

KLASIFIKASI DATA PERKAWINAN ANAK DI MALUKU UTARA DENGAN METODE *KERNEL REGRESSION* DAN *SUPPORT VECTOR MACHINE*

Child Marriage Data Classification in Maluku Utara Using Kernel Regression and Support Vector Machine

Arifin M. Kahar*

Badan Pusat Statistik Provinsi Maluku Utara
Jl. Stadion No. 65, Ternate, 97712, Indonesia

Corresponding author e-mail: *ipinputera@gmail.com

Abstrak

Salah satu tujuan *Sustainable Development Goals* (SDGs) yang ingin dicapai pada 2030 adalah menghapus semua praktek perkawinan anak. Di Maluku Utara, kasus perkawinan anak pada perempuan masih cukup tinggi sesuai dengan laporan Badan Pusat Statistik (BPS) yaitu sebesar 14,36 persen pada 2019, melebihi angka nasional yaitu 10,82 persen. Sehingga, informasi mengenai determinan perkawinan anak diperlukan Pemerintah untuk melaksanakan program yang bertujuan untuk menekan kasus. Untuk itu, metode klasifikasi *Kernel Regression* dan *Support Vector Machine* (SVM) dapat digunakan untuk mengetahui determinannya. Hasil penelitian menunjukkan bahwa metode SVM memberikan ketepatan klasifikasi yang lebih tinggi yaitu 99,17 persen pada 70% *data training* dan 100 persen pada 30% *data testing* dibandingkan metode *Kernel Regression*. Melalui SVM, diperoleh determinan perkawinan anak di Maluku Utara yaitu, lama sekolah, status pekerjaan, akses terhadap internet, daerah tempat tinggal, jumlah ART, lama sekolah KRT yang telah diselesaikan dan pengeluaran per kapita rumah tangga.

Kata Kunci : Umur Kawin Pertama, Perkawinan Anak, *Kernel Regression*, *Support Vector Machine*.

Abstract

One of the aims of *Sustainable Development Goals* (SDGs) wants to be achieving in 2030 is eliminating all practices of child marriage. Based on the BPS-Statistics Indonesia, child marriage cases in Maluku Utara were high of 14.36 percent, exceeds the national of 10.82 percent in 2019. Thus, information about the determinants of child marriage is need to driving the programs by the Government to eliminate cases. Therefore, to find out the determinants can be used by *Kernel Regression* and *Support Vector Machine* (SVM) methods. The results showed that the SVM method provided the highest classification accuracy of 99.17 percent on 70% data training and 100 percent on 30% data testing compared to the *Kernel Regression*. Through the SVM, the determinants of child marriage in Maluku Utara were years of schooling, employment status, internet access, residence area, household members, years of schooling of the head of household, and per capita expenditure.

Keywords: Age at First Marriage, Child Marriage, *Kernel Regression*, *Support Vector Machine*.

Article info:

Submitted: 18th February 2021

Accepted: 02nd July 2021

How to cite this article:

A. M. Kahar, "KLASIFIKASI DATA PERKAWINAN ANAK DI MALUKU UTARA DENGAN METODE *KERNEL REGRESSION* DAN *SUPPORT VECTOR MACHINE*", *BAREKENG: J. Il. Mat. & Ter.*, vol. 15, no. 03, pp. 401-408, September 2021.



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).
Copyright © 2021 Arifin M. Kahar

1. PENDAHULUAN

Salah satu ukuran fertilitas yang sering digunakan pada analisis kependudukan adalah Umur Kawin Pertama (UKP). Menurut Badan Pusat Statistik, UKP adalah umur pertama kali melakukan hubungan suami istri [1]. Pada *Sustainable Development Goals* (SDGs), UKP menjadi salah satu prioritas pembangunan yang tertuang dalam target 5.3 yaitu menghapus semua praktik berbahaya seperti perkawinan usia anak, perkawinan dini dan paksa serta sunat perempuan. Selanjutnya, target tersebut didefinisikan melalui indikator 5.3.1 yaitu proporsi perempuan umur 20-24 tahun yang berstatus kawin atau berstatus hidup bersama sebelum umur 15 tahun dan sebelum umur 18 tahun. Penduduk yang melakukan perkawinan pada usia kurang dari 18 tahun disebut sebagai perkawinan anak [2].

