



PENGELOMPOKKAN DAN KLASIFIKASI PENGGUNAAN KONTRASEPSI DI INDONESIA

Gede Suwardika
UPBJJ-UT Denpasar
e-mail: isuwardika@ut.ac.id

ABSTRACT

The use of contraception is crucial, given the growth rate can decrease. The election to use or not to use contraceptives is one of the problems of classification. Classification problem can be solved by methods of binary Logistic Regression, Support Vector Machine (SVM), and CART (Classification and Regression Trees). However, due to the variable response will be made binary then the done method of grouping in advance with its own assumptions, the k-means cluster and cluster the kernel k-means. In this study data about the use of classification taken contraceptives that were affected by the 9 free variables. Such data is data the secondary National Contraceptive Prevalence Survey results Indonesia in 1987. The number of respondents is 1.473 people. The end results were obtained grouping the results turned out to be assuming themselves better than other methods. While on the method of classification SVM classification that results obtained are better than the other two methods. However, if the method of grouping and classification are merged, acquired that classification with SVM where the variable response is obtained based on the k-means cluster groupings or clusters of the kernel k-means can make Apperant Error Rate (APER) the least.

Keywords: binary logistic regression, CART, contraception, SVM

ABSTRAK

Penggunaan kontrasepsi merupakan hal yang penting, mengingat dapat menurunkan laju pertumbuhan. Pemilihan menggunakan atau tidak menggunakan alat kontrasepsi merupakan salah satu permasalahan klasifikasi. Permasalahan klasifikasi dapat diselesaikan salah satunya dengan metode Regresi logistik biner, *Support Vector Machine* (SVM), dan CART (*Classification and Regression Trees*). Namun dikarenakan variabel respon akan dijadikan biner maka dilakukan metode pengelompokan terlebih dahulu dengan asumsi sendiri, *cluster k-means* dan *cluster kernel k-means*. Dalam penelitian ini diambil data mengenai klasifikasi penggunaan kontrasepsi yang dipengaruhi oleh 9 variabel bebas. Data tersebut adalah data sekunder hasil Survey Prevalensi Kontrasepsi Nasional Indonesia tahun 1987. Jumlah respondennya adalah 1.473 orang. Hasil akhir yang diperoleh ternyata hasil pengelompokan dengan asumsi sendiri lebih baik dibandingkan metode yang lain. Sedangkan pada metode pengklasifikasian diperoleh bahwa hasil klasifikasi SVM lebih baik dibanding dua metode yang lain. Namun apabila metode pengelompokan dan klasifikasi digabung, diperoleh bahwa pengklasifikasian dengan SVM dimana variabel respon diperoleh berdasarkan hasil pengelompokan *cluster k-means* atau *cluster kernel k-means* dapat menghasilkan Apperant Error Rate (APER) yang paling kecil.

Kata kunci: CART, kontrasepsi, regresi logistik biner, SVM

Indonesia merupakan salah satu negara yang jumlah penduduknya banyak yaitu 254,9 juta pada tahun 2015. Salah satu upaya untuk mengurangi laju pertumbuhan adalah dengan diselenggarakannya program Keluarga Berencana. Salah satu metode dalam KB adalah dengan menggunakan metode atau alat kontrasepsi. Memilih metode atau alat kontrasepsi bukan merupakan hal yang mudah karena efek yang berdampak terhadap tubuh tidak akan diketahui selama belum menggunakannya. Selain itu tidak ada metode atau alat kontrasepsi yang selalu cocok bagi semua orang karena situasi dan kondisi tubuh dari setiap individu selalu berbeda, sehingga perlunya pengetahuan yang jelas mengenai kekurangan dan kelebihan dari masing-masing metode atau alat kontrasepsi yang disesuaikan dengan kondisi tubuh pengguna.

Pemilihan alat kontrasepsi secara efektif harus berdasarkan tingkat keberhasilannya. Pemilihan metode kontrasepsi yang tepat dan rasional merupakan salah satu hal yang perlu diperhatikan dalam pemakaian alat kontrasepsi. Pemilihan metode kontrasepsi dapat dipengaruhi oleh beberapa faktor antara lain karakteristik akseptor (meliputi umur, pendidikan, pendapatan keluarga per bulan, jumlah anak, dan sebagainya) dan pengetahuan akseptor tentang alat kontrasepsi.

