{(x_7 - x_{min})}{(x_{max} - x_{min})}$$

$$x'_7 = \frac{(50 - 4)}{(70 - 4)}$$

$$x'_7 = 0,69$$

$$x'_8 = \frac{(x_8 - x_{min})}{(x_{max} - x_{min})}$$

$$x'_8 = \frac{(48 - 4)}{(70 - 4)}$$

$$x'_8 = 0,66$$

$$x'_9 = \frac{(x_9 - x_{min})}{(x_{max} - x_{min})}$$

$$x'_9 = \frac{(54 - 4)}{(70 - 4)}$$

$$x'_9 = 0,75$$

Hasil dari nilai normalisasi disajikan pada tabel 4.3

**Tabel 3 Data Normalisasi Penderita Tuberkulosis**

Tanggal	Penderita TB
01/04/2019	0,59
01/05/2019	0,66
01/06/2019	0,62
01/07/2019	0,78
01/08/2019	1,0
01/09/2019	0,59
01/10/2019	0,69
01/11/2019	0,66
01/12/2019	0,75

#### 4. Membagi Data Training dan Testing

Data awal dibagi menjadi data training dan data testing. Proporsinya yaitu 80% data training atau 39 observasi dan 20% data testing atau 9 observasi. Data training digunakan untuk membuat model SVR baik Non-PSO maupun PSO. Data testing digunakan untuk memvalidasi model.



## 5. Membuat Model SVR

Pada pemodelan SVR pada penelitian kali ini menggunakan fungsi kernel RBF. Menurut (Hsu et al., 2016) fungsi kernel yang direkomendasikan adalah fungsi kernel RBF, karena memiliki kinerja yang sama dengan kernel linier pada beberapa parameter, dan memiliki perilaku tertentu, seperti fungsi kernel tipe-S dengan parameter tertentu dan rentang nilai yang kecil [0,1]. dalam model SVR, fungsi kernel gaussian (RBF) memiliki tiga parameter dalam modelnya, yaitu *cost* ( $c$ ), *gamma* ( $\gamma$ ) dan *epsilon* ( $\epsilon$ ). Untuk mendapatkan hasil parameter perlu dilakukan pemilihan parameter secara tepat. Dalam metode SVR terdapat metode optimasi didalamnya yang disebut dengan metode *Grid Search*. Metode *Grid Search* melakukan pemodelan dengan menggunakan SVR dengan *range* parameter tertentu. Parameter yang digunakan dilakukan secara bertahap dengan merubah parameternya selama di dalam range. Range parameter yang digunakan dalam penelitian ini adalah  $\{(10^{-4}:10), (10^{-4}:10), (10^{-4}:1)\}$  secara berturut-turut untuk parameter *cost*, *gamma*, *epsilon*.

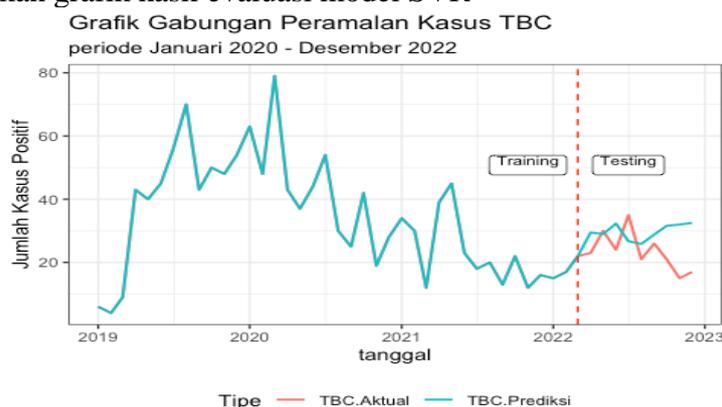
**Tabel 4 Nilai Parameter**

<i>Cost</i> ( $C$ )	<i>Gamma</i> ( $\gamma$ )	<i>Epsilon</i> ( $\epsilon$ )
2.0001	1.0001	1e04

Berdasarkan tabel di atas, dapat dilihat bahwa parameter optimum untuk model SVR tanpa optimasi PSO adalah  $cost = 2.0001$ ,  $gamma = 1.0001$ , dan  $epsilon = 1e-04$ . Selanjutnya membuat evaluasi model dengan melihat nilai MAPE data training dan MAPE data testing.

