

# KLASIFIKASI STATUS KEMISKINAN RUMAH TANGGA DENGAN METODE *SUPPORT VECTOR MACHINES* (SVM) DAN *CLASSIFICATION AND REGRESSION TREES* (CART) MENGGUNAKAN GUI R (Studi Kasus di Kabupaten Wonosobo Tahun 2018)

Lutfia Nuzula<sup>1</sup>, Alan Prahutama<sup>2</sup>, Arief Rachman Hakim<sup>3</sup>

<sup>1,2,3</sup>Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro  
*e-mail* : [lutfianuzula@students.undip.ac.id](mailto:lutfianuzula@students.undip.ac.id)

## ABSTRACT

The poor are people who have average monthly expenditures per capita below the poverty line. Wonosobo District became the poorest district in Central Java in 2011-2018, although the percentage of poor people has decreased every year. It cannot be separated from the efforts of the Wonosobo District Government to overcome poverty through various programs. This study classified households in Wonosobo District in 2018 as poor and non-poor based on influencing factors. This study used the *Support Vector Machines* (SVM) method to be compared with the *Classification and Regression Trees* (CART) method. It used the data from the 2018 National Socio-Economic Survey of Central Java with a total of 795 observations. Result of the research using the SVM method and the RBF kernel, the classification accuracy reaches 89.82% then the classification accuracy using the CART method reaches 87.08%. GUI designed by RShiny package can make easier for users to analyze the SVM and CART with the valid output.

**Keywords** : Poverty, *Support Vector Machines* (SVM), *Classification and Regression Trees* (CART), Accuracy

## 1. PENDAHULUAN

Menurut Nugroho (1995) definisi kemiskinan dibentuk berdasarkan identifikasi dan pengukuran terhadap sekelompok masyarakat atau golongan yang selanjutnya disebut miskin. Setiap negara memiliki definisi sendiri mengenai suatu masyarakat dikategorikan miskin atau tidak miskin, hal ini dikarenakan kondisi yang disebut miskin bersifat relatif untuk setiap negara. Definisi kemiskinan ditentukan menurut kriteria atau ukuran yang didasarkan pada pendapatan rata-rata, kemampuan konsumsi rata-rata, akses terhadap fasilitas kesehatan maupun pendidikan.

Dalam tujuh tahun terakhir, Kabupaten Wonosobo menjadi kabupaten yang memiliki presentase angka kemiskinan tertinggi di Jawa Tengah. Berbagai program dan kegiatan yang dicanangkan pemerintah belum memberikan kontribusi yang signifikan dalam menurunkan angka kemiskinan. Hal ini dikarenakan data terpadu sebagai basis data penanggulangan kemiskinan belum dipakai sebagai dasar penerima sasaran, sehingga banyak keluarga miskin yang belum menerima bantuan perlindungan sosial.

Salah satu upaya yang dapat dilakukan untuk menanggulangi masalah kemiskinan adalah dengan melakukan klasifikasi status kemiskinan rumah tangga di Kabupaten Wonosobo tahun 2018 berdasarkan faktor-faktor yang mempengaruhi. Pengklasifikasian status rumah tangga ini bertujuan untuk mengetahui layak atau tidaknya suatu rumah tangga menerima bantuan program penanggulangan kemiskinan. Menurut Prasetyo (2012), klasifikasi merupakan suatu pekerjaan menilai objek untuk memasukkannya kedalam kelas tertentu dari sejumlah kelas yang tersedia. Metode pengklasifikasian yang akan digunakan adalah *Support Vector Machines* (SVM) dan *Classification and Regression Trees* (CART), serta dilakukan perancangan *interface* dengan R-Shiny untuk memudahkan *user* yang ingin melakukan analisis klasifikasi SVM dan CART tanpa membuat pemrograman terlebih dahulu.

## 2. TINJAUAN PUSTAKA

### 2.1 Kemiskinan

Definisi kemiskinan yang tertuang dalam Peraturan Presiden Republik Indonesia Nomor 7 Tahun 2005 tentang Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2004-2009 bahwa kemiskinan merupakan suatu kondisi dimana seseorang atau sekelompok orang, laki-laki dan perempuan, tidak terpenuhi hak-hak dasarnya untuk mempertahankan dan mengembangkan kehidupan yang bermartabat. Sedangkan definisi kemiskinan menurut Badan Pusat Statistik (BPS) dan Departemen Sosial, kemiskinan adalah ketidakmampuan individu dalam memenuhi kebutuhan dasar minimal untuk hidup layak (baik makanan maupun non makanan).

