

KLASIFIKASI STATUS *HUMAN IMMUNODEFICIENCY VIRUS* (HIV) MENGUNAKAN METODE *SUPPORT VECTOR MACHINE* (SVM) DAN REGRESI LOGISTIK BINER

(Studi Kasus: Rumah Sakit Tiom Kabupaten Lanny Jaya Provinsi Papua)

Kasmawati¹, Rokhana Dwi Bekti^{2*}

^{1,2}Jurusan Statistika, FST, Institut Sains & Teknologi AKPRIND Yogyakarta
Email: kasmakhaffa777@gmail.com¹. rokhana@akprind.ac.id²

*corresponding author

Abstract. *There are several statistical methods in data mining that are used to classify between two classes, namely the support vector machine method and binary logistic regression, both of these methods have the same type of data variable, namely the response variable which uses the nominal variable value. Classification is the process of predicting labels / classes to be carried out on data that has not been given a label / class based on the function or pattern that has been formed according. This study discusses HIV status in Tiom Hospital in Lanny Jaya Papua Regency and uses support vector machine methods and binary logistic regression. This method serves to see the classification of the accuracy of HIV status based on predictor variables including sex (X1), age (X2), marital status (X3), education level (X4), area of residence (X5), and income (X6), with the dependent variable is that HIV status is categorized into two, namely 1 and -1, with value 1 stating that HIV positive and value -1 state negative HIV while logistic regression is a method of data analysis used to find the relationship between the response variable (Y) which is binary or dichotomy with predictor variables (X) that are polycotomous. The Y response variable consists of 2 categories namely "success" and "failure" denoted 1 and 0. This study discusses classification using the support vector machine method and binary logistic regression in determining the factors that affect the status of Human Immunodeficiency Virus (HIV) on Tiom. The Tiom Hospital data in 2018 has 37 HIV positive people or 25% and 113 HIV negative people or 75%. From the results of the classification analysis it was found that the best model was a classification support vector machine model using a linear kernel with a parameter cost (C) of 0.1 and an error of 0.268657.*

Keywords: *HIV status, Support Vector Machine, Binary Logistic Regression.*

Abstrak. Terdapat beberapa metode statistik dalam data mining yang digunakan untuk mengklasifikasikan antar dua kelas yaitu metode *support vector machine* dan regresi logistik biner, kedua metode ini memiliki persamaan jenis variabel data yaitu pada variabel responnya bernilai nominal. Klasifikasi adalah proses memprediksi label/kelas yang akan dilakukan pada data yang belum diberikan label/kelas berdasarkan fungsi atau pola yang telah dibentuk. Penelitian ini membahas tentang status HIV di RS Tiom Kabupaten Lanny Jaya provinsi Papua dan menggunakan metode *support vector machine* dan regresi logistik biner. Metode ini berfungsi untuk melihat klasifikasi ketepatan status HIV berdasarkan variabel prediktor diantaranya jenis kelamin (X_1), umur (X_2), status pernikahan (X_3), tingkat pendidikan (X_4), daerah tempat tinggal (X_5), dan pendapatan (X_6), dengan variabel dependen adalah status HIV dikategorikan menjadi dua yaitu 1 dan -1, dengan nilai 1 menyatakan bahwa positif HIV dan nilai -1 menyatakan negatif HIV sedangkan regresi logistik merupakan suatu metode analisis data yang digunakan untuk mencari hubungan antara variabel respon (Y) yang bersifat biner atau dikotomi dengan variabel prediktor (X) yang bersifat polikotomus. Dari variabel respon Y terdiri dari 2 kategori yaitu "sukses" dan "gagal" yang dinotasikan 1 dan 0. Penelitian ini membahas pengklasifikasian dengan menggunakan metode *support vector machine* dan regresi logistik biner dalam menentukan faktor-faktor yang mempengaruhi status *Human Immunodeficiency Virus* (HIV) di Tiom. Data RS Tiom tahun 2018 jumlah yang berstatus positif HIV sebanyak 37 orang atau 25% dan jumlah status negatif HIV sebanyak 113 orang atau 75%. Dari hasil analisis klasifikasi didapatkan bahwa model terbaik adalah model klasifikasi support vector machine dengan menggunakan kernel linier dengan parameter cost (C) sebesar 0.1 dan error sebesar 0.268657.

