

Received: 10 Juni 2020

Revised: 25 Juni 2020

Accepted: 27 Juni 2020

Published: 30 Juni 2020

## **Penerapan Metode *Support Vector Machine (SVM)* dan *Backpropagation Neural Network (BPNN)* dalam Pengklasifikasian Pasangan Usia Subur di Jawa Barat**

Istiqomatul Fajriyah Yuliaty<sup>1, a)</sup>, Septie Wulandary<sup>2, b)</sup>, Pardomuan Robinson Sihombing<sup>3, c)</sup>

<sup>1</sup>*Badan Kependudukan dan Keluarga Berencana Nasional, Jalan Permata No. 1, Jakarta Timur*

<sup>2</sup>*Badan Pusat Statistik Provinsi Jambi, Jalan A Yani No.4, Telanaipura, Kota Jambi*

<sup>3</sup>*Badan Pusat Statistik, Jalan Dr. Sutomo No 6-8, Jakarta Pusat*

Email: <sup>a)</sup>[istiqomatul.fy@bkkbn.go.id](mailto:istiqomatul.fy@bkkbn.go.id), <sup>b)</sup>[septie@bps.go.id](mailto:septie@bps.go.id), <sup>c)</sup>[robinson@bps.go.id](mailto:robinson@bps.go.id)

### **Abstract**

The use of contraception is very vital for controlling the population, especially in West Java, because the population in West Java Province is the highest in Indonesia. The classification of contraceptive use in fertile age couples (PUS) based on determinants of contraceptive use is an interesting thing to study. This study examines the comparison of the Support Vector Machine (SVM) method with kernel and Neural Network functions with the backpropagation algorithm (BPNN) in classifying fertile aged couples based on contraceptive use (following the family planning program or not). The method that had a high accuracy value is the best in classifying fertile age couples based on the use of contraception. Both the BPNN classification and SVM classification can describe the classification of fertile age couples based on determinants of contraceptive use with the same level of accuracy, which is equal to 99.84%. It means that the two classification models are very good in predicting classes that follow family planning or not.

**Keywords:** family planning, West Java, SVM, BPPN.

### **Abstrak**

Penggunaan kontrasepsi menjadi sangat vital bagi pengendalian jumlah penduduk terutama di Jawa Barat karena jumlah penduduk di Provinsi Jawa Barat tertinggi di Indonesia. Klasifikasi penggunaan kontrasepsi pada pasangan usia subur (PUS) berdasarkan faktor determinan penggunaan kontrasepsi di Provinsi Jawa Barat merupakan hal yang menarik untuk dikaji. Penelitian ini mengkaji perbandingan metode *Support Vector Machine (SVM)* dengan fungsi kernel dan Neural Network (NN) dengan algoritma *backpropagation (BPNN)* dalam pengklasifikasian PUS berdasarkan penggunaan kontrasepsi (mengikuti program Keluarga Berencana/KB atau tidak) di Provinsi Jawa Barat. Metode yang memiliki nilai ketepatan yang tinggi merupakan metode terbaik dalam pengklasifikasian PUS berdasarkan penggunaan kontrasepsi. Kedua metode klasifikasi baik BPNN dan SVM Fungsi Kernel dapat menggambarkan klasifikasi PUS berdasarkan determinan penggunaan kontrasepsi di Provinsi Jawa Barat dengan tingkat akurasi yang sama, yaitu sebesar 99.84%. Hal ini berarti kedua model klasifikasi sangat baik dalam memprediksi kelas yang mengikuti program keluarga berencana atau tidak mengikuti program tersebut.

**Kata-kata kunci:** keluarga berencana, Jawa Barat, SVM, BPNN.

## PENDAHULUAN

Indonesia merupakan salah satu negara dengan jumlah penduduk tertinggi di dunia yaitu sebesar 268 juta jiwa. Berdasarkan data yang dirilis Badan Pusat Statistik, jumlah penduduk tertinggi di Indonesia adalah Provinsi Jawa Barat sebesar 48 juta jiwa atau 18.34%, dengan kepadatan penduduk tertinggi kedua di Indonesia setelah DKI Jakarta sebesar 1358/km<sup>2</sup> dan laju pertumbuhan penduduk sebesar 1,52%.

Program Kependudukan Keluarga Berencana dan Pembangunan Keluarga (KKBP) merupakan salah satu upaya yang telah dilakukan oleh pemerintah melalui Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) dalam menangani masalah kependudukan terutama dalam aspek kuantitas, yaitu dengan mengendalikan jumlah penduduk melalui penurunan fertilitas. Berdasarkan hasil Survei Demografi dan Kesehatan Indonesia tahun 2017, Provinsi Jawa Barat memiliki median umur kawin pertama wanita terendah se-Jawa, yaitu 21,3 tahun, begitupun angka fertilitas atau *total fertility rate* (TFR) di Jawa Barat tertinggi se-Jawa, yaitu sebesar 2,4. Akan tetapi, angka pemakaian kontrasepsi atau *contraceptive prevalence rate* (CPR) cukup tinggi di Jawa, yaitu 63,3%.