Pemerintah menggunakan garis kemiskinan berdasarkan ukuran dari BPS (Badan Pusat Statistik) yang dihitung berdasarkan data Survei Sosial-Ekonomi Nasional (SUSENAS) sebagai indikator kemiskinan. Dari pengukuran dengan garis kemiskinan, diketahui bahwa penduduk miskin merupakan penduduk yang memiliki rata-rata pengeluaran perkapita perbulan dibawah garis kemiskinan.

### 2.2 *Support Vector Machines*

#### 2.2.1 Klasifikasi Linier Separable

Misalkan diberikan himpunan  $X = \{x_1, x_2, \dots, x_t\}$  dengan  $x_i \in R^n$ , dimana  $i = 1, 2, \dots, t$  data *training*, untuk  $y_i \in \{-1, +1\}$  menyatakan label kelas. Sehingga data berupa pasangan  $(x_1, y_1), (x_2, y_2), \dots, (x_t, y_t)$  merupakan himpunan data *training* dari dua kelas yang akan diklasifikasikan dengan *Support Vector Machine*. Diasumsikan kedua kelas -1 dan +1 dapat terpisah secara sempurna oleh *hyperplane* berdimensi  $n$ , yang didefinisikan sebagai :  $(\mathbf{w} \cdot \mathbf{x}) + b = 0$ ,  $\mathbf{w}$  dan  $b$  merupakan parameter model. Untuk mendapatkan fungsi pemisah terbaik adalah dengan mencari fungsi pemisah yang terletak ditengah-tengah antara dua bidang pembatas kelas. Untuk mendapatkan fungsi pemisah terbaik itu, sama dengan memaksimalkan margin atau jarak antara dua set objek dari kelas yang berbeda (Santosa, 2007). Selanjutnya diformulasikan kedalam persamaan *quadratic programming* dengan meminimalkan invers persamaan  $\min \frac{1}{2} \|\mathbf{w}\|^2$  dimana  $\|\mathbf{w}\|^2 = \mathbf{w}^T \mathbf{w}$  dengan syarat :  $y_i [(\mathbf{w} \cdot \mathbf{x}) + b] \geq 1 ; i = 1, 2, \dots, t$ . Optimalisasi dapat diselesaikan dengan *Lagrange Multiplier*.

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^t \alpha_i \{y_i [\mathbf{w}^T \cdot \mathbf{x}_i + b] - 1\} \quad (1)$$

Syarat optimal dari fungsi *lagrange multiplier* adalah  $\frac{\delta L}{\delta b} = 0$  dan  $\frac{\delta L}{\delta \mathbf{w}} = 0$

dimana  $\mathbf{w}^T \mathbf{w}$  dapat dijabarkan  $\mathbf{w}^T \mathbf{w} = \sum_{i=1}^t \alpha_i y_i (\mathbf{w}^T \cdot \mathbf{x}_i) = \sum_{i=1}^t \sum_{j=1}^t y_i y_j \alpha_i \alpha_j (\mathbf{x}_i^T \cdot \mathbf{x}_j)$

Maka persamaan (1) akan berubah menjadi dualitas *Lagrange Multiplier* sebagai berikut:

$$L_d = \sum_{i=1}^t \alpha_i - \frac{1}{2} \sum_{i=1}^t \sum_{j=1}^t y_i y_j \alpha_i \alpha_j (\mathbf{x}_i^T \cdot \mathbf{x}_j) \quad (2)$$

$$\max_d L_d = \max \sum_{i=1}^t \alpha_i - \frac{1}{2} \sum_{i=1}^t \sum_{j=1}^t y_i y_j \alpha_i \alpha_j (\mathbf{x}_i^T \cdot \mathbf{x}_j) \quad (3)$$

dengan batasan  $\alpha_i \geq 0 ; i = 1, 2, \dots, t$  dan  $\sum_{i=1}^t \alpha_i y_i = 0$

Fungsi keputusan yang dihasilkan hanya dipengaruhi oleh *Support Vector. Hyperplane* (fungsi pemisah) diperoleh dengan rumus :

$$f(\mathbf{z}) = \text{sign}(\mathbf{w} \cdot \mathbf{z} + b) = \left( \sum_{i=1}^p \alpha_i y_i \mathbf{x}_i \cdot \mathbf{z} \right) + b \quad (4)$$

dengan  $p$  merupakan jumlah data yang menjadi *Support Vector*,  $\mathbf{x}_i$  merupakan *Support Vector*,  $\mathbf{z}$  merupakan data *testing* yang akan diprediksi kelasnya, dan  $\mathbf{x}_i \cdot \mathbf{z}$  merupakan *dot-product* antara  $\mathbf{x}_i$  dan  $\mathbf{z}$  (Prasetyo, 2012).