Kata Kunci: *Status HIV, Support Vector Machine, Regresi Logistik Biner.*

1. PENDAHULUAN

Terdapat beberapa masalah kesehatan didunia hingga saat ini belum bisa terselesaikan, salah satu permasalahan kesehatan yang sekarang menjadi *Global Issues* adalah *Human Immunodeficiency Virus* (HIV) dan AIDS. Kasus yang menyerang sistem kekebalan tubuh senantiasa meningkat dari tahun ke tahun bahkan Indonesia merupakan negara dengan penyebaran HIV/AIDS paling cepat di Asia Tenggara (Merati dalam Noer, 2004).

Menurut data laporan terbaru Dinas Kesehatan Provinsi Papua, kasus HIV/AIDS di Provinsi Papua meningkat dimana pada bulan Desember 2017 jumlah kasus sebanyak 32.000 lebih dan pada bulan Juni 2018 sebanyak 37.991 orang. Kendala yang dihadapi dalam penanganan masalah HIV/AIDS di Papua, yaitu tidak semua orang dengan HIV/AIDS mengakses layanan obat anti Retvorial/ARV untuk menekan pertumbuhan virus HIV/AIDS dalam tubuh seseorang.

Kasus HIV/AIDS pertama kali ditemukan pada tahun 1987 di Bali yaitu seorang penderita AIDS warga negara Belanda, pada tahun berikutnya HIV/AIDS ditemukan di Jakarta dan Surabaya, dan semakin banyak provinsi yang melaporkan adanya kasus HIV/AIDS. Jumlah penderita HIV/AIDS cenderung meningkat terus dan daerah yang terinfeksi pun cenderung meluas. Penyebaran di Indonesia terutama sangat dipengaruhi oleh perilaku seksual berisiko dan penyalahgunaan narkotika, psikotropika dan zat adiktif (Napza) (Nasution, dkk, 2000).

HIV adalah jenis parasit obligat yaitu virus ini hanya dapat menginfeksi manusia dan virus ini hidup dan air berkembangbiak pada sel darah putih manusia. HIV akan ada pada cairan tubuh yang mengandung sel darah putih, seperti darah, cairan plasenta, air mani atau cairan sperma, cairan sumsum tulang, cairan vagina, susu ibu (tempointeraktif.com, 2004). HIV menggambarkan tingkat permasalahan kesehatan masyarakat yang berkaitan dengan faktor penyebab kematian pasien diantaranya yaitu: jenis kelamin, umur, status pernikahan, tingkat pendidikan, daerah tempat tinggal, pendapatan dan transfusi darah. Sehingga perlu dilakukan kegiatan analisis dan studi kasus untuk melihat klasifikasi ketepatan pasien HIV (Disa Novianti, 2015)

Metode klasifikasi yang digunakan dalam penelitian ini adalah metode *Support Vektor Machine* (SVM) dan akan dibandingkan dengan metode Regresi Logistik Biner. Kedua metode ini memiliki persamaan jenis nilai variabel data yaitu pada variabel responnya dimana menggunakan nilai variabel nominal. SVM merupakan salah satu bagian dari data mining yang digunakan untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi (Santosa, 2007).

SVM merupakan suatu teknik untuk menemukan hyperplane yang bisa memisahkan dua set data dari dua kelas yang berbeda (Vapnik, 1999). SVM memiliki kelebihan diantaranya adalah dalam menentukan jarak menggunakan support vector sehingga proses komputasi menjadi cepat (Vapnik, 1995). Penelitian tentang SVM telah dilakukan oleh Rustam, et al (2003) yaitu membandingkan metode klasifikasi K-Nearest Neighbor (KNN) dengan metode SVM diperoleh kesimpulan bahwa SVM memiliki kinerja yang lebih unggul, karena telah mampu 100% mengklasifikasikan data aroma berdasarkan kelas yang tepat. Selain itu Rachman dan Purnami (2012) yang melakukan penelitian mengenai klasifikasi tingkat keganasan kanker dengan menggunakan metode regresi logistik dan SVM yang akhirnya diperoleh hasil bahwa tingkat akurasi menggunakan SVM lebih tinggi, yaitu sebesar 98.11%.