Secara statistik pemilihan alat kontrasepsi tersebut merupakan salah satu permasalahan klasifikasi. Klasifikasi merupakan salah satu bentuk peramalan yang memiliki nilai keluaran diskrit, dan bertujuan untuk menemukan suatu fungsi keputusan $f(x)$ yang secara akurat memprediksi kelas dari data (Santosa, 2007). Apabila suatu bentuk data memiliki variabel dependen (respon) yang memiliki lebih dari dua kategori maka untuk dibawa ke dalam bentuk biner maka perlu dilakukan pengelompokan. Pengelompokan tersebut diperoleh dengan cara asumsi sendiri, hasil pengelompokan berdasarkan metode *cluster k-means* dan *cluster kernel k-mean*. Apabila klasifikasi tersebut memiliki variabel dependen dengan dua kategori dan variabel independennya beragam (memiliki kategori atau kontinu), maka digunakan regresi logistik biner (Johnson and Winchern, 2007). Namun, seiring perkembangan zaman ada metode baru yakni *Support Vector Machine* (SVM) dan *CART (Classification and Regression Trees)*. Dikarenakan ada tiga metode yang dapat digunakan, maka penelitian ini dilakukan guna mendapatkan metode yang lebih sesuai untuk pengklasifikasian pemilihan alat kontrasepsi.

TINJAUAN PUSTAKA

Berkaitan dengan penelitian tentang klasifikasi penggunaan alat kontrasepsi, diperlukan landasan teori yang mendukung. Adapun landasan teori tersebut adalah tentang *Support Vector Machine*, Regresi Logistik Biner, *Classification and Regression Trees*, *cluster k-means*, *cluster kernel k-mean*, *Apperent Error Rate*, dan Kontrasepsi.

Support Vector Machine (SVM)

Menurut Gunn (1998), SVM adalah suatu teknik yang dikembangkan oleh Vapnik pada tahun 1995 untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. SVM termasuk dalam kelas *supervised learning*. Konsep dasar SVM adalah menemukan fungsi pemisah (*hyperplane*) yang paling baik (optimum), (Epa dan Santi, 2015). *Hyperplane* terbaik dapat dicari dengan memaksimalkan margin atau jarak dari dua set obyek dari dua kelas yang berbeda. *Hyperplane* pemisah dalam bentuk kanonikal harus memenuhi konstrain seperti persamaan (1) (Gunn, 1998).

$$y_i (w^T x_i + b) \geq 1, i = 1, 2, \dots, n \quad (1)$$

dimana x_i adalah data input, y_i adalah output yang nilainya =1 atau -1, w dan b adalah parameter yang kita cari nilainya. *Hyperplane* yang mampu memisahkan data secara optimal adalah *hyperplane* yang meminimasi persamaan (2).

$$\Phi(w) = \frac{1}{2} \|w\|^2 \quad (2)$$

Gunn (1998) menambahkan bahwa solusi masalah optimasi dari persamaan (2) di bawah konstrain persamaan (1) diberikan melalui *saddle point* dari fungsi Lagrange (Lagrangian) seperti pada persamaan (3).

$$\Phi(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^n \alpha_i \left[y_i (w^T x_i + b) - 1 \right] \quad (3)$$

Dimana α merupakan Lagrange multiplier. Dikarenakan pada saat meminimumkan fungsi tujuan pada persamaan (2) terdapat batasan persamaan (1), maka digunakan fungsi Lagrangian yang kemudian diminimalkan terhadap variabel w dan b dan dimaksimalkan terhadap variabel $\alpha \geq 0$. Secara matematis fungsi tersebut dijelaskan pada persamaan (4).

$$\text{minimize } L_p(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i (y_i (w^T x_i + b) - 1) \text{ dengan batasan } \alpha_i \geq 0 \quad (4)$$

Dengan demikian akan diperoleh kondisi optimal sebagai berikut:

- kondisi 1 $\rightarrow \frac{\partial L_p}{\partial w} = 0 \rightarrow w = \sum_{i=1}^n \alpha_i y_i x_i$ atau $w^* = \sum_{i=1}^n \alpha_i y_i x_i$
- kondisi 2 $\rightarrow \frac{\partial L_p}{\partial b} = 0 \rightarrow \sum_{i=1}^n \alpha_i y_i = 0$

Selain dua kondisi tersebut apabila fungsi Lagrangian dimaksimalkan terhadap variabel α maka akan diperoleh bentuk permasalahan dual. Dual problem diberikan oleh persamaan (5).

$$\text{maximize } \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (5)$$

dengan batasan $\sum_{i=1}^n \alpha_i y_i = 0$ dan $\alpha_i \geq 0$. Berdasarkan persamaan (5) diperoleh nilai b^* seperti pada persamaan (6).