### 2.2.2 Klasifikasi Linier Non-Separable

Struktur data yang terpisahkan linier tidak secara sempurna adalah data yang berada di dalam margin atau berada pada sisi yang salah dari batas keputusan. Hal ini menyebabkan proses optimasi tidak dapat diselesaikan, sehingga masalah optimasi pada rumusan SVM ditambah dengan mengikutsertakan variabel *slack* ( $\xi > 0$ ), sehingga diperoleh rumus sebagai berikut :

$$y_i \left[ (\mathbf{w} \cdot \mathbf{x}_i) + b \right] + \xi_i \geq 1 \quad ; i = 1, 2, \dots, t \quad (5)$$

$$(\mathbf{w}, \xi) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^t \xi_i \quad (6)$$

dalam formulasi ini, merupakan usaha untuk meminimalkan kesalahan klasifikasi yang dinyatakan dengan variabel *slack*  $\xi_i$  sementara dalam waktu yang bersamaan juga memaksimalkan margin (Santoso, 2007).

### 2.2.3 Klasifikasi Non Linier

Menurut Prasetyo (2012) untuk menyelesaikan problem *non*-linier, SVM dimodifikasi dengan memasukkan fungsi kernel. Metode kernel memetakan data ke dalam ruang lain dengan dimensi yang lebih tinggi sehingga data pada ruang input yang awalnya bersifat *non*-linier akan menjadi linier pada dimensi baru. Berikut merupakan fungsi kernel menunjukkan pemetaan *non*-linier pada *feature space* :

$$K(\mathbf{x}_i, \mathbf{z}) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{z}) \quad (7)$$

Prediksi sekumpulan data dengan dimensi fitur yang baru diformulasikan dengan : (Prasetyo, 2012)

$$f(\mathbf{z}) = \text{sign} \left( \sum_{i=1}^p \alpha_i y_i K(\mathbf{x}_i, \mathbf{z}) + b \right) \quad (8)$$

Berikut merupakan fungsi kernel yang populer dan sering digunakan :

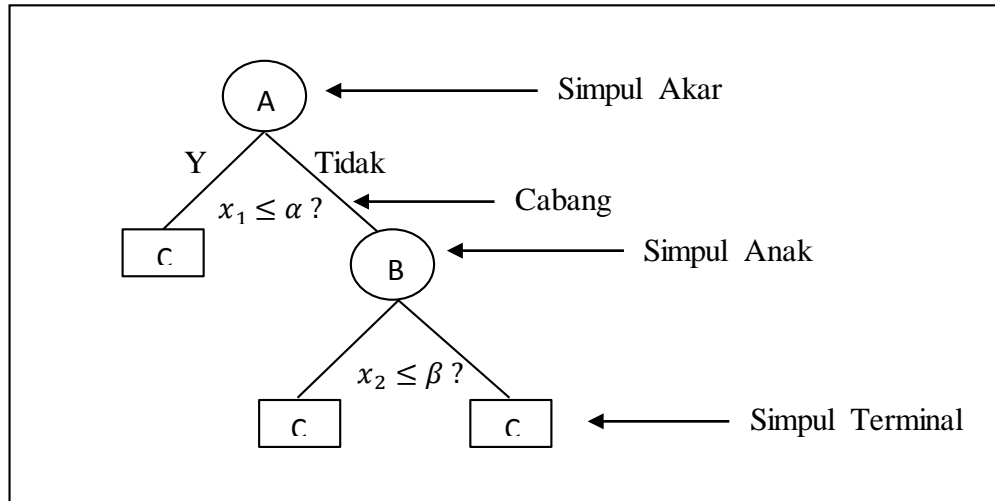
**Tabel 1** Fungsi kernel yang sering digunakan

Jenis Kernel	Definisi Fungsi
Linear	$K(\mathbf{x}_i, \mathbf{z}) = \mathbf{x}_i^T \cdot \mathbf{z}$
Polinomial	$K(\mathbf{x}_i, \mathbf{z}) = (\mathbf{x}_i^T \cdot \mathbf{z} + 1)^d$
Radial Basis Function (RBF)	$K(\mathbf{x}_i, \mathbf{z}) = \exp\left(-\frac{1}{2\sigma^2} \ \mathbf{x}_i - \mathbf{z}\ ^2\right)$

## 2.3 Classification and Regression Trees

*Classification and Regression Trees* atau yang dikenal dengan CART merupakan salah satu algoritma dari metode klasifikasi *Decision Tree*. Menurut Roger dan Lewis (2000) CART merupakan teknik klasifikasi yang menggunakan algoritma *binary recursive partitioning*

(penyekataan ulang secara biner). Hal ini dikarenakan setiap simpul akan selalu mengalami pemecahan menjadi dua simpul anak, dimana proses pemecahan akan terus berulang hingga tidak ada kesempatan untuk pemecahan berikutnya, dan proses klasifikasi dilakukan dengan memilah sekumpulan data menjadi beberapa bagian atau partisi yang lebih kecil. Struktur dari pohon klasifikasi adalah sebagai berikut :