Klasifikasi adalah suatu proses pengelompokan objek berdasarkan kaidah atau standar tertentu yang ditetapkan. Dalam statistika banyak metode yang dapat digunakan dalam mengklasifikasikan suatu objek seperti *decision tree*, *naïve bayes classifier*, *k-nearest neighbor*, *support vector machine*, dan lainnya. Namun dalam beberapa penelitian menghasilkan kesimpulan bahwa metode SVM merupakan metode yang paling efektif dalam melakukan klasifikasi, seperti pada penelitian yang dilakukan oleh Niwayan Sumartini Saraswati (2013) yang melakukan analisis sentiment dengan menggunakan metode *naïve*

bayes dan *support vector machine* dan menghasilkan klasifikasi *support vector machine* lebih baik dibandingkan dengan metode *naïve bayes*.

Teknik *Support Vector Machine* (SVM) mulai diperkenalkan oleh Cortes dan Vapnik Pada tahun 1995. Ide utama dari SVM adalah membagi data menjadi dua kelompok melalui *hyperplane* yang optimal dengan cara memaksimalkan margin untuk meminimalkan *error*. *Hyperplane* adalah garis batas pemisah data antar kelas, sedangkan margin adalah jarak antara *hyperplane* dengan data terdekat pada masing-masing kelas. Adapun data terdekat dengan *hyperplane* pada masing-masing kelas inilah yang disebut *support vector*.

Analisis Regresi adalah teknik analisis yang menjelaskan hubungan antara dua atau lebih khususnya hubungan antara variabel-variabel yang mengandung sebab akibat (Nirwana, 2015). Regresi logistik biner adalah suatu metode analisis data yang digunakan untuk mencari hubungan antara variabel respon yang bersifat biner dengan variabel prediktor (Hosmer dan Lemeshow, 2000). Penelitian Becti, Pratiwi, Jatipaningrum, dan Auliana (2017) telah menggunakan metode regresi logistik untuk analisis pengaruh dan klasifikasi Konsumen dalam memilih minimarket. Kemudian Noeryanti, Suryowati, Setyawan, dan Aulia (2018) juga menggunakan regresi logistik untuk menganalisis studi mahasiswa.

Regresi logistik merupakan pendekatan pemodelan matematika yang dapat digunakan untuk mendeskripsikan hubungan beberapa variabel independen dengan variabel dependen. Model regresi logistik dibuat untuk mendeskripsikan peluang variabel dependen antara 0 dan 1 (Kleinbaum & Klein, 2010). Kondisi tersebut dapat ditemukan pada pengujian Statistika dengan menggunakan metode Regresi Logistik Biner. Berdasarkan penelitian Rahman (2014) dan Imaslihkah (2013), regresi logistik mempunyai ketepatan klasifikasi yang akurat. Menurut Antipov & Pokryshevskaya (2009), regresi logistik sangat menarik karena beberapa hal, yaitu (1) secara konsep sederhana, (2) mudah diinterpretasikan, dan (3) terbukti dapat menyediakan hasil yang akurat dan baik.

2. METODE

2.1. Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang akan diperoleh dari Rumah Sakit Tiom Kabupaten Lanny Jaya Provinsi Papua. Data sekunder merupakan sumber data penelitian yang diperoleh secara tidak langsung melalui media perantara. Data tersebut merupakan data HIV tahun 2018.

2.2. Variabel Penelitian

Variabel-variabel yang digunakan dalam penelitian ini diuraikan dalam tabel berikut:

Tabel 1. Variabel Penelitian

No.	Variabel	Jenis Skala	Keterangan	Definisi Operasional
Variabel Dependen				
1.	Y : Status HIV	Nominal	1. Positif (+) 2. Negatif (-)	Status HIV yang dimaksud adalah pasien yang positif HIV/ negatif HIV
Variabel Independen				
1.	X ₁ : Jenis Kelamin	Nominal	1. Laki-laki 2. Perempuan	Jenis kelamin pasien yang berada di Rumah Sakit Tiom
2.	X ₂ : Umur	Rasio	-	Umur pasien
3.	X ₃ : Status Pernikahan	Nominal	1. Menikah 2. Belum Menikah	Status pernikahan pasien
4.	X ₄ : Tingkat Pendidikan	Ordinal	1. Belum Sekolah 2. SD 3. SMP/MTS 4. SMA/SMK Perguruan Tinggi	Pendidikan terakhir pasien
5.	X ₅ : Daerah Tempat Tinggal	Nominal	1. Longgalo 2. Malagai 3. Popome	Daerah tempat tinggal berdasarkan 5 desa

No.	Variabel	Jenis Skala	Keterangan	Definisi Operasional
			4. Wiringgambut 5. Werime	
6	X ₆ : Pendapatan	Ratio	-	Jumlah Pendapatan per bulan dalam rupiah