$$b^* = -\frac{1}{2} \langle w^*, x_r + x_s \rangle \quad (6)$$

Dimana x_r dan x_s adalah sembarang *support vector* dari tiap kelas yang memenuhi

$W^* = \alpha_r, \alpha_s > 0, y_r = -1, y_s = 1$. Fungsi klasifier yang dapat digunakan untuk memisahkan data adalah persamaan (7).

$$f(x) = \text{sgn}(\langle w^*, x \rangle + b) \quad (7)$$

Dalam kasus dimana titik-titik dari dua kelas tidak bisa dipisahkan maka perlu ditambahkan variabel slack ξ_i , sehingga persamaan (1) dimodifikasi menjadi persamaan (8).

$$y_i [(w^T x_i) + b] \geq 1 - \xi_i, \quad i = 1, 2, \dots, n \quad (8)$$

dimana $\xi \geq 0$. *Hyperplane* pemisah yang optimal ditentukan oleh vektor w , yang meminimasi persamaan (9)

$$\Phi(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (9)$$

dimana C adalah nilai biaya yang ditentukan sebagai penalti kesalahan dan sedapat mungkin diminimasi, dengan fungsi pembatas persamaan (8). Nilai C harus ditentukan dari awal perhitungan.

Ketika batasan linier tidak dapat memisahkan kelas data, permasalahan SVM dapat dimodifikasi dengan memasukkan fungsi kernel. Beberapa fungsi kernel dasar yang biasadipakai dalam literatur SVM (Gunn, 1998) adalah:

a. Linear $\rightarrow K(x_i, x_j) = x_i^T x_j$

b. Polinomial $\rightarrow K(x_i, x_j) = [(x_i x_j^T) + 1]^d$

c. Radial Basis Function (RBF) $\rightarrow K(x_i, x_j) = \exp\left[-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right]$

Regresi Logistik Biner

Regresi logistik biner merupakan salah satu pendekatan model matematis yang digunakan untuk menganalisis hubungan beberapa faktor dengan sebuah variabel yang bersifat dikotomus (biner). Pada regresi logistik jika variabel responnya terdiri dari dua kategori misalnya $Y = 1$ menyatakan hasil yang diperoleh “sukses” dan $Y = 0$ menyatakan hasil yang diperoleh “gagal” maka regresi logistik tersebut menggunakan regresi logistik biner. Menurut Agresti (2007) variabel y yang demikian lebih tepat dikatakan sebagai variabel indikator dan memenuhi distribusi Bernoulli. Fungsi distribusi peluang untuk y dengan parameter π_i adalah persamaan (10).

$$f(y_i, \pi_i) = \begin{cases} \pi_i (1 - \pi_i)^{1-y_i} & , y_i = 0, 1 \\ 0 & , \text{yang lain} \end{cases} \quad (10)$$

dimana $\pi_i = P(Y_i = 1)$. Adapun model regresi logistik biner dapat dituliskan seperti persamaan (11).

$$P(Y = 1) = p_i = \frac{e^{\beta_0 + \beta_i x_i}}{1 + e^{\beta_0 + \beta_i x_i}} \quad (11)$$

Nilai parameter β_i ditaksir dengan $\hat{\beta}_i$ menggunakan metode *Maximum Likelihood* (MLE). Selain Menggunakan MLE juga dapat digunakan iterasi *Newton Rhapson*. Setelah nilai $\hat{\beta}_i$ diperoleh diuji secara parsial apakah nilai tersebut sudah signifikan. Adapun hipotesa pengujiannya sebagai berikut:

$$H_0 : \hat{\beta}_i = 0$$

$$H_1 : \hat{\beta}_i \neq 0$$

Statistik uji yang digunakan adalah uji Wald, secara matematis ditulis pada persamaan (12).

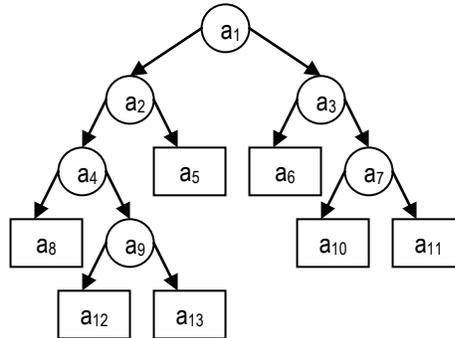
$$W = \frac{\hat{\beta}_i - E(\hat{\beta}_i)}{\sqrt{var(\hat{\beta}_i)}} \quad (12)$$

Pengambilan keputusan dilihat dengan menggunakan tingkat kesalahan sebesar α , dimana H_0 akan ditolak apabila nilai Wald lebih besar daripada $Z_{\alpha/2}$.