**Gambar 1.** Diagram CART

Berdasarkan Gambar 1 dapat dijelaskan bahwa A B dan C merupakan variabel respon yang terpilih untuk menjadi simpul. A merupakan simpul induk, sementara B dan C merupakan simpul anak, dimana C juga merupakan simpul akhir yang tidak bercabang lagi. Sementara  $\alpha$  dan  $\beta$  merupakan suatu nilai yang merupakan nilai tengah antara dua nilai amatan peubah  $x_j$  secara berurutan. Berikut tahapan dalam analisis CART :

1. Proses Pemilihan Simpul

Pada proses pemilihan simpul, akan ditentukan pemilah dari setiap simpul yang menghasilkan penurunan tingkat heterogenitas tertinggi. Adapun untuk fungsi heterogenitas yang digunakan adalah indeks gini karena akan selalu memisahkan kelas dengan anggota paling besar/kelas terpenting dalam simpul terlebih dahulu, dengan persamaan  $I(t) = \sum_{i \neq j} p(j|t)p(i|t)$ . Proses pemecahan pada masing-masing simpul induk didasarkan pada aturan *goodness of split* (kriteria pemecahan terbaik), dirumuskan sebagai berikut:

$$\square_i(s,t) = I(t) - P_R I(t_R) - P_L I(t_L) \tag{9}$$

Pemilah yang menghasilkan  $\square_i(s,t)$  lebih tinggi merupakan pemilah terbaik karena mampu mereduksi heterogenitas lebih tinggi (Kuswanto dan Mubarak, 2019).

2. Proses Pelabelan Kelas

Pelabelan kelas dilakukan mulai dari awal pemilahan simpul hingga simpul akhir terbentuk, penentuan label kelas pada simpul terminal berdasarkan aturan jumlah terbanyak, yaitu jika  $p(j_0|t) = \max_j p(j|t) = \max_j \frac{N_j(t)}{N(t)}$ .

3. Proses Penghentian Pemilahan

Proses pemilahan simpul dihentikan ketika hanya ada satu pengamatan di setiap simpul anak, semua pengamatan dalam setiap simpul anak memiliki distribusi variabel prediktor yang identik, dan batas level pemilahan yang ditetapkan oleh peneliti (Roger dan Lewis, 2000)

4. Proses Pemangkasan Pohon Klasifikasi

Proses pemangkasan pohon bertujuan untuk mencegah terbentuknya pohon klasifikasi yang berukuran besar. Ukuran pemangkasan yang digunakan untuk memperoleh ukuran pohon yang layak adalah *Cost Complexity Minimum*, yaitu:

$$R(T) = \sum_{t \in \bar{T}} r(t) p(t) = \sum_{t \in \bar{T}} R(t) \quad ; \text{dimana} \quad r(t) = 1 - \max_j p(j|t) \quad (10)$$

Jika diperoleh dua simpul anak dan simpul induk yang memenuhi persamaan  $R(t) = R(t_L) + R(t_R)$  maka simpul anak  $t_L$  dan  $t_R$  dipangkas.

### 2.1. K-Fold Cross Validation

*Cross Validation* merupakan salah satu teknik untuk menilai/menvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu. Data yang digunakan dalam proses pembangunan model disebut data *training*, sedangkan data yang akan digunakan untuk menvalidasi model disebut data *testing*. Menurut Prasetyo (2012), dalam pendekatan *k-fold cross validation* yaitu memecah set data menjadi  $k$  bagian set data dengan ukuran yang sama.

### 2.2. Pengukuran Kinerja Klasifikasi

Sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua set data dengan benar. Namun, tidak dapat dipungkiri bahwa kinerja suatu sistem tidak bisa bekerja 100% benar. Oleh karena itu, sebuah sistem klasifikasi juga harus diukur kinerjanya. Umumnya cara mengukur kinerja klasifikasi menggunakan matriks konfusi. Matriks konfusi merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi.