Penggunaan kontrasepsi menjadi sangat vital bagi pengendalian jumlah penduduk terutama di Jawa Barat karena jumlah penduduk di Provinsi Jawa Barat tertinggi di Indonesia. Klasifikasi penggunaan kontrasepsi pada Pasangan Usia Subur (PUS) berdasarkan faktor determinan penggunaan kontrasepsi di Provinsi Jawa Barat merupakan hal yang penting untuk dikaji. Penelitian sebelumnya tentang klasifikasi penggunaan kontrasepsi berdasarkan faktor determinan penggunaan kontrasepsi telah dilakukan antara lain oleh Johra (2018) yang melakukan penerapan non-linier SVM pada penggunaan alat kontrasepsi di Provinsi Maluku Utara, dan Suwardika (2016) melakukan pengelompokan dan klasifikasi penggunaan kontrasepsi di Indonesia.

Ketepatan dalam pengklasifikasian variabel sangat penting, metode klasifikasi yang baik adalah metode yang menghasilkan ketepatan yang tinggi (Hakim, 2016). Ada beberapa metode data mining untuk mengetahui ketepatan klasifikasi baik klasifikasi linear maupun non linear. Metode klasifikasi yang akan digunakan pada penelitian ini adalah Support Vector Machine (SVM) dan Neural Network. Metode SVM merupakan metode klasifikasi non parametrik yang tidak harus memenuhi asumsi dan distribusi tertentu (Sahitayakti dan Fithriasari, 2015), sedangkan metode *Neural Network* merupakan metode yang umum dalam pengolahan data non linear serta memiliki proses paralel yang kuat dalam menangani kesalahan toleransi.

Berdasarkan keterangan sebelumnya, maka pada penelitian ini akan dilakukan perbandingan metode SVM dengan fungsi kernel dan Neural Network dengan algoritma Backpropagation dalam pengklasifikasian PUS berdasarkan penggunaan kontrasepsi (mengikuti program KB atau tidak) di Provinsi Jawa Barat. Metode yang memiliki nilai ketepatan (akurasi) yang tinggi merupakan metode terbaik dalam pengklasifikasian PUS berdasarkan penggunaan kontrasepsi di Provinsi Jawa Barat.

## METODOLOGI

### Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari hasil Survei Demografi dan Kesehatan Indonesia (SDKI) Tahun 2017. Objek pengamatannya adalah PUS yang ada di Provinsi Jawa Barat sebanyak 5.090 PUS. Pasangan usia subur (PUS) merupakan sasaran program KB. PUS adalah wanita dan pria kawin yang menyatakan sedang dalam status kawin atau hidup bersama dengan pasangannya pada saat survei dilakukan, dimana wanita berumur 15-49 tahun dan pria kawin berumur 15-54 tahun (BKKBN, Kemenkes, BPS dan USAID, 2018).

*Label variable* yang digunakan dalam penelitian ini adalah penggunaan kontrasepsi pada PUS, dibagi menjadi dua kategori yaitu mengikuti program KB dan tidak mengikuti program KB. Menurut Peraturan Pemerintah Nomor 87 Tahun 2014 tentang Perkembangan Kependudukan dan Pembangunan Keluarga, Keluarga Berencana, dan Sistem Informasi Keluarga, Keluarga Berencana

(KB) adalah upaya dalam mengatur jumlah kelahiran anak, jarak dan usia ideal melahirkan, mengatur kehamilan, melalui promosi, perlindungan, dan bantuan sesuai dengan hak reproduksi untuk mewujudkan keluarga yang berkualitas. Dalam upaya mengatur kelahiran anak, jarak dan usia ideal melahirkan dan mengatur kehamilan inilah dilakukan dengan menggunakan alat/obat/cara kontrasepsi. Jenis Alat/obat/cara kontrasepsi terdiri dari kontrasepsi modern (sterilisasi pria dan sterilisasi wanita, suntik KB, IUD, Pil, implan, kondom, metode amenore laktasi (MAL), diafragma dan kontrasepsi darurat) dan kontrasepsi tradisional (pantang berkala, sanggama terputus, dan tradisional lainnya).