Badan Pusat Statistik (BPS) melaporkan bahwa rata-rata UKP penduduk Maluku Utara 2015 mencapai 26 tahun pada laki-laki dan 23 tahun pada perempuan [3]. Lebih spesifik, BPS mencatat bahwa penduduk perempuan Maluku Utara berusia 10 tahun ke atas dengan UKP kurang dari 18 tahun mencapai 12,15 persen [2]. Selanjutnya, *baseline* indikator 5.3.1 SDGs menunjukkan bahwa penduduk perempuan Maluku Utara berusia 20-24 tahun dengan UKP kurang dari 18 tahun mencapai 14,36 persen pada 2019 [2]. Angka ini lebih besar dari angka nasional yang mencapai 10,82 persen [2]. Data tersebut mengindikasikan bahwa masih banyak perempuan di Maluku Utara yang melangsungkan perkawinan pertamanya pada usia anak.

Determinan perkawinan anak umumnya dapat ditentukan dari karakteristik sosial ekonomi individu dan rumah tangga. Marshaan dkk. menjelaskan bahwa karakteristik sosial ekonomi rumah tangga yang dominan berpengaruh terhadap perkawinan usia muda di Indonesia adalah akses terhadap internet, pengeluaran per kapita rumah tangga, pendidikan Kepala Rumah Tangga (KRT), jumlah anak yang mengkases pendidikan, fasilitas perumahan dan kepemilikan jaminan Kesehatan [4]. Widiyanti dan Yuhana juga menyatakan bahwa karakteristik sosial ekonomi rumah tangga yang mempengaruhi keputusan perempuan untuk melakukan perkawinan pada usia muda di Indonesia adalah tipe daerah tempat tinggal, status bekerja dan jenis pekerjaan, tingkat pendidikan, tingkat pendidikan orang tua, UKP orang tua dan status kemiskinan [5]. Penelitian serupa juga dilakukan oleh Qibtiyah yang menyatakan bahwa tipe daerah tempat tinggal dan tingkat pendidikan memiliki pengaruh yang besar terhadap perkawinan muda pada perempuan [6]. Penelitian lain oleh Putri dan Nelonda disebutkan bahwa tipe daerah tempat tinggal, tingkat pendidikan KRT, dan jumlah Anggota Rumah Tangga (ART) berpengaruh signifikan terhadap perkawinan usia muda perempuan di Sumatera Barat [7]. Sejalan dengan hasil penelitian di Indonesia, Mpilambo dkk. juga menyatakan bahwa indikator yang menjadi penentu perkawinan muda pada perempuan di Republik Congo diantaranya adalah status pekerjaan, tipe daerah tempat tinggal, akses terhadap media dan tingkat kesejahteraan keluarga [8]. Berdasarkan data aktual dan hasil penelitian sebelumnya, maka penelitian ini bertujuan untuk mengklasifikasikan kasus perkawinan anak pada perempuan di Maluku Utara.

2. METODE PENELITIAN

2.1 Sumber Data

Data yang digunakan bersumber dari BPS Provinsi Maluku Utara melalui Survei Sosial Ekonomi Nasional (Susenas) 2019. Mengacu pada Indikator 5.3.1 SDGs [2], maka unit analisis penelitian ini dibatasi pada perempuan berusia 15-24 tahun yang berstatus kawin atau pernah kawin. Adapun variabel yang akan digunakan disajikan pada Tabel 1, berikut:

Tabel 1. Deskripsi Variabel Penelitian

| Notasi | Deskripsi Variabel | Skala Ukur | Kategori |
|--------|-------------------------|------------|---|
| Y | Umur Kawin Pertama | Nominal | 1 = UKP < 18 Tahun; 0 = UKP \geq 18 Tahun |
| X_1 | Lama Pendidikan | Kontinu | - |
| X_2 | Status Bekerja | Ordinal | 1 = Tidak Bekerja; 0 = Bekerja |
| X_3 | Akses Terhadap Internet | Nominal | 1 = Ya; 0 = Tidak |
| X_4 | Daerah Tempat Tinggal | Nominal | 1 = Perdesaan; 0 = Perkotaan |
| X_5 | Jumlah ART | Kontinu | - |
| X_6 | Lama Pendidikan KRT | Kontinu | - |
| X_7 | Pengeluaran per Kapita | Kontinu | - |