CART (*Classification and Regression Trees*)

CART (*Classification and Regression Trees*) adalah salah satu metode nonparametrik dari salah satu teknik eksplorasi data yaitu suatu teknik pohon keputusan (*decisions tree*). Pohon keputusan dibentuk dengan menggunakan algoritma penyekatan rekursif secara biner (Lewis, 2000).

Metode CART dikembangkan oleh Leo Breiman, dkk pada tahun 1980-an dimana CART merupakan cara pemilihan sekelompok data yang terkumpul dalam suatu ruang yang disebut simpul (*node*) menjadi dua simpul anak dan setiap simpul anak dapat dipilah lagi menjadi dua simpul anak berikutnya, begitu seterusnya dan berhenti jika telah mendapatkan sekelompok observasi yang relatif homogen. CART digunakan untuk menggambarkan hubungan antara variabel respon dengan satu atau lebih variabel prediktor. Tujuan utama dari CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai salah satu ciri dari suatu pengklasifikasian.



Gambar 1. Contoh struktur pohon klasifikasi

Gambar 1 menjelaskan bahwa simpul utama (*root node*) dinotasikan sebagai a_1 , sedangkan simpul a_2, a_3, a_4, a_7 , dan a_9 adalah simpul dalam (*internal nodes*). Simpul terminal (*terminal node*) atau yang biasa disebut simpul akhir dinotasikan $a_5, a_6, a_8, a_{10}, a_{11}, a_{12}$, dan a_{13} . Kedalaman pohon dihitung dari simpul utama dimana a_1 berada pada kedalaman 1, a_2 dan a_3 berada pada kedalaman 2 dan seterusnya hingga simpul terminal yang berada pada kedalaman 5. Proses penentuan pohon klasifikasi dimulai dari simpul utama yang berisi data yang akan dipilah. Pemilahan tersebut dilakukan guna memilah data menjadi dua kelompok, yaitu kelompok yang masuk simpul kiri dan kelompok yang masuk simpul kanan. Pemilahan tersebut dilakukan pada tiap simpul hingga didapatkan suatu simpul terminal. Variabel yang nantinya akan memilah pada simpul utama adalah variabel yang paling penting dalam pendugaan kelompok dari amatan.

Cluster K-means

Pengelompokan menggunakan metode *cluster K-means* didasarkan pada nilai fungsi keanggotaannya, dimana fungsi keanggotaan tersebut didasarkan oleh jarak minimum antara objek dengan pusat *cluster*. Fungsi keanggotaan setiap observasi diperoleh melalui iterasi maksimal sehingga tidak ada anggota yang masuk maupun keluar lagi. Tujuan utama algoritma *k-means* adalah untuk meminimasi fungsi objektif yang merupakan fungsi eror kuadrat. Jarak yang digunakan untuk pengelompokan adalah jarak Euclidean yang dirumuskan sebagai berikut :

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{13}$$

Adapun algoritma dari pengelompokkan dengan K-means adalah sebagai berikut.

1. Memilih jumlah klaster.
2. Menentukan inisialisasi k pusat klaster (diberi nilai-nilai random)

3. Menempatkan setiap data/obyek ke kluster terdekat. Kedekatan dua obyek ditentukan berdasarkan jarak kedua obyek tersebut. Jarak paling dekat antara satu data dengan satu kluster tertentu akan menentukan suatu data masuk dalam kluster mana.
4. Menghitung kembali pusat kluster dengan anggota kluster yang sekarang. Pusat kluster adalah rata-rata semua data/obyek dalam kluster
5. Menugaskan lagi setiap obyek memakai pusat kluster yang baru. Jika pusat kluster sudah tidak berubah lagi, maka proses pengklasteran selesai.
6. Kembali ke langkah 3 sampai pusat kluster tidak berubah lagi.

Cluster Kernel K-means

Metode *cluster Kernel K-means* merupakan metode pengembangan dari metode *cluster K-means*. Jarak pengelompokan yang digunakan juga jarak Euclidean seperti pada persamaan (13). Metode ini lebih baik daripada *cluster K-means* ketikapada data terdapat outlier (sifat nonlinear). Dengan menggunakan metode *cluster Kernel K-means*, pernyataan jarak tersebut direalisasikan dalam fungsi Kernel. Fungsi kernel yang digunakan pada penelitian ini adalah kernel RBF. Anggaphlah $u_i = \phi(x_i)$ menunjukkan transformasi x_i , sehingga persamaan (13) menjadi persamaan (14).

$$d(x_i, x_j) = \|\phi(x_i) - \phi(x_j)\|^2 \quad (14)$$

Apperent Error Rate (APER)

APER (*apperent error rate*) merupakan fraksi observasi dalam sampel yang salah diklasifikasikan atau *misclassified* pada fungsi klasifikasi (Johnson and Winchern, 2007). Perhitungan APER terlebih dahulu dibuat matriks konfusinya yang diperlihatkan dalam Tabel 1.