2.3. Tahapan Analisis Data

Adapun tahapan analisis yang dilakukan pada penelitian ini adalah sebagai berikut:

- 1) Mengambil data HIV.
- 2) Melakukan analisis deskriptif dengan menyajikan frekuensi dan persentase untuk setiap variabel-variabel yang diamati.
- 3) Melakukan analisis klasifikasi dengan SVM dengan tahapan sebagai berikut:
 - a. Memasukkan data sesuai format *software*.
 - b. Membagi data training dan data testing dengan proposrsi data training sebesar 89(134 pemeriksa HIV)% dari populasi sebanyak 150 pemeriksa HIV dan sisanya sebesar 11(16 pemeriksa HIV)% sebagai data testing.
 - c. Menentukan fungsi kernel terbaik yang ingin digunakan sebagai permodelan *hyperplane* SVM yaitu kernel linier pada data training.
 - d. Menentukan nilai parameter C dan nilai parameter kernel linier yang akan digunakan sebagai permodelan *hyperplane* SVM. Penentuan nilai adalah dengan menggunakan grid search dan bernilai positif, semakin besar nilai C akan mempengaruhi pelanggaran yang akan dikenakan untuk tiap klasifikasi. Untuk nilai parameter kernel linier C akan digunakan adalah 0.1, 0.5, 1, 5, 10, 50, 100.
 - e. Mendapatkan nilai alpha dan b
 - f. Membentuk persamaan *hyperplane*
 - g. Melakukan prediksi klasifikasi pada data testing.
- 4) Evaluasi performansi model klasifikasi menggunakan matriks konfusi:
 - a. Menghitung akurasi klasifikasi hasil prediksi berdasarkan nilai APER dan nilai CCR.
 - b. Memilih nilai parameter dan fungsi kernel terbaik.
- 5) Melakukan analisis regresi logistik biner
- 6) Perbandingan metode *support vector machine* dan regresi logistik biner.

3. HASIL DAN PEMBAHASAN

3.1. Analisis Klasifikasi Status HIV dengan *Support Vector Machines*

Pada tahap klasifikasi akan diuji dengan menggunakan model kernel *Linier*, untuk klasifikasi menggunakan *support vector machine* ada beberapa nilai C (Cost) yang digunakan untuk mengetahui ketepatan klasifikasi yang terbaik pada kernel *linier*, nilai cost yang digunakan dalam analisis ini yaitu 0.1, 0.5, 1, 5, 10, 50 dan 100. Adapun nilai error model klasifikasi pada masing-masing nilai cost dapat dilihat pada tabel.

Tabel 2. Nilai Error Klasifikasi Pada Model Kernel Linier

Nilai Cost	Error	Jumlah Support Vector
0.1	0.268657	85
0.5	0.268657	80
1	0.268657	84
5	0.268657	86
10	0.268657	85
50	0.268657	87
100	0.268657	87

Berdasarkan Tabel 2. Dapat dilihat bahwa pada semua nilai cost dengan menggunakan kernel linier menghasilkan nilai error yang sama, sehingga model yang terbaik akan dilihat dari nilai cost yang paling rendah yaitu model klasifikasi *support vector machine* kernel linier dengan nilai cost sebesar 0.1 dan jumlah *support vector* sebanyak 85, untuk nilai alpha model kernel linier dengan parameter $c = 0.1$ dapat dilihat pada tabel berikut:

Tabel 3. Nilai Alpha Kernel Linier $C = 0.1$

No	Nilai Alpha	Data Ke
1	0.1	1
2	0.1	2
3	0.1	3
4	0.1	4
5	0.1	5
...
85	0.1	134

Dari nilai alpha atau *lagrange multiplier* di atas dapat membangun model klasifikasi *non-separable* dengan kernel *trick* linier dengan mengubah persamaan menjadi bentuk kernel dengan memanfaatkan nilai alpha, sehingga didapat fungsi sebagai berikut:

$$f(x) = (w^T \cdot x + b)$$

$$f(x) = \sum_{i=1}^n \alpha_1 y_i (x^T \cdot x) + b$$

Dengan $w = \sum_{i=1}^{85} \alpha_1 y_i \phi(x_i)$ dimana $\phi(x_i)$ = Kernel sehingga didapat persamaan optimal *hyperplane* untuk kernel *trick* linier sebagai berikut:

$$f(x) = \sum_{i=1}^{85} \alpha_1 y_i K(x, x) + b$$

$$f(x) = \sum_{i=1}^{85} \alpha_1 y_i (x^T \cdot x) + b$$

Dengan:

- α = Nilai Alpha (*Lagrange Multiplier*)
- y = Label/ Kelas Data
- N_{sv} = Jumlah Support Vector
- $(x_i^t x_i)$ = Fungsi Kernel Linier
- B = Nilai Bias/Constanta

Sehingga diperoleh jumlah *support vector machine* sebanyak 85, maka output dapat dibentuk dengan fungsi akhir sebagai berikut:

$$f(x) = \sum_{i=1}^{85} \alpha_1 y_i (x^T \cdot x) + 0.9998975$$

Model di atas diterapkan pada data testing pada Lampiran 8 untuk pasien pertama sehingga didapat hasil sebagai berikut:

$$f(x) = \sum_{i=1}^{85} \alpha_1 y_i (x' \cdot x) + 0.9998975$$

$$f(x) = 0.1 (-1) \left([1 \ 3 \ 1 \ 2.941 \ 1 \ 3] * \begin{bmatrix} 1 \\ 3 \\ 1 \\ 2.941 \\ 1 \\ 3 \end{bmatrix} \right) + 0.9998975$$

$$f_1 = -1.965$$

3.3. Analisis Regresi Logistik Biner

Dalam analisis ini telah dilakukan transformasi data dari bentuk ordinal menjadi interval atau biasa disebut dengan *Method of Succesive Interval_(MSI)*, Data yang ditranformasi adalah data tingkat pendidikan.

Tabel 3. Pembagian Data Training dan Testing

Pembagian	Persentase	Total
Data Pelatihan	89%	134
Data Pengujian	11%	16
Total	100%	150

Tabel 4. Uji Signifikan Variabel Bebas Secara Individu

Variabel	Koefisien	Z Value	P-Value
Jenis Kelamin1	-0.20863	-0.430	0.667
Umur	0.02364	0.051	0.959
Status Pernikahan1	0.30101	0.486	0.627
Tingkat Pendidikan	-1.34998	-4.511	0.016
Daerah Tempat Tinggal2	-1.20717	-1.544	0.123
Daerah Tempat Tinggal3	-0.22185	-0.322	0.748
Daerah Tempat Tinggal4	-0.44909	-0.660	0.509
Daerah Tempat Tinggal5	-0.10106	-0.129	0.897
Pendapatan	-0.31647	-0.898	0.369

3.4. Model Regresi Logistik Biner

Dari hasil uji Signifikan dapat dinyatakan bahwa semua variabel yang diteliti terdapat satu variabel yang berpengaruh signifikan terhadap Status HIV di RS Tiom Lanny Jaya Papua, Maka dapat dinyatakan dengan persamaan regresi logistik, sebagai berikut:

$$L_n = \left(\frac{p}{1-p} \right) = 2.6428 - 1.2289X_1$$

Berdasarkan persamaan regresi logistik maka, hasil persamaan regresi logistik kemudian di eksponenkan untuk mengetahui kecocokan dengan nilai Status HIV yaitu 1 (Positif HIV) atau 0 (Negatif HIV). Peluang kejadian terjadinya ketepatan Status HIV dapat dihitung menggunakan persamaan berikut:

$$\pi(x) = \frac{\exp(2.6428 - 1.2289X_1)}{1 + \exp(2.6428 - 1.2289X_1)}$$

Dari persamaan diatas, maka peneliti dapat menghitung nilai odds ratio. Berikut ini dapat dilihat peluang kejadian Status HIV dan nilai odds ratio pada tabel.

Tabel 5. Odds Ratio

Variabel	Exp (B)
Tingkat Pendidikan	0,292614

Berdasarkan Tabel 5. Diketahui bahwa nilai odds ratio didapatkan dari nilai eksponensial (B). Berikut ini akan dijelaskan nilai odds ratio berdasarkan variabel-variabel yang berpengaruh secara signifikan terhadap Status HIV yaitu:

Variabel Tingkat Pendidikan diketahui nilai odds ratio sebesar 0,292614. Angka tersebut menjelaskan bahwa peluang pengaruh variabel Tingkat Pendidikan terhadap pasien terkena HIV sebesar 0,292614 kali.