2.2 Kernel Regression

Misalkan $f(\cdot)$ dan $m(\cdot)$ merupakan fungsi marjinal gabungan dari (X, Y) dan X , dimana Y adalah variabel respon biner dan X merupakan variabel prediktor bertipe kontinu dan kategori (terurut dan tidak terurut), maka fungsi densitas Y bergantung kepada X didefinisikan sebagai berikut [9]:

$$g(y|x) = \frac{f(x,y)}{m(x)}, \quad (1)$$

Dimana fungsi ini tidak diketahui bentuknya, sehingga untuk melakukan pendugaan, persamaan (1) dapat didefinisikan kembali menjadi [9], berikut:

$$\hat{g}(y|x) = \frac{\hat{f}(x,y)}{\hat{m}(x)}. \quad (2)$$

Penduga $m(\cdot)$ dapat didekati dengan fungsi kernel. Apabila variabel X merupakan gabungan beberapa tipe data yaitu kontinu dan kategori (terurut dan tidak terurut), maka digunakan penduga *Generalized Product Kernel* [9].

Misalkan $X = (X^c, X^d, \tilde{X}^d)$ adalah komponen variabel prediktor bertipe kontinu sebanyak p , tipe kategori tidak terurut sebanyak q , dan tipe kategori terurut sebanyak r , maka penduga $m(\cdot)$ untuk X yang dinotasikan dengan $x = (x^c, x^d, \tilde{x}^d)$ dirumuskan sebagai berikut [9]:

$$\begin{aligned} \hat{m}(x) &= \hat{m}(x^c, x^d, \tilde{x}^d) \\ &= \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^p W(X_{ij}^c, x_j^c) \prod_{j=1}^q l(X_{ij}^d, x_j^d) \prod_{j=1}^r \tilde{l}(\tilde{X}_{ij}^d, \tilde{x}_j^d), \end{aligned} \quad (3)$$

Dimana untuk setiap variabel prediktor dengan tipe berbeda digunakan fungsi kernel yang juga berbeda yaitu,

a. Untuk variabel kontinu x_j^c , fungsi kernel $W(\cdot)$ yang digunakan adalah [9]:

$$W(X_{ij}^c, x_j^c) = \frac{1}{h_j} K\left(\frac{X_{ij}^c - x_j^c}{h_j}\right), \quad (4)$$

dengan $K(\cdot)$ adalah fungsi kernel dan h_j adalah parameter pemulus (*bandwidth*). Jenis fungsi kernel yang dapat digunakan diantaranya adalah [9]:

- *Uniform*, $K(X_{ij}^c, x_j^c) = \frac{1}{2} \mathbf{1}_{[-1,1]}(X_{ij}^c - x_j^c)$;
- *Epanechnikov*, $K(X_{ij}^c, x_j^c) = \frac{3}{4} \left(1 - (X_{ij}^c - x_j^c)^2\right) \mathbf{1}_{[-1,1]}(X_{ij}^c - x_j^c)$;
- *Gaussian*, $K(X_{ij}^c, x_j^c) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(X_{ij}^c - x_j^c)^2}{2}\right)$.

b. Untuk variabel tidak terurut x_j^d dengan c_j kategori, digunakan fungsi kernel Aitchison & Aitken [9], [10], yaitu:

$$l(X_{ij}^d, x_j^d) = \begin{cases} 1 - \lambda_j & X_{ij}^d = x_j^d \\ \frac{\lambda_j}{c_j - 1} & X_{ij}^d \neq x_j^d \end{cases} \quad (5)$$

dengan λ_j adalah *bandwidth* dengan interval $[0, (c_j - 1)/c_j]$. Jika $\lambda_j = 0$ maka kernel l menjadi indikator fungsi. Saat $\lambda_j = (c_j - 1)/c_j$ maka kernel $l(X_{ij}^d, x_j^d) = 1/c_j$ untuk semua nilai X_{ij}^d dan x_j^d .

c. Untuk variabel bertipe kategori terurut, digunakan fungsi kernel Wang & Van Ryzin yaitu [9]:

$$\tilde{l}(\tilde{X}_{ij}^d, \tilde{x}_j^d) = \begin{cases} 1 & \tilde{X}_{ij}^d = \tilde{x}_j^d \\ \gamma_j^{|\tilde{X}_{ij}^d - \tilde{x}_j^d|} & \tilde{X}_{ij}^d \neq \tilde{x}_j^d \end{cases} \quad (6)$$

dengan γ_j adalah *bandwidth* dengan interval $[0,1]$. Saat $\gamma_j = 0$, maka kernel \tilde{l} menjadi indikator fungsi dan saat $\gamma_j = 1$, maka kernel \tilde{l} merupakan fungsi pembobot seragam.