**Tabel 2.** Matriks Konfusi

Akurasi		Hasil Prediksi ( <i>Predicted Class</i> )	
		Kelas 1	Kelas 0
Hasil Observasi ( <i>Actual Class</i> )	Kelas 1	F <sub>11</sub>	F <sub>10</sub>
	kelas 0	F <sub>01</sub>	F <sub>00</sub>

Akurasi klasifikasi dapat dihitung dengan formula sebagai berikut :

$$\text{Akurasi klasifikasi (\%)} = \frac{F_{11} + F_{00}}{F_{11} + F_{01} + F_{10} + F_{00}}$$

### 2.3. Graphical User Interface dengan R-Shiny

R-Shiny merupakan salah satu paket pada R yang merupakan *toolkit* disusun oleh grup Rstudio. R-Shiny mengizinkan penggunanya membangun *web apps* yang interaktif yang menggabungkan antara kekuatan komputasi statistika R dan interaksinya dengan web modern. Struktur dari aplikasi R-Shiny terdiri dari 3 komponen, yaitu :

1. *User Intercafe* (UI), merupakan fungsi yang mendefinisikan tampilan web dari aplikasi yang dibangun. UI berfungsi untuk memuat seluruh input dan output yang akan ditampilkan pada aplikasi. Untuk membuat web interaktif, kontrol dilakukan melalui file *ui.r*
2. *Server* merupakan fungsi yang mendefinisikan logika kerja analisis dari sisi server aplikasi, yang berisi beberapa perintah terkait dengan permintaan input atau output pada file. Format *syntax* pada fungsi server berupa *server.r*.
3. *ShinyApp* merupakan fungsi dari aplikasi yang memanggil UI dan Server untuk menjalankan aplikasi yang telah dibangun.

### 3. METODOLOGI PENELITIAN

#### 3.1 Data dan Variabel Penelitian

Jenis data yang digunakan dalam penelitian ini merupakan data sekunder, diperoleh dari Survei Sosial Ekonomi Nasional (SUSENAS) yang dilakukan oleh Badan Pusat Statistik (BPS) Provinsi Jawa Tengah pada Maret 2018. Data yang digunakan terdiri dari 795 amatan rumah tangga. Variabel dalam penelitian ini terdiri atas 1 variabel respon (Y) adalah status kemiskinan rumah tangga, dimana terbagi menjadi dua kelas yaitu rumah tangga miskin dan rumah tangga tidak miskin. Sedangkan variabel penjelas (X) yang digunakan terdiri dari : jenjang pendidikan tertinggi kepala rumah tangga ( $X_1$ ), bahan bangunan utama atap rumah terluas ( $X_2$ ), bahan bangunan utama dinding rumah terluas ( $X_3$ ), bahan bangunan utama lantai rumah terluas ( $X_4$ ), kepemilikan fasilitas tempat buang air besar ( $X_5$ ), sumber air utama yang digunakan untuk minum ( $X_6$ ), kepemilikan emas ( $X_7$ ), kepemilikan mobil ( $X_8$ ), kepemilikan sepeda motor ( $X_9$ ), kepemilikan kulkas ( $X_{10}$ ), dan luas lantai bangunan tempat tinggal ( $X_{11}$ ).

#### 3.2 Metode Analisis Data

Analisis data dalam penelitian ini dibantu dengan software Microsoft Excel 2010 dan RStudio. Tahapan analisis yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Mempersiapkan data klasifikasi kemiskinan rumah tangga yang terdiri dari variabel independen dan variabel dependen
2. Melakukan klasifikasi data rumah tangga miskin dan tidak miskin dengan metode *Support Vector Machines* dengan tahapan sebagai berikut :
  - a. Membagi data menjadi data *training* dan data *testing*
  - b. Menentukan fungsi kernel yang digunakan beserta nilai-nilai parameter
  - c. Menentukan parameter terbaik untuk masing-masing kernel dengan menerapkan metode *cross validation* yang diterapkan pada data *training*
  - d. Menentukan *hyperplane* dengan menggunakan parameter terbaik pada masing-masing kernel
  - e. Evaluasi hasil klasifikasi pada data *testing* untuk mengukur ketepatan klasifikasi
3. Melakukan klasifikasi data rumah tangga miskin dengan metode *Classification and Regression Trees* dengan tahapan sebagai berikut :
  - a. Membagi data menjadi data *training* dan data *testing*
  - b. Menentukan simpul awal untuk pohon klasifikasi CART
  - c. Pembentukan pohon klasifikasi
  - d. Pemangkasan pohon, jika dinilai pohon klasifikasi yang dihasilkan terlalu kompleks
  - e. Melakukan interpretasi terhadap pohon klasifikasi yang terbentuk dan mengukur ketepatan klasifikasinya.
4. Membandingkan ketepatan klasifikasi yang diperoleh dari SVM dan CART

### 4. HASIL DAN PEMBAHASAN

#### 4.1. Analisis Deskriptif

Analisis deskriptif digunakan untuk memperoleh gambaran data secara umum, dari 795 amatan rumah tangga didapat 87% atau sebanyak 693 rumah tangga di Kabupaten Wonosobo adalah rumah tangga tidak miskin, sedangkan sisanya 13% atau sebanyak 102 merupakan rumah tangga miskin.