Tabel 1. Matrik Konfusi

Data Aktual	Data Prediksi		Total
	Y ₁	Y ₂	
Y ₁	n _{1C}	n _{1M}	n ₁
Y ₂	n _{2M}	n _{2C}	n ₂

Berdasarkan Tabel 1 nilai APER diperoleh berdasarkan persamaan (15).

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2} \quad (15)$$

Kontrasepsi

Kontrasepsi berasal dari kata kontra berarti mencegah atau melawan, sedangkan konsepsi adalah pertemuan antara sel telur (sel wanita) yang matang dan sel sperma (sel pria) yang mengakibatkan kehamilan. Maksud dari kontrasepsi adalah menghindar dan mencegah terjadinya kehamilan sebagai akibat pertemuan antara sel telur yang matang dengan sel sperma tersebut. Cara kerja kontrasepsi bermacam-macam tetapi pada umumnya mempunyai fungsi mengusahakan agar tidak terjadi ovulasi, melumpuhkan sperma, menghalangi pertemuan sel telur dengan sperma. Kontrasepsi yang ideal harus dapat bekerja dalam waktu yang tahan lama, mempunyai efektifitas

yang tinggi, aman, mudah dalam menggunakan dan melepaskannya dan memiliki beberapa atau tidak sama sekali efek samping (Nancy, 1999).

Berbagai jenis metode atau alat kontrasepsi dibagi menjadi (Sobirin, 2006)

- a. Kontrasepsi sterilisasi, seperti tubektomi dan vasektomi
- b. Kontrasepsi teknik, dibagi menjadi *Coitus Interruptus* (senggama terputus), Sistem kalender (pantang berkala), *Prolonged lactation* (menyusui)
- c. Kontrasepsi mekanik, terdiri dari kondom, *Spermatisida*, *Vaginal diafragma*, IUD (*Intra Uterine Device*) atau spiral
- d. Kontrasepsi hormonal seperti pil KB, suntikan, susuk (Implan), koyo KB atau spiral berhormon.

METODOLOGI

Data yang digunakan adalah data sekunder yang diambil di *website UCI Machine Learning Respository*. Data yang disumbangkan oleh Lim (1997) merupakan data hasil Survey Prevalensi Kontrasepsi Nasional Indonesia (*the National Indonesia Contraceptive Prevalence Survey*) pada tahun 1987. Jumlah respondennya adalah 1.473 orang. Sedangkan variabel independen (X) yang digunakan ada 9 buah dengan 1 variabel dependen Y. Secara rinci ditampilkan oleh Tabel 2. Adapun metode yang digunakan dalam penentuan pengelompokan adalah asumsi sendiri (tercantum pada Tabel 2), *cluster K-means* dan *cluster kernel K-means*. Untuk metode klasifikasi digunakan Regresi Logistic Biner, SVM dan CART. Regresi Logistik Biner menggunakan *software* SPSS sedangkan SVM dengan Matlab dan CART dengan *software* CART 5. Hasil dari masing-masing metode kemudian dibandingkan berdasarkan nilai APER yang paling kecil.

Tabel 2. Deskripsi Variabel

Variabel	Keterangan	Skala data	Kategori
X ₁	Umur Istri	Interval	-
X ₂	Pendidikan Istri	Ordinal	1 = Rendah 2 = Sedang 3 = Agak Tinggi 4 = Tinggi
X ₃	Pendidikan Suami	Ordinal	1 = Rendah 2 = Sedang 3 = Agak Tinggi 4 = Tinggi
X ₄	Jumlah Anak	Rasio	-
X ₅	Agama Istri	Nominal	0 = Bukan Islam 1 = Islam
X ₆	Status Pekerjaan istri	Nominal	0 = Bekerja 1 = Tidak bekerja
X ₇	Tingkat Kesuburan Suami	Ordinal	1 = Rendah 2 = Sedang 3 = Agak Tinggi 4 = Tinggi
X ₈	Standar indeks kehidupan	Ordinal	1 = Rendah 2 = Sedang 3 = Agak Tinggi 4 = Tinggi
X ₉	Pengetahuan tentang akseptor KB	Nominal	0 = Baik 1 = Tidak baik
Y _(sendiri)	Penggunaan Kontrasepsi	Nominal	0 = Tidak Menggunakan 1 = Menggunakan

HASIL DAN PEMBAHASAN

Pembahasan awal sebelum menginjak pada tujuan utama penelitian ini, terlebih dahulu dilihat tentang statistika deskriptif dari masing-masing variabel. Deskripsi tersebut terlihat pada Tabel 3.