Selanjutnya, pendugaan terhadap fungsi densitas gabungan $f(\cdot)$ untuk (X, Y) yang dinotasikan dengan $x = (x^c, x^d, \tilde{x}^d, y)$ dirumuskan sebagai berikut [9]:

$$\begin{aligned}\hat{f}(x, y) &= \hat{f}(x^c, x^d, \tilde{x}^d, y) \\ &= \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^p W(X_{ij}^c, x_j^c) \prod_{j=1}^q l(X_{ij}^d, x_j^d) \prod_{j=1}^r \tilde{l}(\tilde{X}_{ij}^d, \tilde{x}_j^d) \times l(Y_i^d, y).\end{aligned}\quad (7)$$

Pada pendugaan $m(x)$, akan melibatkan pemilihan *bandwidth* h_j untuk $j = 1, \dots, p$, λ_j untuk $j = 1, \dots, q$, γ_j untuk $j = 1, \dots, r$. Pemilihan *bandwidth* ini dapat dilakukan dengan beberapa cara namun pada penelitian ini digunakan metode *Normal-Reference* (NR) [11] dan *Least Square Cross Validation* (LSCV) [9]. Pada metode NR, *bandwidth* optimal untuk variabel prediktor bertipe kategori ditetapkan 0, sedangkan pada variabel prediktor bertipe kontinu, *bandwidth* optimal dicari melalui formula $h_j = 1,06\sigma_j n^{-1/(2P+L)}$ dimana σ_j merupakan nilai yang didefinisikan dari $\min(\text{standar deviasi}, \text{rata-rata deviasi mutlak}/1,4826, \text{jarak antar kuartil}/1,349)$, n adalah jumlah observasi, P adalah orde kernel dan L adalah jumlah variabel kontinu. Pada metode LSCV, digunakan formula berikut [9]:

$$\sum_{x_d} \int \{\hat{g}(y|x) - g(y|x)\}^2 m(x) M(x^c) dx dy \quad (8)$$

dengan $M(\cdot)$ adalah fungsi pembobot. Kemudian, untuk mendapatkan *bandwidth* optimal maka dilakukan dengan menimalkan fungsi tersebut.

2.3 Support Vector Machine

Support Vector Machine (SVM) umumnya digunakan untuk melakukan prediksi pada data klasifikasi maupun pada model regresi. SVM dapat dilakukan pada variabel respon bertipe biner, multinomial, dan kontinu. Data variabel respon bertipe biner dikenal dengan klasifikasi dimana cara kerjanya yaitu menemukan suatu fungsi pemisah (*hyperplane*) yang optimal antara dua kelas secara linier maupun nonlinier.

Pada kasus data respon biner, fungsi pemisah yang akan dicari dapat didefinisikan sebagai berikut [12]:

$$f(x; w; b) = \langle w, x \rangle + b \quad (9)$$

dengan w dan b adalah parameter fungsi yang akan dicari nilainya dan \langle, \rangle merupakan notasi dari perkalian dalam dari dua vektor. Fungsi pemisah tersebut tidak terbatas jumlahnya, sehingga untuk mendapatkan fungsi pemisah terbaik dilakukan dengan memaksimalkan margin atau jarak antara dua kelas data yang berbeda. Misalkan suatu set data (x_i, y_i) , $i = 1, \dots, l$, dimana $y_i \in \{-1, +1\}$ merupakan kelas data dan harus memenuhi batasan $y(\langle w, x \rangle + b) \geq 1$, $\forall i = 1, \dots, l$, maka, fungsi pemisah optimal dapat diperoleh dengan meminimalkan margin $\min \frac{1}{2} \|w\|^2$ atau memaksimalkan $\|w\|^2$. Apabila kelas data $y_i = +1$, maka fungsi pembatasnya menjadi $\langle w, x \rangle + b \geq 1$. Sebaliknya, apabila kelas data $y_i = -1$, maka fungsi pembatasnya menjadi $\langle w, x \rangle + b \leq 1$. Pada kasus tertentu, beberapa data mungkin tidak dapat dikelompokkan dengan tepat, sehingga formulanya dapat dirumuskan kembali menjadi [12]:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \quad (10)$$