*Attribute variable* yang digunakan mengacu penelitian Fienalia (2012) yang terdiri dari: (1) Usia istri ( $X_1$ ), dibagi menjadi tujuh kategori yaitu 15-19 tahun, 20-24 tahun, 25-29 tahun, 30-34 tahun, 35-39 tahun, 40-44 tahun, dan 45-49 tahun; (2) Paritas atau jumlah anak ( $X_2$ ); (3) Pendidikan terakhir istri ( $X_3$ ) dan suami ( $X_4$ ), dibagi menjadi enam kategori yaitu tidak sekolah, tidak tamat SD, tamat SD, tidak tamat SLTA, tamat SLTA, dan perguruan tinggi; (4) Pengetahuan KB baik cara modern atau tradisional ( $X_5$ ), dibagi menjadi dua kategori yaitu tidak tahu dan tahu; (5) Keterpaparan informasi KB melalui media massa (radio, televisi, majalah/ koran atau *handphone/mobile phone*) dalam beberapa bulan terakhir ( $X_6$ ), dibagi menjadi dua kategori yaitu tidak terpapar dan terpapar; (6) Keterpaparan pesan KB melalui Petugas KB dalam 12 bulan terakhir ( $X_7$ ), dibagi menjadi dua kategori yaitu tidak terpapar dan terpapar; (7) Keterpaparan pesan KB melalui Faskes KB dalam 12 bulan terakhir ( $X_8$ ), dibagi menjadi dua kategori yaitu tidak terpapar dan terpapar; (8) Pekerjaan istri ( $X_9$ ) dan suami ( $X_{10}$ ), dibagi menjadi sepuluh kategori yaitu tidak bekerja, profesional/teknisi, manajer dan administrasi, tata usaha, tenaga usaha penjualan, tenaga usaha jasa, pertanian, pekerja industri, lainnya, dan tidak tahu; (9) Kuantil kekayaan menggambarkan tingkat kesejahteraan rumah tangga, dibagi menjadi lima kategori yaitu terbawah, menengah bawah, menengah, menengah atas, dan teratas; (10) Keputusan ikut KB, dibagi menjadi empat kategori yaitu sendiri, suami, bersama, dan lainnya; (11) Tempat tinggal, dibagi menjadi dua kategori yaitu kota dan desa.

### Data Preprocessing

*Data preprocessing* dilakukan untuk menjamin kualitas data yang akan diolah dalam pemodelan klasifikasi. Ada beberapa teknik *data preprocessing*, yaitu *data cleaning*, *data integration*, *data reduction*, dan *data transformation* (Han, Kamber dan Pei, 2012). Pada penelitian ini, tahap *preprocessing* dilakukan dengan *data cleaning* untuk menghilangkan objek pengamatan yang mengandung *missing value* agar tidak mengganggu analisis data yang akan dilakukan.

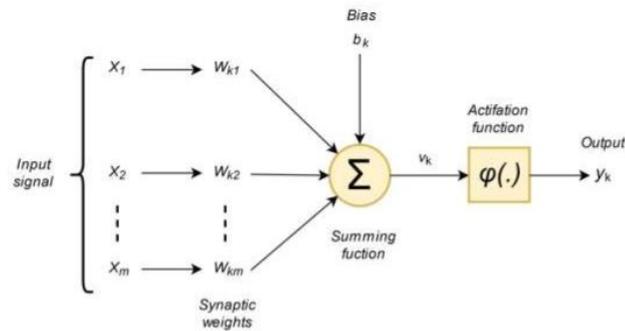
### Pemodelan Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang dapat menggambarkan dan membedakan kelas data atau konsep, dengan tujuan agar model tersebut dapat digunakan untuk memprediksi kelas yang belum diketahui dari suatu objek pengamatan (Han dan Kamber, 2000).

### *Backpropagation Neural Network* (BPNN)

Metode *neural network* tercipta sebagai suatu generalisasi model matematis dari pemahaman manusia (*human cognition*) yang didasarkan atas asumsi sebagai berikut:

- Pemrosesan informasi terjadi pada elemen sederhana yang disebut neuron.
- Isyarat mengalir di antara sel saraf atau neuron melalui suatu sambungan penghubung.
- Sambungan penghubung memiliki bobot yang bersesuaian.
- Sel saraf merupakan fungsi aktivasi terhadap isyarat hasil penjumlahan berbobot yang masuk kepadanya untuk menentukan isyarat keluarannya (Kamil, 2017).



GAMBAR 1. Model Neural Network

Menurut Shukla, Tiwari dan Kala (2010) dalam Rifa'i (2013), setiap unit neuron mengambil *input* dari satu set neuron, kemudian memproses *input* tersebut dan melewati *output* untuk satu set neuron. Keluaran dikumpulkan oleh neuron lain untuk diproses lebih lanjut.

Metode *neural network* masuk ke dalam kelas *supervised training*. Ada beberapa jenis desain *neural network*, namun dalam pada penelitian ini akan digunakan *Backpropagation Neural Network* (BPNN).

BPNN pertama kali diperkenalkan oleh Rumelhart, Hinton dan William pada tahun 1986, kemudian Rumelhart dan Mc Clelland mengembangkan pada tahun 1988. BPNN ini tersusun atas sekumpulan elemen neuron atau simpul atau sel yang terinterkoneksi dan terorganisasi dalam lapisan-lapisan. Setiap neuron memproses sinyal dengan fungsi aktivasinya yaitu fungsi logistik :

$$f(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

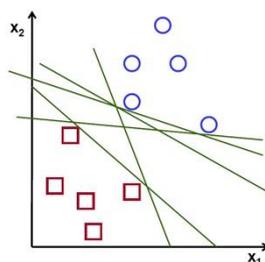
dan turunannya

$$f'(x) = f(x)(1 - f(x)) \tag{2}$$

Algoritma *backpropagation* proses *training* terdiri dari dua proses, yaitu *feedforward* dan *backpropagation* dari *error*-nya.