Tabel 3. Statistika Deskriptif Masing-Masing Variabel

Variabel	Minimum	Maximum	Mean	Variance
Umur Istri	16	49	32,538	67,688
Pendidikan Istri	1	4	2,959	1,030
Pendidikan Suami	1	4	3,430	0,666
Jumlah Anak	0	16	3,261	5,563
Agama Istri	0	1	0,851	0,127
Status Pekerjaan istri	0	1	0,750	0,188
Tingkat Kesuburan Suami	1	4	2,138	0,748
Standar indeks kehidupan	1	4	3,134	0,953
Pengetahuan tentang akseptor KB	0	1	0,074	0,069
Penggunaan Kontrasepsi	0	1	0,573	0,245

Dengan mengacu pada Tabel 3 terlihat bahwa umur istri berkisar diantara 16 sampai 49 tahun, Namun dengan variasi yang tinggi ternyata rata-rata umur istri adalah 33 tahun. Untuk jumlah anak yang dimiliki rata-rata responden memiliki 4 anak, dimana responden ada yang tidak mempunyai anak, dan ada yang memiliki anak hingga 6 anak. Terlihat bahwa variansi mengenai jumlah anak ini cukup tinggi atau beragam. Variabel yang lain merupakan variabel yang memiliki kategori sesuai pada pembahasan sebelumnya, atau dapat dilihat kembali pada Tabel 2. Untuk variabel-variabel kategori tersebut variansinya cukup kecil. Adapun karakteristik responden berdasarkan ke-8 variabel yang lain akan dilihat dari nilai rata-ratanya. Untuk pendidikan istri rata-rata sudah agak tinggi (rata-rata ≈ 3) dan rata-rata pendidikan suami lebih tinggi. Secara agama, rata-rata responden beragama Islam. Status pekerjaan istri rata-rata banyak yang bekerja. Ditinjau dari tingkat kesuburan suami rata-rata sudah agak tinggi yang diimbangi dengan standar indeks kehidupan yang rata-rata juga sudah tinggi. Hal ini juga didukung dengan banyaknya responden yang rata-rata memiliki pengetahuan tentang akseptor KB dengan baik. Apabila dilihat dari penggunaan alat kontrasepsi ternyata jumlah responden yang menggunakan dengan yang tidak menggunakan hampir sama.

Setelah mengetahui karakteristik responden, maka selanjutnya dilakukan pengelompokan berdasarkan asumsi sendiri, *cluster K-means* dan *cluster kernel K-means*. Hasil dari pengelompokan tersebut disajikan pada Tabel 4.

Tabel 4. Hasil Pengelompokan Pengguna Alat Kontrasepsi

Kelompok	Sendiri	K-mean	Kernel K-means
I (Tidak Menggunakan)	629	682	682
II (Menggunakan)	844	791	791

Berdasarkan hasil yang disajikan pada Tabel 4 diperoleh bahwa untuk setiap metode menghasilkan hasil yang sama, yakni responden yang menggunakan alat kontrasepsi lebih banyak

daripada yang tidak menggunakan. Selain itu ternyata baik dengan metode *cluster k-means* maupun *cluster kernel k-means* diperoleh hasil yang sama untuk setiap kelompok.

Selanjutnya menginjak pada tujuan inti penelitian yakni tentang pengklasifikasian. Metode klasifikasi yang akan dibahas pertama adalah dengan Regresi Logistik Biner. Pada metode ini setiap variabel independen yang memiliki kategori mengambil variabel kontrol pada variabel dengan kategori terakhir. Adapun penaksiran parameter dari model regresi logistik biner diperoleh $\hat{\beta}_i$ yang terdapat pada Tabel 4. Selain itu hasil uji signifikansi parameter secara parsial juga ditampilkan pada Tabel 4 yang dilihat dari nilai Wald dengan hipotesa sebagai berikut:

$$H_0 : \hat{\beta}_i = 0$$

$$H_1 : \hat{\beta}_i \neq 0$$

Pengambilan keputusan dilihat dengan menggunakan $\alpha = 5\%$, dimana H_0 akan ditolak apabila nilai Wald lebih besar daripada $Z_{(0,025)} = 1,96$. Hasil regresi logistik biner mengenai variabel-variabel yang signifikan untuk masing-masing hasil pengelompokan pada pembahasan sebelumnya disajikan pada Tabel 5.