dengan ξ_i merupakan variabel *slack* yang berfungsi untuk mengontrol kesalahan klasifikasi dari pembatas $y(\langle w, x \rangle + b) \geq 1$ dengan cara memberi penalti C pada data yang tidak memenuhi kriteria tersebut. Untuk mengoptimalkan fungsinya dapat dilakukan dengan pendekatan *Lagrange Multiplier*, sehingga fungsinya diformulasikan menjadi [12]:

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i [y_i(\langle w, x_i \rangle + b) - 1] \quad (11)$$

dimana fungsi tersebut dapat optimal ketika turunan pertama dari w dan b adalah $w = \sum_{i=1}^l y_i \alpha_i x_i$ dan $\sum_{i=1}^l \alpha_i y_i = 0$. Kemudian, untuk mencapai kondisi yang optimal maka digunakan pendekatan *duality problem* yaitu turunan pertama dari $L(w, b, \alpha)$ terhadap w dan b disamakan dengan 0, sehingga fungsi akhir yang diperoleh sebagai berikut [12]:

$$\max W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle \quad (12)$$

dengan $\sum_{i=1}^l y_i \alpha_i = 0$, $\alpha_i \geq 0$, $\forall i = 1, \dots, l$.

Pada beberapa kasus, pemisahan pada dua atau lebih kelas data sering tidak dapat dilakukan dengan fungsi pemisah linier, sehingga untuk mengatasi permasalahan ini digunakan fungsi pemisah nonlinier dengan metode kernel. Pada metode kernel, suatu data x akan dipetakan pada dimensi yang lebih tinggi melalui ϕ yaitu, $\phi: x \rightarrow \phi(x)$. Umumnya, $\phi(x)$ sering kali tidak dapat dihitung sehingga dapat didekati dengan

teorema Mercer bahwa suatu fungsi $K(x, z)$ merupakan perkalian dalam dari dua vektor x dan z [12]. Pada kondisi ini, $K(x, z) = \langle \phi(x), \phi(z) \rangle$ dimana K menyatakan fungsi kernel yang dapat di hitung. Selanjutnya, fungsi pemisah optimalnya dapat diformulasikan sebagai berikut [12]:

$$\max W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j K(x_i, x_j) \quad (13)$$

dengan $\sum_{i=1}^l y_i \alpha_i = 0$, $\alpha_i \geq 0$, $\forall i = 1, \dots, l$ dan $f(x) = \sum_{i=1}^l y_i \alpha_i K(x_i, x)$ menyatakan fungsi klasifikasi. Beberapa fungsi kernel yang dapat digunakan adalah [12], [13]:

- *Linear*: $K(x, z) = \langle x, z \rangle$;
- *Polynomial*: $K(x, z) = (\langle x, z \rangle + 1)^p$;
- *Radial Basis Function (RBF)/Gaussian*: $K(x, z) = \exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right)$;
- *Sigmoid*: $K(x, z) = \tanh(\beta \langle x, z \rangle + \beta_1)$,

dengan $\beta, \beta_1 \in \mathfrak{R}$.

Salah satu cara untuk memilih fungsi kernel yang terbaik adalah dengan metode validasi silang (*cross validation*). Pada penelitian ini, digunakan metode *K-Fold Cross Validation* sebanyak 10 kali pengulangan seperti yang dikembangkan oleh Bengio dan Grandvalet [14]. Selanjutnya, untuk mengevaluasi ketepatan klasifikasi baik pada model *Kernel Regression* maupun pada model SVM digunakan nilai akurasi serta nilai sensitifitas dan spesifitas melalui tabel *confusion matrix*, berikut [9]:

Tabel 2. Notasi *Confusion Matrix*

| Observasi | Prediksi | | Total |
|-----------|----------|----|-------|
| | 0 | 1 | |
| 0 | TN | FP | n_0 |
| 1 | FN | TP | n_1 |
| Total | | | n |