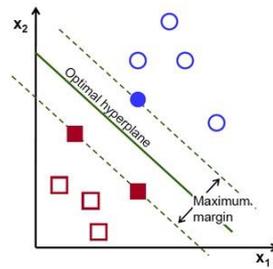
### Support Vector Machine (SVM)

*Support Vector Machine* (SVM) adalah metode *machine learning* yang digunakan untuk klasifikasi baik data linear maupun non-linear. SVM untuk data non-linear merupakan SVM yang dimodifikasi dengan memasukkan fungsi kernel trick, yaitu fungsi yang memetakan fitur data dari dimensi awal yang rendah ke fitur data baru dengan dimensi yang relatif lebih tinggi sehingga kedua kelas dapat dipisahkan secara linear oleh sebuah *hyperplane* (Han, Kamber dan Pei, 2012). SVM berada dalam satu kelas dengan *neural network* dalam hal fungsi dan kondisi permasalahan yang bisa diselesaikan. Kedua metode tersebut masuk ke dalam kelas *supervised learning*. Ide dasar SVM adalah memaksimalkan batas *hyperplane* seperti digambarkan pada gambar di bawah ini:



GAMBAR 2. Kemungkinan Hyperplane dalam model klasifikasi

Pada Gambar 2 menunjukkan ada sejumlah pilihan *hyperplane* yang mungkin untuk set data, sedangkan pada Gambar 3 merupakan *hyperplane* dengan margin paling maksimum.



GAMBAR 3. Optimal *Hyperplane* dengan *Maximum Margin*

*Hyperplane* (batas keputusan) pemisah terbaik antara kedua kelas didapatkan dengan mengukur margin *hyperplane* tersebut dan mencari titik maksimalnya. Margin adalah jarak antara *hyperplane* tersebut dengan data terdekat dari masing masing kelas. Data yang paling dekat itu disebut sebagai *support vector*.

Pada mulanya teknik *machine learning* dikembangkan dengan asumsi kelinieran, sehingga algoritma yang dihasilkan terbatas untuk kasus-kasus yang linear saja. Sedangkan, untuk domain data secara *real* sangat jarang ditemukan data yang bersifat linier. Untuk menyelesaikan problem data yang tidak linier, SVM dapat dimodifikasi dengan memasukkan fungsi kernel. Fungsi kernel memberikan berbagai kemudahan, karena dalam proses pembelajaran SVM, untuk menentukan *support vector*, maka cukup dengan mengetahui fungsi kernel yang dipakai, dan tidak perlu mengetahui wujud dari fungsi non-linier. Berikut fungsi kernel yang digunakan pada penelitian ini adalah:

Kernel Linier:

$$K(x, y) = (x^T y) \quad (3)$$

Kernel Polynomial:

$$K(x, y) = (\gamma x^T y + coef. 0)^d \quad (4)$$

Kernel Radial:

$$K(x, y) = \exp \left\{ -\frac{|x - y|^2}{2\gamma^2} \right\} \quad (5)$$

Kernel Anova:

$$K(x, y) = \sum_{k=1}^n \exp (-\sigma(x^k - y^k)^2) \quad (6)$$

dengan  $d$  adalah kernel *degree* (derajat) dan  $k$  adalah kernel *gamma*.

### Tahapan Pemodelan Klasifikasi

Tahapan pemodelan klasifikasi menggunakan BPNN dan SVM diawali dengan membagi data menjadi dua bagian yaitu data *training* yang digunakan untuk pemilihan model dan *testing* untuk melihat kebaikan model. Pembagian data menggunakan metode *cross validation* dengan  $k\text{-fold} = 10$  dan jenis sampling adalah *stratified sampling*. Kohavi (1995) pada Raschka (2018) menyatakan *10-fold cross validation* dapat memberikan *trade off* antara bias dan varians terbaik sehingga cenderung memberikan estimasi kinerja dengan bias yang menurun dalam pemilihan model terbaik. Pada *10-fold cross validation*, data *training set* dipartisi secara acak menjadi 10 subset berukuran sama. Pada setiap bagian akan digunakan untuk *training* dan *testing* secara bergantian. Pada subset 1 digunakan sebagai data *testing* dan sisanya (9 subset) digunakan sebagai data *training* yang dilakukan secara berulang sebanyak 10 kali sampai semua bagian digunakan *testing*.

Selanjutnya melakukan klasifikasi data PUS berdasarkan determinan penggunaan kontrasepsi di Provinsi Jawa Barat dengan metode BPNN dan SVM.