Tabel 5. Nilai Signifikansi $\hat{\beta}_i$ pada Masing-masing Hasil Pengelompokan

Variabel	Sendiri	<i>K-means</i>	<i>Kernel K-means</i>
Konstanta (β_0)	Tolak H_0	Gagal Tolak H_0	Gagal Tolak H_0
Umur Istri (β_1)	Tolak H_0	Gagal Tolak H_0	Gagal Tolak H_0
Pendidikan Istri (1) $\rightarrow (\beta_{21})$	Tolak H_0	Gagal Tolak H_0	Gagal Tolak H_0
Pendidikan Istri (2) $\rightarrow (\beta_{22})$	Tolak H_0	Gagal Tolak H_0	Gagal Tolak H_0
Pendidikan Istri (3) $\rightarrow (\beta_{23})$	Tolak H_0	Gagal Tolak H_0	Gagal Tolak H_0
Pendidikan Suami (1) $\rightarrow (\beta_{31})$	Gagal Tolak H_0	Gagal Tolak H_0	Gagal Tolak H_0
Pendidikan Suami (2) $\rightarrow (\beta_{32})$	Gagal Tolak H_0	Gagal Tolak H_0	Gagal Tolak H_0
Pendidikan Suami (3) $\rightarrow (\beta_{33})$	Gagal Tolak H_0	Gagal Tolak H_0	Gagal Tolak H_0
Jumlah Anak (β_4)	Tolak H_0	Gagal Tolak H_0	Gagal Tolak H_0
Agama Istri (0) $\rightarrow (\beta_{51})$	Tolak H_0	Gagal Tolak H_0	Gagal Tolak H_0
Status Pekerjaan istri (0) $\rightarrow (\beta_{61})$	Gagal Tolak H_0	Gagal Tolak H_0	Gagal Tolak H_0
Tingkat Kesuburan Suami (1) $\rightarrow (\beta_{71})$	Gagal Tolak H_0	Gagal Tolak H_0	Gagal Tolak H_0
Tingkat Kesuburan Suami (2) $\rightarrow (\beta_{72})$	Gagal Tolak H_0	Gagal Tolak H_0	Gagal Tolak H_0
Tingkat Kesuburan Suami (3) $\rightarrow (\beta_{73})$	Gagal Tolak H_0	Gagal Tolak H_0	Gagal Tolak H_0
Standar index kehidupan (1) $\rightarrow (\beta_{81})$	Tolak H_0	Gagal Tolak H_0	Gagal Tolak H_0
Standar index kehidupan (2) $\rightarrow (\beta_{82})$	Tolak H_0	Gagal Tolak H_0	Gagal Tolak H_0
Standar index kehidupan (3) $\rightarrow (\beta_{83})$	Tolak H_0	Gagal Tolak H_0	Tolak H_0
Pengetahuan tentang akseptor KB (0) $\rightarrow (\beta_{91})$	Tolak H_0	Gagal Tolak H_0	Gagal Tolak H_0

Mengacu pada Tabel 5 diperoleh kesimpulan bahwa dengan pengelompokan asumsi sendiri diperoleh hanya 2 variabel yang tidak berpengaruh secara signifikan, yakni Pendidikan Suami dan Tingkat Kesuburan Suami. Hasil klasifikasi dimana variabel respon diperoleh dari *cluster k-means* diperoleh tidak ada satu pun variabel yang berpengaruh secara signifikan. Hal yang tidak jauh berbeda juga terjadi pada hasil klasifikasi dimana variabel respon diperoleh dari *cluster kernel k-means*, dimana hanya terdapat satu variabel yang berpengaruh secara signifikan yakni standar indeks kehidupan responden yang agak tinggi. Hasil prediksi klasifikasi dengan metode ini disajikan

pada Tabel 6. Berdasarkan Tabel 6 dapat diketahui bahwa yang lebih banyak mengklasifikasikan dengan tepat adalah dengan asumsi sendiri.