Ukuran ketepatan klasifikasi diformulasikan, sebagai berikut [9]:

$$\text{Akurasi} = (TN + TP)/n,$$

$$\text{Sensifitas} = TP/n_1,$$

$$\text{Spesifitas} = TN/n_0.$$

3. HASIL DAN PEMBAHASAN

Karakteristik data yang digunakan pada penelitian ini dapat dilihat pada Gambar 1. Pada 2019, sekitar 36,68 persen perempuan di Maluku Utara berusia 15-24 tahun yang pernah kawin telah melangsungkan perkawinan pertamanya pada usia kurang dari 18 tahun (anak), sisanya 63,32 persen melakukan perkawinan pertama pada usia 18 tahun ke atas.



Gambar 1. Proporsi Perkawinan Anak pada Perempuan di Maluku Utara, 2019

Untuk melakukan analisis klasifikasi, data dipartisi menjadi data *training* dan data *testing*. Pada penelitian ini, dilakukan dua kali partisi dengan perbandingan 70:30 dan 80:20. Selanjutnya, dilakukan pengolahan data untuk metode *Kernel Regression* menggunakan program *R* versi 4.0.3 melalui *np Package*

[11], [15]. Adapun hasil pengolahan disajikan pada Tabel 3. Pada tabel tersebut, terlihat bahwa klasifikasi pada model regresi kernel dengan penduga *Generalized Product Kernel* menghasilkan akurasi klasifikasi tertinggi sebesar 100% pada 20% data *testing* dengan metode pemilihan *bandwidth* N-R pada orde kernel 4. Secara umum, metode N-R memberikan akurasi klasifikasi yang lebih tinggi, baik pada partisi 70:30 maupun 80:20. Selain itu, metode ini juga dapat memprediksi data masuk ke dalam kelas positif (sensifitas) dan kelas negatif (spesifitas) secara benar cukup tinggi pada kedua partisi data. Dengan demikian, dapat dikatakan bahwa model klasifikasi *Kernel Regression* dengan metode pemilihan *bandwidth Normal-Reference* dapat dipilih untuk mengklasifikasikan data perkawinan anak pada perempuan di Maluku Utara 2019.

Tabel 3. Ketepatan Klasifikasi Model *Kernel Regression*

| Partisi Data | Metode Pemilihan <i>Bandwidth</i> | Orde Kernel (d) | Akurasi (%) | Spesifitas (%) | Sensitifitas (%) | |
|-------------------------|-----------------------------------|-----------------|---------------|----------------|------------------|-------|
| 70% Data Training | LSCV | 2 | 78,45 | 86,94 | 65,00 | |
| | | 4 | 68,23 | 75,68 | 56,43 | |
| | N-R | 2 | 91,99 | 95,95 | 85,71 | |
| | | 4 | 92,82 | 95,05 | 89,29 | |
| | 30% Data Testing | LSCV | 2 | 71,79 | 99,06 | 14,00 |
| | | | 4 | 73,72 | 99,06 | 20,00 |
| N-R | | 2 | 96,79 | 99,06 | 92,00 | |
| | | 4 | 98,08 | 99,06 | 96,00 | |
| 80% Data Training | | LSCV | 2 | 83,33 | 96,28 | 59,31 |
| | | | 4 | 70,77 | 99,63 | 17,24 |
| | N-R | 2 | 91,55 | 98,14 | 79,31 | |
| | | 4 | 92,75 | 97,03 | 84,83 | |
| | 20% Data Testing | LSCV | 2 | 98,08 | 98,31 | 97,78 |
| | | | 4 | 90,38 | 88,14 | 93,33 |
| N-R | | 2 | 97,12 | 96,61 | 97,78 | |
| | | 4 | 100,00 | 100,00 | 100,00 | |

Selanjutnya, dilakukan pengklasifikasian dengan metode SVM melalui *e1071 Package* [16]. Adapun hasil pengolahannya disajikan pada Tabel 4. Pada tabel tersebut terlihat bahwa parameter optimal (*Cost* dan *Gamma*) yang diperoleh berbeda pada masing-masing jenis fungsi kernel, baik pada partisi data 70:30 maupun partisi data 80:20. Secara umum, terlihat bahwa model SVM dengan fungsi kernel RBF memiliki akurasi tertinggi pada partisi 70:30 maupun 80:20. Hal ini dapat dilihat dari akurasi klasifikasi pada data *training* sebesar 99,17% berbanding 99,03% dan pada data *testing* keduanya memiliki akurasi yang sama yaitu 100%. Selain itu, metode ini juga dapat mengklasifikasikan data untuk kelas positif dan kelas negatif secara benar dengan persentase yang sangat tinggi pada kedua partisi data.