## 6. Evaluasi Model SVR

Berikut merupakan grafik hasil evaluasi model SVR



Gambar 1 Grafik Hasil Prediksi

Berdasarkan grafik di atas dapat dilihat bahwa TBC hasil prediksi pada training sangat mirip dengan data aktual, sehingga grafik pada training untuk TBC aktual dan TBC prediksi saling tumpang tindih. hal ini diperkuat dengan nilai MAPE data training yang sangat kecil yaitu 0.017. Sedangkan untuk hasil prediksi data testing dapat dilihat bahwa grafik TBC prediksi cenderung memiliki trend ke atas, sedangkan grafik TBC data aktual cenderung memiliki trend ke bawah sehingga nilai MAPE yang didapatkan cenderung besar yaitu 42.09. Oleh karena itu model SVR tanpa PSO mengalami *overfitting* (perilaku machine learning yang tidak diinginkan yang terjadi saat model machine learning memberikan prediksi yang akurat untuk data pelatihan tetapi tidak untuk data testing). Oleh karena itu dibuat model SVR-PSO untuk mengatasi masalah *overfitting* tersebut.

## 7. Membuat Model SVR dengan PSO

Optimasi dengan menggunakan PSO memiliki lima langkah utama yang harus dilakukan, yaitu sebagai berikut:

1. Mendefinisikan parameter-parameter yang digunakan dalam PSO, yaitu jumlah partikel, batas iterasi maksimum, bobot inersia, batas maksimum kecepatan, koefisien komponen tiap partikel ( $C_1$ ) dan koefisien komponen semua partikel ( $C_2$ ). Jumlah partikel yang

- digunakan adalah 100 partikel dengan batas iterasi maksimum adalah 100 kali iterasi. Bobot inersia ( $w$ ) bernilai 0,9 dan Batasan kecepatan maksimum 1.  $C_1$  dan  $C_2$  bernilai sama dengan 0,2.
- Menginisialisasikan vektor posisi dari tiap partikel. Dimensi dari partikel PSO ditentukan dari jumlah parameter yang akan dioptimasi. Pada penelitian ini parameter yang akan dioptimasi adalah parameter dari Model SVR dengan fungsi kernel Gaussian (RBF), yaitu cost, gamma dan epsilon. Jadi vektor posisi terbentuk dalam ukuran 3x1, karena terdapat tiga parameter yang akan dioptimasi. Posisi parameter akan di generate secara random diantara range parameter yang digunakan. Range parameter yang digunakan berturut-turut untuk parameter cost, gamma dan epsilon adalah  $\{(10-3:4),(10-3:1),(10-6:1)\}$ .
  - Menentukan fitness function. Fitness function yang digunakan adalah menggunakan SMAPE. Dari 100 vektor posisi dihitung nilai fitnessnya dan kemudian diurutkan dari nilai paling kecil hingga paling besar.
  - Langkah selanjutnya adalah melakukan pergeseran partikel dengan menggunakan rumus berikut:

$$x_i(t + 1) = x_i(t) + v_i(t + 1)$$

Vektor posisi suatu partikel akan ditambah dengan vektor kecepatan dan menjadi sebuah posisi baru. Vektor kecepatan tersebut dihitung dengan menggunakan rumus berikut.

$$v_i(t) = v_i(t - 1) + c_1r_1(localbest(t) - x_i(t - 1)) + c_2r_2(globalbest(t) - x_i(t - 1))$$

Dimana local best merupakan nilai *fitness* terbaik dari suatu partikel dan *global best* merupakan nilai *fitness* terbaik dari keseluruhan partikel. Setelah dilakukan pergeseran, selanjutnya dihitung nilai *fitness* dari posisi partikel yang baru.

Kemudian langkah 1 hingga 4 diulang hingga mencapai batas iterasi atau setelah semua partikel menghasilkan *fitness* yang konvergen.

Setelah didapatkan hasil yang konvergen maka dapat dilakukan peramalan dan kemudian dibandingkan dengan data testing dan training.

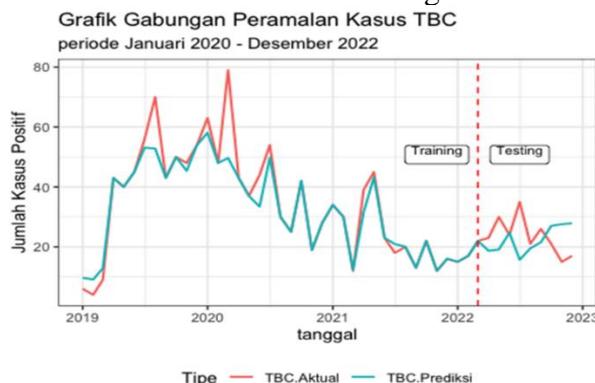
**Tabel 5 Parameter Optimum**

$C$	$\gamma$	$\epsilon$
0.9402773	0.3600042	0.001397562

Berdasarkan tabel pada 4.7 di atas, dapat dilihat bahwa parameter optimum untuk model SVR dengan optimasi PSO adalah cost = 0.9403, gamma = 0.3600, dan epsilon = 0.0014. Selanjutnya membuat evaluasi model dengan melihat nilai MAPE data training dan MAPE data testing.