*Tahapan Pengklasifikasian Data Menggunakan BPNN*

1. Menentukan nilai parameter *learning rate* dan *momentum* yang optimal menggunakan metode *grid search*. Menurut tutorial *RapidMiner* di [www.neuralmarkettrends.com](http://www.neuralmarkettrends.com) (2011), pada pengujian optimasi nilai parameter *learning rate* dan *momentum* dapat dibandingkan mulai dari 0.1 hingga 0.9. Sedangkan, nilai parameter lainnya sesuai *default* yang ada pada *RapidMiner Studio Trial 9.2.001*.
2. Menentukan banyaknya neuron pada lapisan *input*. Banyak neuron yang digunakan pada lapisan *input* adalah 13 buah neuron karena jumlah *input* yang digunakan sebanyak 13 *attribute variable*.
3. Menganalisis banyaknya neuron pada lapisan tersembunyi. Pada penelitian ini jumlah neuron pada lapisan tersembunyi akan dianalisis sesuai dengan hasil analisis *RapidMiner Studio Trial 9.2.001*.
4. Menentukan jumlah neuron pada lapisan *output*. *Output* yang dihasilkan adalah 2 kategori yaitu tidak dan menggunakan kontrasepsi, sehingga pada lapisan *output* terdapat dua buah neuron.
5. Melakukan analisis BPNN pada data *training*, dengan menggunakan nilai parameter *learning rate* dan *momentum* yang optimal.
6. Melakukan pengujian model BPNN menggunakan data *testing*. Proses pengujian hanya menggunakan tahap *feedforward* dan menggunakan bobot-bobot yang diperoleh dari proses *training*.

*Tahapan Pengklasifikasian Data Menggunakan SVM*

1. Menentukan nilai parameter *gamma* dan *cost* yang optimal menggunakan metode *grid search*. Menurut Puspitasari dkk (2018), pada pengujian optimasi nilai parameter *gamma* dapat dibandingkan mulai dari 0.001 hingga 50, serta nilai parameter *cost* mulai dari 0.01 hingga 100. Sedangkan, nilai parameter lainnya sesuai *default* yang ada pada *RapidMiner Studio Trial 9.2.001*.
2. Menentukan fungsi kernel (Linier, Radial, Polynomial dan Anova), nilai parameter kernel dan *cost* untuk optimasi *hyperplane* pada data *training*.
3. Menentukan *hyperplane* dengan menggunakan nilai parameter *gamma* dan *cost* optimal.
4. Menggunakan *hyperplane* dengan nilai parameter *gamma* dan *cost* optimal yang diperoleh untuk setiap fungsi kernel pada klasifikasi data *testing*.

**Evaluasi Kinerja Klasifikasi**

*Confusion matrix* adalah alat yang berguna untuk menganalisis seberapa baik atau seberapa akurat metode klasifikasi dapat mengenali objek pengamatan dari kelas yang berbeda (Han, Kamber dan Pei, 2012). Tabel 1 menunjukkan *confusion matrix* untuk klasifikasi biner. Kelas positif (1) disebut juga kelas minoritas dan kelas negatif (0) disebut juga kelas mayoritas. Pada matriks diperoleh hasil klasifikasi antara lain TP adalah jumlah objek pengamatan pada kelas positif yang diklasifikasikan benar; FN yaitu jumlah objek pengamatan pada kelas positif yang diklasifikasikan sebagai kelas negatif; FP adalah jumlah objek pengamatan pada kelas negatif yang diklasifikasikan sebagai kelas positif; serta TN adalah jumlah objek pengamatan pada kelas negatif yang diklasifikasikan dengan benar (Maalouf dan Trafalis, 2011).

**TABEL 1.** *Confusion Matrix* dari dua kelas

<i>Confusion Matrix</i>		<b>Kelas Prediksi</b>	
		<b>1</b>	<b>0</b>
<b>Kelas Aktual</b>	<b>1</b>	TP ( <i>True Positive</i> ) <i>Correct result</i>	FP ( <i>False Positive</i> ) <i>Unexpected result</i>
	<b>0</b>	FN ( <i>False Negative</i> )	TN ( <i>True Negative</i> )

<i>Missing result</i>	<i>Correct absence of result</i>
-----------------------	----------------------------------

Sumber: Maalouf dan Trafalis (2011)

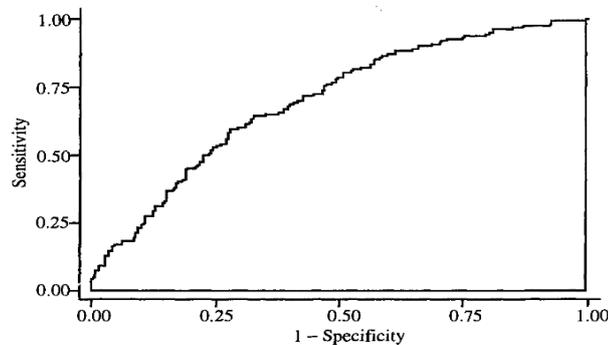
Beberapa ukuran evaluasi kinerja klasifikasi dapat dihitung berdasarkan *confusion matrix*, antara lain akurasi, *precision*, dan *recall*, dengan persamaan sebagai berikut:

**TABEL 2.** Ukuran Evaluasi Kinerja Klasifikasi

Ukuran	Rumus
Akurasi	$\frac{TP + TN}{TP + FP + FN + TN}$
Recall	$\frac{TP}{TP + FN}$
Precision	$\frac{TN}{TN + FP}$

Sumber : Maalouf dan Trafalis (2011)

Ukuran evaluasi kinerja klasifikasi lain adalah kurva *Receiver Operating Characteristic* (ROC). Kurva ROC adalah kurva analisis yang menggambarkan kinerja suatu model klasifikasi pada dua dimensi antara *sensitivity* sebagai sumbu y dan  $(1 - \text{specificity})$  sebagai sumbu x (Fawcett, 2006). Kurva ROC tidak tergantung pada distribusi objek pengamatan antar kelas (Kubat dan Matwin, 1997).