#### 4.2. Klasifikasi Status Kemiskinan Rumah Tangga dengan Metode *Support Vector Machines*

Pada klasifikasi status rumah tangga di Kabupaten Wonosobo pada tahun 2018 dengan metode SVM kali ini dititik beratkan pada penggunaan kernel *Radial Basis Function* (RBF). Parameter C yang akan dicobakan adalah 0.001, 0.01, 0.1, 1, 10, dan 100, sedangkan untuk parameter  $\gamma$  (gamma) yang akan dicobakan adalah 0.5, 0.6, 0.7, 0.8, dan 0.9. Data penelitian akan dibagi menjadi data *training* dan data *testing* dengan proporsi 80:20, kemudian diuji menggunakan *10-Fold Cross Validation* sebagai alat untuk menyeleksi parameter terbaik, diperoleh hasil sebagai berikut:

**Tabel 3.** Nilai Error Klasifikasi Setiap Parameter C dan  $\gamma$

Cost	Gamma	Error	Cost	Gamma	Error	Cost	Gamma	Error
0.001	0.5	0.13512	10	0.6	0.17468	0.1	0.8	0.13512
0.01		0.13512	100		0.20613	<b>1</b>		<b>0.12728</b>
0.1		0.13512	0.001	0.7	0.13512	10		0.18251
1		0.13046	0.01		0.13512	100	0.21084	
10		0.16059	0.1		0.13512	0.001	0.13512	
100		0.21081	1		0.12884	0.01	0.13512	
0.001		0.6	0.13512		10	0.1778	0.1	0.13512
0.01	0.13512		100	0.20766	1	0.12889		
0.1	0.13512		0.001	0.8	0.13512	10	0.1841	
1	0.12887		0.01		0.13512	100	0.21399	

Berdasarkan Tabel 3 diketahui bahwa nilai parameter terbaik yang dapat digunakan pada model SVM dengan fungsi kernel RBF adalah  $C = 1$  dan  $\gamma = 0.8$  dengan error terkecil sebesar 0.12728. diperoleh *hyperplane* :  $f(z) = \text{sign}(\sum_{i=1}^p \alpha_i y_i K(x_i, z) - 0.87005$

Matriks konfusi yang terbentuk adalah :

**Tabel 4.** Matriks Konfusi Klasifikasi dengan Metode SVM

Klasifikasi		Kelas Prediksi	
		Miskin	Tidak Miskin
Kelas Data	Miskin	0	16
	Tidak Miskin	0	143

Berdasarkan Tabel 4, didapatkan nilai akurasi sebagai berikut :

$$\begin{aligned} \text{Akurasi klasifikasi} &= \frac{0+143}{0+0+16+143} \\ &= 0.8994 \end{aligned}$$

Hasil evaluasi didapatkan nilai akurasi dari metode klasifikasi SVM dengan fungsi kernel RBF adalah 0.8994 atau 89,94% dengan laju error sebesar 0.1006 atau 10.06%. performansi SVM pada data rumah tangga di Kabupaten Wonosobo ini tidak dapat bekerja dengan baik untuk memprediksi rumah tangga miskin, namun sangat baik untuk memprediksi rumah tangga tidak miskin.

#### 4.3. Klasifikasi Status Kemiskinan Rumah Tangga dengan Metode *Classification and Regression Trees*

Hasil analisis status kemiskinan rumah tangga menggunakan metode CART akan menghasilkan suatu pohon klasifikasi yang dapat mengidentifikasi status kemiskinan rumah tangga dengan variabel yang mempengaruhi. Sebelumnya perlu dilakukan pembagian data

menjadi dua bagian, yang akan dibagi menjadi data *training* dan data *testing* dengan proporsi 80:20.

Berdasarkan hasil pengujian, diperoleh informasi bahwa variabel-variabel independen yang mempengaruhi variabel dependen diantaranya adalah variabel jenjang pendidikan tertinggi kepala rumah tangga ( $X_1$ ), kepemilikan fasilitas buang air besar ( $X_5$ ), sumber air utama yang digunakan rumah tangga untuk minum ( $X_6$ ), kepemilikan sepeda motor ( $X_9$ ), kepemilikan kulkas ( $X_{10}$ ), dan luas lantai bangunan tempat tinggal ( $X_{11}$ ) dengan masing-masing tingkat kepentingan variabel disajikan dalam tabel berikut:

**Tabel 5.** Tingkat Kepentingan Variabel pada CART

Variabel	Tingkat Kepentingan (%)	Variabel	Tingkat Kepentingan (%)	Variabel	Tingkat Kepentingan (%)
$X_1$	16.74%	$X_5$	9.65%	$X_9$	13.56%
$X_2$	2.87%	$X_6$	5.71%	$X_{10}$	10.03%
$X_3$	3.00%	$X_7$	0.00%	$X_{11}$	23.42%
$X_4$	0.18%	$X_8$	0.72%		

sisanya sebesar 14.13% dipengaruhi oleh variabel diluar penelitian.