### 8. Evaluasi Model SVR dengan PSO

Berikut merupakan grafik hasil evaluasi model SVR dengan PSO



Gambar 2 Grafik Perbandingan antar Grafik Aktual dan Grafik Prediksi

Berdasarkan grafik pada 4.8 dapat dilihat bahwa pada data training, grafik TBC aktual dan grafik TBC prediksi memiliki pola yang sama. Hal ini terbukti dengan nilai MAPE pada data training yang kecil yaitu 9.98. Sedangkan pada data testing grafik TBC aktual dan TBC prediksi cenderung



memiliki pola yang sama walau tidak semirip data traing. Hal ini dibuktikan oleh nilai MAPE lebih besar dari data training yaitu 34.86.

#### D. Kesimpulan

Berdasarkan hasil percobaan pada penelitian yang telah dilakukan, dapat disimpulkan bahwa peramalan jumlah kasus Tuberkulosis di RSUD Haji Medan menggunakan metode *support vector regression* (SVR) tanpa optimasi memperoleh MAPE data Training sebesar 0.28 dan MAPE data Testing sebesar 41.68. Sedangkan *Support Vector Regression* (SVR) dengan optimasi *Particle Swarm Optimization* (PSO) memperoleh MAPE data Training 9.97 dan MAPE data Testing 34.85

Berdasarkan nilai evaluasi di atas dapat dilihat bahwa model *Support Vector Regression* (SVR) tanpa optimasi *Particle Swarm Optimization* (PSO) memiliki nilai overfitting yang sangat tinggi bila dibandingkan model *Support Vector Regression* (SVR) dengan optimasi *Particle Swarm Optimization* (PSO). Oleh karena itu peramalan Kasus TBC di rumah sakit umum haji Medan lebih baik menggunakan *Support Vector Regression* (SVR) dengan optimasi *Particle Swarm Optimization* (PSO).

## DAFTAR PUSTAKA

- Andayani, Sri. (2020). Prediksi Kejadian Penyakit Tuberkulosis Paru Berdasarkan Jenis Kelamin. *Jurnal Keperawatan Muhammadiyah Bengkulu*. 8(2). 135- 140.
- Departement Kesehatan RI 2021. *Profil Kesehatan Indonesia 2021*.
- Hsu, C., Chang, C., & Lin, C. (2016). *A Practical Guide to Support Vector Classification*. 1(1), 1–16.
- Montgomery, D, C., Jennings, C, L., Kulachi, M. (2015). *Introduction To Time Series Analysis and Forecasting*. 2nd ed. S. I.: Wiley.
- Ramadhan, M, R., Stevanus, B, W., Muhammad, K. (2018). *Pemodelan Matematika Penyebaran Penyakit Tuberkulosis Dengan Strategi DOTS*. *UNNES Journal of Mathematics*. 7(2). 130-141.
- Sugiyono. (2011). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta.
- Sun, Jun. Choi, H, L., Xiao, J, W. (2012). *Particle Swarm Optimization Classical and Quantum Perspectives*. New York: Taylor & Francis Group, LLC.
- Suyono, A, A., Kusriani. Radiyanto, A. (2022). *Prediksi Indeks Harga Konsumen Komoditas Makanan di Kota Surabaya Menggunakan Metode Support Vektor Regression*. *Jurnal Metik*. 6(1). 45-51.
- Trismanjaya Hulu, Victor. Dkk. (2020). *Epidemiologi Penyakit Menular: Riwayat, Penularan dan Pencegahan*. Medan: Yayasan Kita Menulis. 23.
- Umiyati, A., Dadan, D., Fitriani, A. (2021). *Peramalan Harga Batubara Acuan Menggunakan Metode PSOSVR dan IPSOSVR*. *JEM*. 9(1). 71-95.
- World Health Organization (WHO) 2022. *Global Tuberculosis Report 2022*.