Sumber: Hosmer dan Lemeshow (2000)

**GAMBAR 4.** Kurva ROC

Nilai tunggal yang dapat digunakan untuk mengukur kinerja klasifikasi pada kurva ROC adalah *Area Under Curve the ROC* (AUC). Tabel 3 menunjukkan standar kategori pengklasifikasian berdasarkan nilai AUC:

**TABEL 3.** Kategori klasifikasi berdasarkan nilai AUC

Nilai AUC	Kategori Pengklasifikasian
0.90 – 1.00	<i>Excellent classification</i>
0.80 - 0.90	<i>Good classification</i>
0.70 – 0.80	<i>Fair classification</i>
0.60 – 0.70	<i>Poor classification</i>
0.50 – 0.60	<i>Failure</i>

Sumber: Gorunescu (2011)

Pengolahan data untuk pemodelan klasifikasi BPNN dan SVM menggunakan *RapidMiner Studio Trial 9.2.001*.

## HASIL DAN PEMBAHASAN

### Pemodelan Klasifikasi BPNN

Sebelum dilakukan analisis model klasifikasi menggunakan metode BPNN, terlebih dahulu dilakukan penentuan parameter optimal menggunakan metode *grid search* yang dipadukan dengan metode *k-fold cross validation* ( $k\text{-fold} = 10$ ) sebagai pengukur kinerja metode klasifikasi BPNN pada

data *training*. Selanjutnya, berdasarkan nilai parameter optimal yang telah diperoleh maka dilakukan analisis model klasifikasi menggunakan BPNN dipadukan dengan metode *k-fold cross validation* ( $k\text{-fold}=10$ ).

Berdasarkan *output RapidMiner Studio Trial 9.2.001* diperoleh nilai parameter *learning rate* sebesar 0.66 dan *momentum* sebesar 0.18 memiliki tingkat akurasi yang paling tinggi dari nilai lainnya yaitu 99.84%, sehingga nilai parameter tersebut merupakan nilai parameter optimal dengan *classification error* sebesar 0.16%. Oleh karena itu, model klasifikasi terbaik menggunakan BPNN adalah model klasifikasi dengan nilai parameter *learning rate* = 0.66, nilai parameter *momentum* = 0.18, serta arsitektur jaringan terdiri dari 1 lapisan *input* dengan 13 *neuron*, 1 lapisan *hidden layer* dengan 9 *neuron*, dan 1 lapisan *output* dengan 2 *neuron*.

### Pemodelan Klasifikasi SVM

Sebelum dilakukan analisis model klasifikasi menggunakan metode SVM, terlebih dahulu dilakukan pengujian optimasi parameter *gamma* dan *cost* menggunakan *grid search* yang dipadukan dengan metode *k-fold cross validation* ( $k\text{-fold}=10$ ) sebagai pengukur kinerja metode klasifikasi SVM pada data *training*. Selanjutnya, berdasarkan nilai parameter optimal yang telah diperoleh maka dilakukan analisis model klasifikasi menggunakan SVM dipadukan dengan metode *k-fold cross validation* ( $k\text{-fold}=10$ ).

Pengujian nilai parameter *cost* pada dasarnya memiliki fungsi untuk meminimalkan nilai *error* dan memperkecil nilai *sloack* variabel, hal itu relatif penting diperlukan untuk memaksimalkan margin dan meminimalkan jumlah *sloack*. Apabila nilai parameter *cost* mendekati angka 0 maka lebar margin yang terdapat pada *hyperplane* menjadi maksimal, sehingga nilai parameter *cost* dapat digunakan untuk memperkecil nilai *error* pada proses *training* saat perhitungan nilai bobot ( $w$ ) dan nilai bias. Oleh karena itu, untuk mendapatkan nilai *error* yang kecil maka nilai parameter *cost* yang digunakan juga harus memiliki nilai yang kecil.

Berdasarkan *output RapidMiner Studio Trial 9.2.001* diperoleh nilai parameter optimal menggunakan metode SVM untuk masing-masing fungsi kernel adalah sebagai berikut:

**TABEL 4.** Nilai parameter optimal pada model klasifikasi SVM

Fungsi Kernel	Gamma	Cost	k-fold	Sampling	Tingkat akurasi
Linier	-	70.003	10	Stratified sampling	77.17%
Radial	10.0008	100			76.25%
Polynomial	-	100			93.61%
Anova	5.0009	0.01			99.84%

Sumber : Hasil Olahan RapidMiner Studio Trial 9.2.001

### Perbandingan Pemodelan Klasifikasi BPNN dan SVM

Perbandingan hasil kinerja klasifikasi pada pemodelan BPNN dan SVM dengan fungsi kernel Linier, Radial, Polynomial dan Anova dapat dilihat sebagai berikut:

**TABEL 5.** Perbandingan Evaluasi Kinerja Klasifikasi Pemodelan BPNN dan SVM

Ukuran Ketepatan Klasifikasi	Metode				
	BPNN	SVM			
		Linier	Radial	Polynomial	Anova
Akurasi	99.84%	70.92%	75.80%	91.98%	99.84%
Precision	99.50%	51.53%	99.21%	82.54%	99.50%
Recall	100%	79.56%	22.41%	99.18%	100%
AUC	0.999	0.813	0.810	0.991	0.999

Sumber : Hasil Olahan RapidMiner Studio Trial 9.2.001

Pada Tabel 5 menunjukkan bahwa model klasifikasi dengan menggunakan metode BPNN dan SVM Fungsi Kernel Anova merupakan model terbaik karena memiliki nilai tingkat akurasi, *precision*,

*recall* dan AUC paling tinggi diantara metode lainnya, yaitu tingkat akurasi sebesar 99.84%, *precision* sebesar 99.50%, *recall* sebesar 100% dan AUC sebesar 0.999. Untuk nilai AUC, berdasarkan Tabel 3 maka dapat disimpulkan bahwa pemodelan klasifikasi dengan menggunakan metode BPNN dan SVM Fungsi Kernel Anova dapat dikategorikan sebagai *excellent classification*. Hal ini mengindikasikan bahwa pemodelan klasifikasi menggunakan BPNN dan SVM Fungsi Kernel Anova dapat sangat baik memprediksi kelas KB dan Tidak KB. Oleh karena itu, dalam penelitian ini, model algoritma BPNN dan SVM Fungsi Kernel Anova adalah pemodelan klasifikasi terpilih. Adapun *confussion matrix* dan kurva ROC pada pemodelan klasifikasi terpilih adalah sebagai berikut:

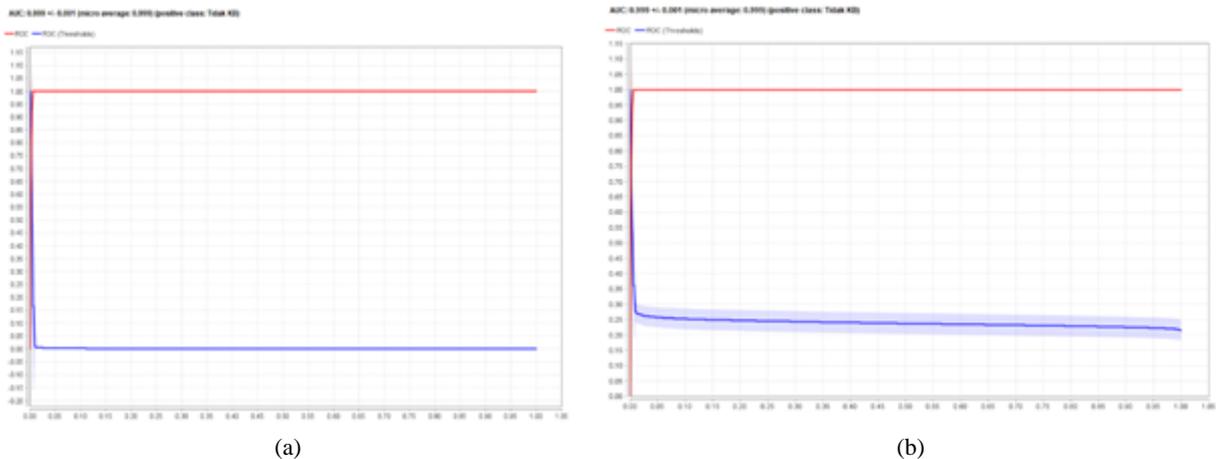
**TABEL 6.** *Confusion Matrix* pada pemodelan klasifikasi BPNN

Prediksi	Aktual		Class Precision
	KB	Tidak KB	
KB	3498	0	100%
Tidak KB	8	1584	99.50%
Class Recall	99.77%	100%	

Sumber : Hasil Olahan RapidMiner Studio Trial 9.2.001

**TABEL 7.** *Confusion Matrix* pada pemodelan klasifikasi SVM Fungsi Kernel Anova

Prediksi	Aktual		Class Precision
	KB	Tidak KB	
KB	3498	0	100%
Tidak KB	8	1584	99.50%
Class Recall	99.77%	100%	



**GAMBAR 5.** Kurva ROC untuk Pemodelan Klasifikasi a) BPNN, b)SVM.

## KESIMPULAN DAN SARAN

### Kesimpulan

Berdasarkan hasil dan pembahasan yang telah diuraikan mengenai analisis kinerja metode BPNN dan SVM dalam mengklasifikasikan PUS berdasarkan determinan penggunaan kontrasepsi (KB dan tidak KB) di Provinsi Jawa Barat beserta perbandingan ketepatan klasifikasinya, maka dapat diambil kesimpulan sebagai berikut:

1. Metode BPNN dapat menggambarkan klasifikasi PUS berdasarkan determinan penggunaan kontrasepsi (KB dan tidak KB) di Provinsi Jawa Barat, dengan parameter *learning rate* sebesar 0.66 dan parameter *momentum* sebesar 0.18. Model klasifikasi tersebut memiliki tingkat akurasi sebesar 99.84%.
2. Metode SVM dapat menggambarkan klasifikasi PUS berdasarkan determinan penggunaan kontrasepsi (KB dan Tidak KB) di Provinsi Jawa Barat. Fungsi kernel yang digunakan adalah Linier, Radial, Polynomial dan Anova. Model klasifikasi terbaik pada penelitian ini adalah Kernel Anova dengan parameter *gamma* sebesar 5.0009 dan parameter *cost* sebesar 0.01. Model klasifikasi tersebut memiliki tingkat akurasi sebesar 99.84%.
3. Kedua metode klasifikasi baik BPNN dan SVM Fungsi Kernel Anova dapat menggambarkan klasifikasi PUS berdasarkan determinan penggunaan kontrasepsi (KB dan Tidak KB) di Provinsi

Jawa Barat dengan tingkat akurasi yang sama, yaitu sebesar 99.84%, artinya model klasifikasi sangat baik dalam memprediksi kelas KB dan Tidak KB.

#### Saran

PUS berdasarkan penggunaan kontrasepsi sebenarnya terdapat empat kategori dimana dua kategori memiliki persentase yang kecil; sehingga data tidak seimbang. Oleh karena itu, penelitian selanjutnya diperlukan kajian tentang data tidak seimbang sehingga dapat mengoptimalkan kinerja klasifikasi.

#### UCAPAN TERIMA KASIH

Terima kasih kepada Bapak Yudhie Andriyana, M.Sc,Ph.D. atas masukan dan bimbingannya untuk penelitian ini.

#### REFERENSI

- BKKBN, Kemenkes, BPS dan USAID. 2018. Survei Demografi dan Kesehatan Indonesia 2017. Jakarta : Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN).
- BPS. 2018. Statistik Indonesia : *Statistical Yearbook of Indonesia*. Jakarta : Badan Pusat Statistik (BPS).
- Fawcett, T.. 2006. An Introduction to ROC Analysis. *Journal of Pattern Recognition Letters*, Vol. 27, Hal. 861-874. Florida: Chapman and Hall/CRC.
- Fienalia. R.A. 2012. *Faktor-Faktor yang Berhubungan dengan Penggunaan Metode Kontrasepsi Jangka Panjang (MKJP) di Wilayah Kerja Puskesmas Pancoran Mas Kota Depok Tahun 2011*. Skripsi. Fakultas Kesehatan Masyarakat. Prodi Studi Kesehatan Reproduksi. Jakarta : Universitas Indonesia.
- Gorunescu, F. 2011. *Data Mining Concept, Models and Techniques*. Verlag Berlin Heidelberg: Springer.
- Hakim. A.R. 2016. *Klasifikasi Keikutsertaan Keluarga dalam Program Keluarga Berencana (KB) di Kota Semarang menggunakan Metode Mars dan FK-NNC*. Skripsi. Semarang : Jurusan Statistika, Universitas diponegoro.
- Han, J. dan M. Kamber. 2000. *Data Mining: Concepts and Techhiques*. San Francisco: Morgan Kaufmann Publishers.
- Han, J., M. Kamber dan Pei J.. 2012. *Data Mining Concepts and Techhiques Third Edition*. Waltham: Elsevier Inc.
- Hosmer, D.W. dan S. Lemeshow. 2000. *Applied Logistic Regression Second Edition*. New Jersey: John Wiley & Sons, Inc.
- Johra. M.B. 2018. *Perbandingan Kernel Trick pada Non-Linier Support Vector Machine (Studi Kasus: Pemilihan Penolong Persalinan di Provinsi Maluku Utara 2016)*. Tesis. Bandung : Jurusan Statistika Terapan, Universitas Padjadjaran.
- Johra, M.B. 2018. Penerapan *Non-Linier Support Vector Machine* pada Penggunaan Alat Kontrasepsi di Provinsi Maluku Utara. *Jurnal Matematika "Mantik", Oktober 2018. Vol 04 No. 02*.
- Kamil. R. 2017. *Implementasi Jaringan Syaraf Tiruan Backpropagation pada Pengenalan Pola Batik Motif Lampung*. Skripsi. Lampung : Jurusan Ilmu Komputer, Universitas Lampung.
- Kubat, M. dan S. Matwin. 1997. Addressing The Curse of Imbalanced Training Sets: One-Sided Selection. *In: Proceedings of The 14th International Conference on Machine Learning, Morgan Kaufman-San Francisco*, Hal. 179-186.

- Maalouf, M. dan T.B. Trafalis. 2011. Rare Events and Imbalanced Datasets: An Overview. *Int. Journal Data Mining, Modelling and Management*, Vol. 3 (4), Hal. 375-385.
- Raschka, S.. 2018. *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*. arXiv:1811.12808.
- Suwardika. G.. 2017. Pengelompokan dan Klasifikasi Penggunaan Kontrasepsi di Indonesia. *Jurnal Matematika, Sains, dan Teknologi 15 Agustus 2017*,. Diunduh dari <https://scholar.google.co.id/citations?user=RpRdQmsAAAAJ&hl=en> diakses tanggal 5 April 2019.