3.4. Pemilihan Model Terbaik dari Kedua Metode

Dengan membandingkan hasil error pada masing-masing metode yang diterapkan pada data testing, akan dipilih model yang menghasilkan nilai error paling rendah.

Tabel 6. Perbandingan Hasil Error Pada Data Train

Ukuran Ketepatan	SVM	Regresi Logistik Biner
APER	26.87	27.61
CCR	73.13	72.39

Berdasarkan Tabel 6. Pada metode SVM dengan ukuran ketepatan pada data training menghasilkan nilai APER yaitu sebesar 26.87 dan nilai CCR sebesar 73.13, sedangkan pada metode regresi logistik biner dengan ukuran ketepatan pada data training menghasilkan nilai APER yaitu sebesar 27.61 dan nilai CCR sebesar 72.39.

Tabel 7. Perbandingan Hasil Error Pada Data Test

Ukuran Ketepatan	SVM	Regresi Logistik Biner
APER	6.25	6.25
CCR	93.75	93.75

Berdasarkan Tabel 7. Pada masing-masing metode ukuran ketepatan pada data testing menghasilkan nilai yang sama yaitu nilai APER sebesar 6.25 dan nilai CCR sebesar 93.75.

4. KESIMPULAN

Berdasarkan hasil analisis klasifikasi dengan menggunakan metode *Support Vector Machine* dan Regresi Logistik Biner menghasilkan kesimpulan sebagai berikut:

1. Berdasarkan data pasien HIV Tahun 2018 diketahui bahwa sebanyak 37 atau (25%) pasien yang positif HIV dan sebanyak 113 atau (75%) pasien yang negatif HIV.
2. Hasil pemodelan status HIV dengan menggunakan metode *support vector machine* dan regresi logistik biner adalah sebagai berikut:
 - a. Metode *Support Vector Machine* menggunakan data training dengan kernel linier menghasilkan nilai error klasifikasi sebesar 0.268657.
 - b. Metode Regresi Logistik Biner menggunakan data training, pada analisis regresi variabel respon yang digunakan ada enam variabel, namun yang signifikan hanya tingkat pendidikan dan selebihnya tidak signifikan, sehingga didapatkan model regresi logistik biner yaitu:

$$\pi(x) = \frac{\exp(2.6428 - 1.2289X_1)}{1 + \exp(2.6428 - 1.2289X_1)}$$

3. Hasil prediksi masing-masing metode yaitu sebagai berikut:

a. *Support Vector Machine*

Hasil prediksi pada metode SVM yaitu ada 1 pasien yang diprediksi berstatus HIV dan 15 pasien yang tidak berstatus HIV.

b. Regresi Logistik Biner

Hasil prediksi pada metode regresi logistik biner yaitu ada 1 pasien yang diprediksi berstatus HIV dan 15 pasien yang tidak berstatus HIV.

Dari kedua metode yaitu SVM dan regresi logistik biner memiliki hasil prediksi yang sama.

4. Adapun nilai APER dan CCR masing-masing metode dengan error klasifikasi pada pengujian dengan menggunakan data testing adalah sebagai berikut:
- a. Metode *Support Vector Machine*
 1. Uji coba penggunaan model pada data training menghasilkan nilai APER sebesar 26.87% dengan tingkat akurasi CCR klasifikasi sebesar 73.13%.
 2. Uji coba penggunaan model pada data testing menghasilkan nilai APER sebesar 6.25% dengan tingkat akurasi CCR klasifikasi sebesar 93.75%.
 - b. Metode Regresi Logistik Biner
 1. Uji coba penggunaan model pada data training menghasilkan nilai APER sebesar 27.61% dengan tingkat akurasi CCR klasifikasi sebesar 72.39%.
 2. Uji coba penggunaan model pada data testing menghasilkan nilai APER sebesar 6.25% dengan tingkat akurasi CCR klasifikasi sebesar 93.75%.

Dari kedua metode yang digunakan pada kasus Status HIV Tahun 2018 dapat disimpulkan bahwa model terbaik untuk klasifikasi pada metode *support vector machine*. Hal ini dikarenakan, dengan menggunakan data training SVM menghasilkan nilai *Appearrent Error Rate* (APER) yang kecil dan nilai CCR yang besar. Namun berdasarkan data testing kedua metode memiliki nilai yang sama.