Hasil dari pohon klasifikasi yang terbentuk, terdapat 29 simpul, 15 diantaranya merupakan simpul anak, yang berarti terdapat 15 karakteristik status kemiskinan rumah tangga di Kabupaten Wonosobo pada tahun 2018. Matriks konfusi yang terbentuk adalah :

**Tabel 6.** Matriks Konfusi dengan Metode CART

Klasifikasi		Kelas Prediksi	
		Miskin	Tidak Miskin
Kelas Data	Miskin	1	15
	Tidak Miskin	2	141

Berdasarkan Tabel 6, didapatkan nilai akurasi sebagai berikut :

$$\begin{aligned} \text{Akurasi klasifikasi} &= \frac{1+141}{1+2+15+141} \\ &= 0.8931 \end{aligned}$$

Hasil evaluasi didapatkan nilai akurasi dari metode klasifikasi CART adalah 0.8931 atau 89.31% dengan laju error sebesar 0.1069 atau 10.69%. Klasifikasi CART mampu memprediksi rumah tangga tidak miskin sebagai rumah tangga tidak miskin dengan baik, namun kurang baik jika diterapkan untuk memprediksikan rumah tangga miskin.

#### 4.4. Perbandingan Ketepatan Klasifikasi metode SVM dengan CART

Berdasarkan hasil evaluasi klasifikasi yang telah dilakukan mendapatkan hasil yang berbeda pada kedua metode. Hal ini dikarenakan pembangkitan data yang dilakukan secara acak pada masing-masing metode juga berbeda. Hasil evaluasi klasifikasi SVM dan CART adalah sebagai berikut :

**Tabel 7.** Perbandingan Evaluasi Klasifikasi SVM-RBF dengan CART

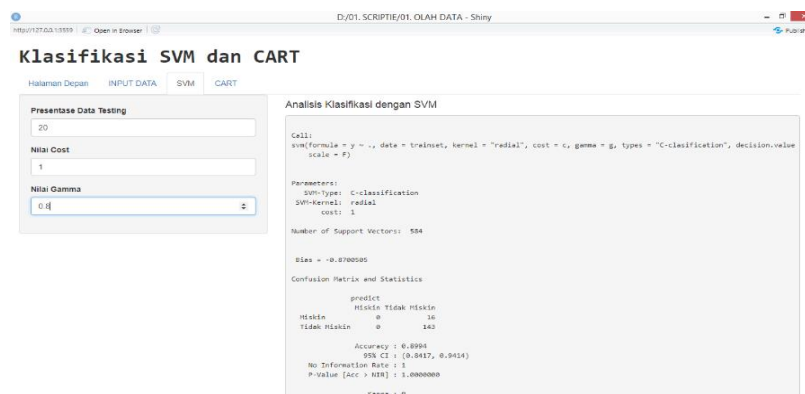
Proporsi Data <i>Testing</i>	Ukuran Performansi	
	SVM	CART
10%	91,14%	88.61%
20%	89.94%	89.31%
30%	89.5%	87.39%
40%	88.68%	83.02%
<b>Rata-rata Akurasi</b>	<b>89.82%</b>	<b>87.08%</b>



Berdasarkan Tabel 7 diperoleh informasi bahwa, klasifikasi terbaik untuk data kemiskinan rumah tangga di Kabupaten Wonosobo pada tahun 2018 adalah dengan menggunakan metode SVM fungsi kernel RBF. Dari data proporsi data training 90:10, 80:20, 30:70, 40:60 diperoleh rata-rata akurasi klasifikasi sebesar 89.82%.

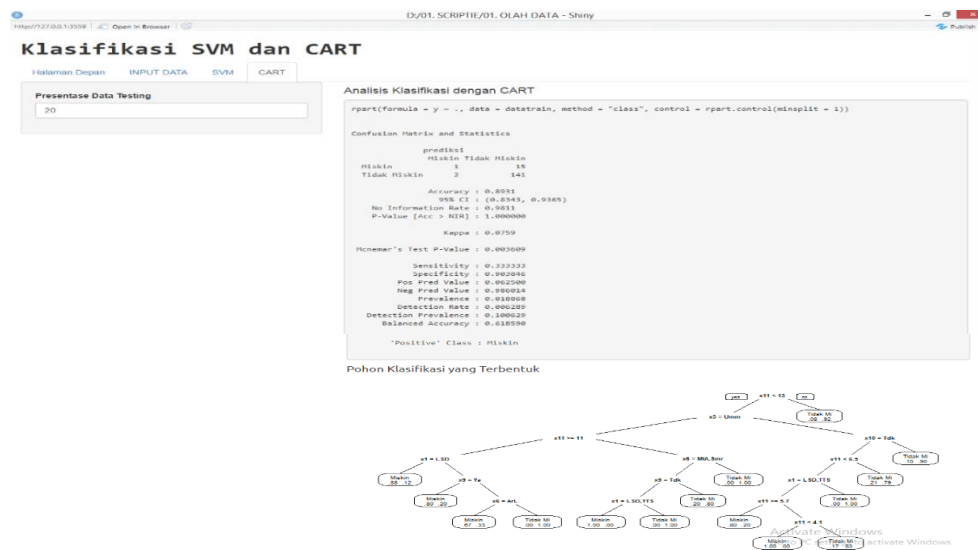
#### 4.5. Graphical User Interface (GUI) dengan R-Shiny