Tabel 4. Ketepatan Klasifikasi Model SVM

| Partisi Data | Fungsi Kernel | <i>Cost</i> | <i>Gamma</i> | Akurasi (%) | Spesifitas (%) | Sensitifitas (%) |
|-------------------------|------------------|-------------|--------------|---------------|----------------|------------------|
| 70% Data Training | Linear | 1000 | 4 | 61,33 | 0,00 | 100,00 |
| | Polynomial (d=2) | 10 | 0,1 | 70,44 | 52,14 | 81,98 |
| | Polynomial (d=3) | 10 | 0,1 | 73,20 | 47,14 | 89,64 |
| | Sygmoid | 10 | 3 | 51,93 | 37,86 | 60,81 |
| | RBF | 100 | 6 | 99,17 | 98,57 | 99,55 |
| 30% Data Testing | Linear | 1000 | 4 | 67,95 | 0,00 | 100,00 |
| | Polynomial (d=2) | 10 | 0,1 | 75,64 | 30,00 | 97,17 |
| | Polynomial (d=3) | 10 | 0,1 | 75,00 | 26,00 | 98,11 |
| | Sygmoid | 10 | 3 | 50,00 | 22,00 | 63,21 |
| | RBF | 100 | 6 | 100,00 | 100,00 | 100,00 |
| 80% Data Training | Linear | 1 | 0,1 | 64,98 | 0,00 | 100,00 |
| | Polynomial (d=2) | 100 | 6 | 69,81 | 25,52 | 93,68 |
| | Polynomial (d=3) | 10 | 0,1 | 71,01 | 26,90 | 94,80 |
| | Sygmoid | 10 | 2 | 55,07 | 35,86 | 65,43 |

| | RBF | 10 | 5,5 | 99,03 | 98,62 | 99,26 |
|------------------------|------------------|-----|-----|--------|--------|--------|
| 20% Data Testing | Linear | 1 | 0,1 | 67,31 | 60,00 | 72,88 |
| | Polynomial (d=2) | 100 | 6 | 83,65 | 75,56 | 89,83 |
| | Polynomial (d=3) | 10 | 0,1 | 81,73 | 68,89 | 91,53 |
| | Sygmoid | 10 | 2 | 47,12 | 40,00 | 52,54 |
| | RBF | 10 | 5,5 | 100,00 | 100,00 | 100,00 |

Tabel 5. Perbandingan Akurasi Model *Kernel Regression* dan SVM (%)

| Partisi Data | | Kernel Regression | | SVM | |
|--------------|-----|-------------------|--------|----------------------|--------|
| 70:30 | 70% | N-R(d=4) | 92,82 | RBF(C=100, Gamma=6) | 99,17 |
| | 30% | N-R(d=4) | 98,08 | RBF(C=100, Gamma=6) | 100,00 |
| 80:20 | 80% | N-R(d=4) | 92,75 | RBF(C=10, Gamma=5.5) | 99,03 |
| | 20% | N-R(d=4) | 100,00 | RBF(C=10, Gamma=5.5) | 100,00 |

Pada Tabel 5, dapat dilihat bahwa pada partisi data 70:30 dan 80:20, model SVM memberikan akurasi klasifikasi yang lebih tinggi dibandingkan dengan model *Kernel Regression*. Terdapat dua kandidat model SVM karena memiliki nilai akurasi yang sama yaitu 100% baik pada data *testing* 30% maupun 20%. Namun, model yang dipilih adalah model pada partisi data 70:30 karena dapat memberikan akurasi tertinggi untuk data *training* (99,17%), sehingga model terpilih adalah model SVM dengan fungsi kernel RBF dan parameter $Cost=100$ dan $Gamma = 6$.