Tabel 6. Hasil Prediksi Klasifikasi Regresi Logistik Biner

Pengklasifikasian	<i>k-mean</i>	<i>Kernel K-means</i>	sendiri
benar	629	874	1.004
salah	844	599	469

Metode yang lainnya yakni dengan menggunakan SVM, dimana digunakan nilai $\sigma = 1$ dan $0,01$ sedangkan $C=2$ dan $0,01$ diperoleh hasil untuk masing-masing hasil pengklasifikasian disajikan pada Tabel 7. Hasil yang diperoleh menunjukkan adanya perbedaan dengan hasil dari regresi logistik biner. Dimana hasil dari pengklasifikasian SVM dengan variabel respon ditentukan berdasarkan *cluster k-means* dan *cluster kernel k-means* lebih baik daripada asumsi sendiri, bahkan mencapai ketepatan 100% untuk masing-masing Nilai Parameter C dan σ . Pada hasil asumsi sendiri apabila nilai σ semakin kecil maka hasil klasifikasi akan semakin baik, sedangkan nilai C yang berubah-ubah tidak berpengaruh terhadap hasil.

Tabel 7. Hasil Prediksi Klasifikasi SVM

Nilai Parameter C dan σ	<i>k-mean</i>		<i>kernel k-means</i>		Sendiri	
	Benar	Salah	Benar	Salah	Benar	Salah
$C=2, \sigma=0,01$	1.473	0	1.473	0	1.418	55
$C=0,01, \sigma=1$	1.473	0	1.473	0	1.223	250
$C=0,01, \sigma=0,01$	1.473	0	1.473	0	1.418	55

Metode terakhir yang digunakan adalah dengan CART. Hasil pengklasifikasian dengan metode CART ditampilkan pada Tabel 8. Berdasarkan Tabel 8 dapat diketahui bahwa yang lebih banyak mengklasifikasikan dengan tepat secara berurutan adalah dengan *cluster k-means*, *cluster kernel k-means* dan asumsi sendiri.

Tabel 8. Hasil Prediksi Klasifikasi CART

Pengklasifikasian	<i>k-mean</i>	<i>kernel k-means</i>	sendiri
benar	1.472	1.087	1.100
salah	1	386	373

Tabel 9. Nilai APER

Metode Pengelompokan	Nilai APER		
	Regresi Logistik	SVM	CART
<i>k-mean</i>	57,298	0,000	0,0007
<i>kernel k-means</i>	40,665	0,000	0,2621
sendiri	31,839	3,734	0,2532

Setelah diketahui pengklasifikasian untuk masing-masing metode, dilakukan pemilihan model terbaik dengan kriteria nilai APER yang paling kecil. Nilai APER ditampilkan pada Tabel 9. Hasil yang diperoleh berdasarkan Tabel 9 metode yang paling baik adalah pengklasifikasian dengan

SVM dimana variabel respon diperoleh berdasarkan hasil pengelompokan *cluster k-means* atau *cluster kernel k-means*.

SIMPULAN

Hasil pemaparan pada pembahasan sebelumnya diperoleh bahwa hasil pengelompokan dengan asumsi sendiri lebih baik dibandingkan yang metode yang lain. Pada metode pengklasifikasian diperoleh bahwa hasil klasifikasi SVM lebih baik dibanding dua metode yang lain. Namun apabila metode pengelompokan dan klasifikasi digabung, diperoleh bahwa pengklasifikasian dengan SVM dimana variabel respon diperoleh berdasarkan hasil pengelompokan *cluster k-means* atau *cluster kernel k-means* dapat menghasilkan APER yang paling kecil.

REFERENSI

- Agresti, A. (2007). *An introduction to categorical data analysis second edition*. USA: A John Wiley & Sons, Inc.
- Epa Suryanto & Santi Wulan Purnami. (2015). Perbandingan reduced support vector machine dan smooth support vector machine untuk klasifikasi large data. *Jurnal sains dan seni ITS*, vol. 4, No.1, (2015)
- Gunn S. R. (1998). *Support vector machines for classification and regression*. Technical Report. University of Southampton.
- Johnson, R.A., & Winchern, D.W. (2007). *Applied multivariate statistical analysis*. USA: Pearson education international.
- Lim, T.S. (1997). *Contraceptive method choice*. Diakses pada tanggal 6 April 2012 dari website <http://archive.ics.uci.edu>.
- Nancy, JA. (1999). Contraception: Present and future. *Medical journal of Indonesia*, vol. 8(1).
- Santosa, B. (2007). *Data Mining: Teknik pemanfaatan data untuk keperluan bisnis*. Yogyakarta: Graha Ilmu.
- Sobirin. (2006). Mengenal lebih dalam aneka alat kontrasepsi. Diakses pada tanggal 6 April 2012 dari website <http://www.kafka.web.id/forum/kesehatan1>.