Graphical User Interface (GUI) R untuk analisis klasifikasi SVM dan CART dibangun dengan menggunakan bantuan *package shiny*. Pada penggunaannya, file data yang akan diklasifikasikan diimpor dengan ekstensi file '.csv', impor data dilakukan dengan klik tombol "Upload File". Setelah impor data, kemudian beralih ke sub menu SVM dan input proporsi data *testing*, nilai  $C$  dan  $\gamma$  sebagai parameter kernel RBF. Nilai parameter  $C$  dibatasi dari 0 sampai 1000, dan nilai parameter  $\gamma$  dibatasi dari 0 sampai 10. Pada contoh digunakan data *testing* 20%,  $C=1$ , dan  $\gamma=0.8$ . sehingga diperoleh hasil sebagai berikut:



Gambar 2. Output pada tab SVM

Setelah analisis klasifikasi dengan SVM selesai, kemudian beralih ke sub menu CART untuk analisis menggunakan metode CART. Pada analisis klasifikasi CART, input proporsi data testing yang akan digunakan dan tunggu hasilnya. Proporsi data testing yang digunakan dibatasi dari 10% sampai dengan 40%. Pada contoh digunakan data *testing* 20%, sehingga diperoleh hasil sebagai berikut:



Gambar 3. Output pada tab CART

## 5. KESIMPULAN

Berdasarkan analisis yang telah dilakukan, bahwa klasifikasi rumah tangga di Kabupaten Wonosobo pada tahun 2018 dengan proporsi data *training* : data *testing* 80:20 diperoleh persamaan fungsi kernelnya  $f(z) = \text{sign}(\sum_{i=1}^p \alpha_i y_i K(\mathbf{x}_i, \mathbf{z}) - 0.87005)$  dengan ketepatan klasifikasi untuk metode SVM yaitu sebesar 89.94%, sedangkan ketepatan klasifikasi untuk metode CART sebesar 89.31%. Metode klasifikasi terbaik pada data status kemiskinan rumah tangga di Kabupaten Wonosobo pada tahun 2018 adalah dengan menggunakan metode SVM, diperoleh rata-rata akurasi klasifikasi tertinggi sebesar 89.82% dengan empat kali percobaan. Kedua metode SVM dan CART tidak dapat bekerja dengan baik untuk memprediksi kelas minoritas, namun dapat bekerja dengan baik pada kelas mayoritas. Hasil pengujian klasifikasi dengan SVM dan CART juga ditunjukkan dalam GUI yang telah dibangun dengan menampilkan output yang valid. Pada penelitian selanjutnya dapat ditambahkan sebuah metode untuk mengatasi imbalance data, dengan harapan dapat meningkatkan performa klasifikasi SVM dan CART.

## DAFTAR PUSTAKA

1. Hsu, C. W., Chang, C. C., Lin, C. J. 2003. *A Practical Guide to Support Vector Classification*. Department of Computer Science National Taiwan University.
2. Kuswanto, H., Mubarak, R. 2019. *Classification of Cancer Drug Compounds of Radiation Protection Optimazation Using CART*. Jurnal Elsevier. Procedia Compute Science No. 161 : Hal 458-465.
3. Nugroho, A.S., Witarto, A.B., Handoko, D. 2003. *Support Vector Machines : Teori dan Aplikasinya dalam Bioinformatika*.
4. Nugroho, Heru. 1995. *Kemiskinan, Ketimpangan dan Kesenjangan*. Yogyakarta: Aditya Media.
5. Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta : Penerbit Andi.
6. Roger, J. Lewis, M.D. 2000. *An Introduction to Classification and Regression Tree (CART) Analysis*. Annual Meeting of the Society for Academic Emergency Medicine in San Fransisco. California : Departement of Emergency Medicine.
7. Santosa, B. 2007. *Data Mining, Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Jakarta : Graha Ilmu.