4. KESIMPULAN

Model SVM pada penelitian ini memberikan akurasi klasifikasi yang sangat tinggi dibandingkan dengan model *Kernel Regression* sehingga dapat dikatakan model SVM lebih cocok untuk mengklasifikasikan data perkawinan anak di Maluku Utara pada 2019. Melalui model SVM, diperoleh karakteristik sosial ekonomi individu dan rumah tangga yang menjadi determinan perkawinan anak pada perempuan usia 15-24 tahun yaitu, lama sekolah, status pekerjaan, akses terhadap internet, daerah tempat tinggal, jumlah ART, lama sekolah KRT yang telah diselesaikan dan pengeluaran per kapita rumah tangga.

DAFTAR PUSTAKA

- [1] Badan Pusat Statistik, Buku 4: Konsep dan Definisi Survei Sosial Ekonomi Nasional Maret 2019, Jakarta: Badan Pusat Statistik, 2019.
- [2] Badan Pusat Statistik, Indikator Tujuan Pembangunan Berkelanjutan Indonesia 2020, Jakarta: Badan Pusat Statistik, 2020.
- [3] Badan Pusat Statistik, Profil Kependudukan Provinsi Maluku Utara Hasil Supas 2015, Ternate: Badan Pusat Statistik Provinsi Maluku Utara, 2016.
- [4] J. N. Marshan, M. F. Rakhmadi dan M. Rizky, "Prevalence of Child Marriage and Its Determinants among Young Women in Indonesia," dalam *Child Poverty and Social Protection Conference*, Jakarta, 2013.
- [5] A. Widiyantara dan R. J. Yuhan, "Pengaruh Variabel Sosial Ekonomi Terhadap Perkawinan Usia Anak pada Wanita di Indonesia Tahun 2017," *Statistika*, vol. 19, no. 2, pp. 139-149, 2019.
- [6] M. Qibtiyah, "Faktor yang Mempengaruhi Perkawinan Muda Perempuan," *Jurnal Bimometrika dan Kependudukan*, vol. 3, no. 1, pp. 50-58, 2014.
- [7] D. Z. Putri dan S. Nelonda, "Socio-Economic Determinants of Age at First Marriage among Women and Early Marriage Women in West Sumatera," dalam *The 1st International Conference on Economics, Business and Accounting 2016*, Padang, 2016.
- [8] J. Mpilambo, S. Susumanappunni, O. Kanayo dan N. Stiegler, "Determinants of early marriage among young women in Democratic Republic of Congo," *Journal of Social Sciences*, vol. 52, no. 1-3, 2017.
- [9] C. Bontemps, J. S. Racine dan M. Simioni, "Nonparametric vs parametric binary choice models: An empirical investigation," dalam *Agricultural and Applied Economics Association (AAEA) Conferences*, Wisconsin, 2009.
- [10] Q. Li dan J. S. Racine, *Nonparametric Econometrics: Theory and Practice*, New Jersey: Princeton Univ. Press, 2007.
- [11] J. S. Racine dan T. Hayfield, "CRAN.R-project.org," 6 Februari 2021. [Online]. Available: <https://CRAN.R-project.org/package=np>. [Diakses 8 Januari 2021].

- [12] D. Fradkin dan I. Muchnik, "Support Vector Machines for Classification," *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 70, pp. 13-20, 2006.
- [13] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua dan A. Lopez, "A Comprehensive Survey on Support Vector Machine Classification: Applications, Challenges and Trends , vol. 408, pp. 189-215, 2020.," *Neurocomputing*, vol. 408, pp. 189-215, 2020.
- [14] Y. Bengio dan Y. Grandvalet, "No Unbiased Estimator of the Variance of K-Fold Cross-Validation," *Journal of Machine Learning Research*, vol. 5, pp. 1089-1105, 2004.
- [15] T. Hayfield dan J. S. Racine, "Nonparametric Econometrics: The np Package," *Journal of Statistical Software*, vol. 27, no. 5, pp. 1-32, 2008.
- [16] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, F. Leisch, C. Chang dan C. Lin, "CRAN.R-project.org," 14 Oktober 2020. [Online]. Available: [CRAN.R-project.org/package=e1071](https://cran.r-project.org/package=e1071). [Diakses 8 Januari 2021].