UCAPAN TERIMA KASIH

Dalam penyusunan makalah ini, banyak pihak yang telah memberikan dukungan kepada penulis. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan terima kasih kepada seluruh dosen dan pimpinan Jurusan Statistika Institut Sains & Teknologi AKPRIND Yogyakarta.

DAFTAR PUSTAKA

- Cotton, S., Puchalki, C.M., Sherman, S.N., et.al. (2006). Spirituality and religion in patients with HIV/AIDS. *Journal Gen Intern Medical*. (21), pp. S5–S13
- Bekti, R. D., Pratiwi, N., Jatipaningrum, M. T., & Auliana, D. (2017). Analisis Pengaruh Lokasi dan Karakteristik Konsumen dalam Memilih Minimarket dengan Metode Regresi Logistik dan Cart. *MEDIA STATISTIKA*, 10(2), 119-130.
- DTREG Predictive Modeling Software, <https://www.dtreg.com/solution/view/20>, diakses 17 september 2017.
- Ika Yuli Kumalasari (2013). Perilaku Berisiko Penyebab *Human Immunodeficiency Virus* (HIV) Positif. Skripsi. Universitas Negeri Semarang.
- Masmuda. M. 2011. Analisis Regresi Logistik Biner dan Aplikasinya untuk Mengidentifikasi faktor-faktor yang Mempengaruhi Tingkat Kesejahteraan Petani. Skripsi. Universitas Negeri Makassar. Makassar.
- Nirwana. S.R.A. 2015. Regresi Logistik Multinomial dan Penerapannya dalam Menentukan Faktor yang Berpengaruh pada Pemilihan Program Studi di Jurusan Matematika UNM. Skripsi. Universitas Negeri Makassar. Makassar.
- Noeryanti, N., Suryowati, K., Setyawan, Y., & Aulia, R. R. (2018, April). Logistic Regression and Path Analysis Method to Analyze Factors influencing Students' Achievement. In *IOP Conference Series: Materials Science and Engineering* (Vol. 335, No. 1, p. 012044). IOP Publishing.
- Notoatmodjo, S. (2007). Ilmu Perilaku Kesehatan. Jakarta: Rineka Cipta
- Nuha, M.U, Arieshanti, I. dan Purwanto, Y. 2012. Pengembangan Perangkat Lunak Prediktor Kebangkrutan Menggunakan Bagging Nearest Neighbor Support Vector Machine. *Jurnal Teknik Pomits* Vol. 1, No. 1, 1-6. Surabaya.

- Nurul Fuadi Yusuf (2017). Karakteristik Penderita Hiv/Aids Dengan Ko-Infeksi Tuberkulosis Paru Di Rumah Sakit Umum Pusat Wahidin Sudirohusodo Makassar Periode Januari Sampai Juni 2016. Skripsi. Universitas Hasanuddin. Makassar.
- Sasmitaningsih. I.P. (2010). Faktor-Faktor yang Mempengaruhi Persepsi Tenaga Keperawatan Perempuan Terhadap Penderita Hiv/Aids.
- Riska Prakasita Sahitayakti dan Kartika Fithriasari (2015). Klasifikasi Kesejahteraan Rumah Tangga di Provinsi Papua dengan Metode Regresi Logistik dan Support Vector Machine. *Jurnal Sains Dan Seni ITS* Vol. 4, No.2, (2015) 2337-3520 (2301-928X Print).
- Rizki. F. Widodo. D. A. A. dan Wulandari. S. P. 2015. Faktor Risiko Anemia Gizi Besi pada Ibu Hamil di Jawa Timur Menggunakan Analisis Regresi Logistik. *Jurnal*. Vol. 4. No.2. Institut Teknologi Sepuluh Nopember. Surabaya.
- Santosa, B.2007. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Graha Ilmu: Yogyakarta.
- UNAID. (2009). *Statistik HIV/AIDS update*. Geneva: UNAIDS
- Vapnik, V dan Cortes,C.1995. *Support Vector Networks*. *Machine Learning*, 20, 273-297.
- Vapnik, V *The Nature of Statistical Learning Theory*, Springer-Verlag, New York Berlin Heidelberg, 1999 (2nd ed).
- Widoyono. 2011. *Penyakit Tropis Epidemiologi, Penularan, Pencegahan, dan Pemberantasannya*. Jakarta: Erlangga
- Zaki, 2014, *Data Mining and Analysis: Fundamental Concepts and Algorithms*, Cambridge University Press, 32 Avenue Of The America, New York.