

KLASIFIKASI JENIS PENYAKIT ERYTHEMATO-SQUAMOUS BERDASARKAN CIRI KLINIS DAN HISTOPATOLOGIS MENGUNAKAN METODE ANALISIS DISKRIMINAN VERTEX

Nurmaleni, Ayu Puji Rahayu dan Nina Fitriyati

Program Studi Matematika, Fakultas Sains dan Teknologi
Universitas Islam Negeri Syarif Hidayatullah Jakarta
Nurmaleniadek@gmail.com dan rayupuji@gmail.com

Abstract: This study discusses the classification of types of erythematous-squamous disease using the Vertex Discriminant Analysis (VDA) method based on clinical and histopathological examination results. We use three penalties in the VDA method i.e. Euclidian, Lasso, and Ridge where misclassification are assessed using Apparent Rate Error (APER) value and loss function minimization using Majorize Minimize Algorithm. There are 366 data which we will classify consists of 34 variables of clinical and histopathological examination results from 6 disease groups: psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, chronic dermatitis, and pityriasis rubra pilaris. The result shows that each penalty in the VDA method forms 5 discriminant functions to classify 6 disease groups. VDA with Euclidian penalty succeed to classify exactly 104 data from 110 training data with 27 explanatory variables consist of 12 clinical and 15 histopathological examination results. VDA with Lasso penalty succeed to classify exactly 102 data from 110 training data with 25 explanatory variables consist of 11 clinical and 14 histopathological examination results, and the VDA with Ridge penalty succeed to classify exactly 107 data from 110 training data with 34 explanatory variables consist of 12 clinical and 22 histopathological examination results.

Keywords: Apparent Rate Error, Euclidian pinalty, Lasso pinalty, Majorize Minimize Algorithm, Ridge penalty.

Abstrak: Penelitian ini membahas mengenai klasifikasi jenis penyakit erythematous-squamous menggunakan metode Vertex Discriminant Analysis (VDA) berdasarkan hasil pemeriksaan klinis dan histopatologis. Digunakan 3 pinalti pada metode VDA yaitu Euclidian, Lasso, dan Ridge dan kesalahan klasifikasi dinilai menggunakan Apparent Rate Error (APER). Data yang digunakan berjumlah 366 terdiri dari 34 buah peubah hasil pemeriksaan klinis dan histopatologis yang berasal dari 6 kelompok penyakit: *psoriasis*, *seboreic dermatitis*, *lichen planus*, *pityriasis rosea*, *cronic dermatitis*, dan *pityriasis rubra pilaris*. Hasil menunjukkan bahwa setiap pinalti pada metode VDA membentuk 5 buah fungsi diskriminan untuk membedakan 6 kelompok penyakit. VDA dengan pinalti Euclidian berhasil mengklasifikasikan dengan tepat 104 data dari 110 data training dengan 27 peubah penjas yang terdiri dari 12 ciri klinis dan 15 ciri hispatologis. VDA dengan pinalti Lasso berhasil mengklasifikasikan dengan tepat 102 data dari 110 data training dengan 25 peubah penjas yang terdiri dari 11 ciri klinis dan 14 ciri hispatologis. Sedangkan VDA dengan pinalti Ridge berhasil mengklasifikasikan dengan tepat 107 data dari 110 data training dengan 34 peubah penjas yang terdiri dari 12 ciri klinis dan 22 ciri hispatologis.

Kata kunci: Apparent Rate Error, pinalti Euclidian, pinalti Lasso, Algoritma Majorize Minimize, pinalti Ridge.

PENDAHULUAN

Erythematous-squamous adalah salah satu kelompok jenis penyakit kulit. Penyakit kulit yang tergabung dalam kelompok ini adalah *psoriasis*, *seboreic dermatitis*, *lichen planus*, *pityriasis rosea*, *chronic dermatitis* dan *pityriasis rubra pilaris*. Penelitian mengenai klasifikasi jenis penyakit erythematous-squamous berdasarkan ciri-ciri klinis erythema dan scaling dengan perbedaan yang sangat kecil pada gejala awal pasien perlu dilakukan untuk mendiagnosa dengan tepat dan akurat [5].

Penelitian ini membahas mengenai pengklasifikasian penyakit erythematous-squamous menggunakan Vertex Discriminant Analysis (VDA). Metode VDA pertama kali dikembangkan oleh Lange dan Wu [5] pada kasus multikategori linier. Untuk kasus multikategori linier, metode ini menghasilkan ketepatan klasifikasi yang lebih baik dibandingkan metode klasifikasi lainnya [10][3]. Metode ini dapat digunakan bahkan pada masalah berdimensi tinggi (*high dimensional*). Klasifikasi untuk membentuk fungsi diskriminan pada metode VDA dilakukan dengan meminimalkan fungsi tujuan menggunakan Algoritma Majorize-Minimize untuk mendekati fungsi kerugian dengan fungsi kuadrat menggunakan reweighted least square [4]. Fungsi tujuan terdiri dari fungsi kerugian ϵ -insensitif dan penambahan penalti. Penambahan penalty yang digunakan pada penelitian ini adalah penalty Euclidean, Lasso dan Ridge dimana masing-masing penalti memiliki karakteristik dan fungsi penalti yang berbeda. Kesalahan klasifikasi akan dinilai menggunakan nilai *Apparent Rate Error* (APER). Terdapat 366 data yang akan diklasifikasikan yang terdiri dari 34 buah peubah hasil pemeriksaan klinis dan histopatologis. Perbandingan hasil klasifikasi akan dilakukan berdasarkan banyaknya data yang berhasil diklasifikasikan dengan tepat menggunakan ketiga penalti pada VDA. Selain itu, akan ditentukan pula peubah penjas dari hasil pemeriksaan klinis dan histopatologis yang berpengaruh terhadap jenis penyakit erythematous-squamous berdasarkan penalti-pinalti tersebut.

TINJAUAN PUSTAKA

Vertex Discriminant Analysis (VDA)

VDA adalah salah satu metode statistik multivariat untuk mengklasifikasikan objek dengan mengembangkan metode analisis diskriminan menggunakan simpul (*vertex*) dari simpleks pada ruang Euclidean [3]. VDA dapat digunakan bahkan pada kasus data berdimensi tinggi (*high dimensional data*) yaitu ketika banyaknya peubah (p) jauh lebih besar dari banyaknya objek (n) [5]. Masalah yang akan terjadi akibat dimensi tinggi adalah overfitting, hal ini dapat ditangani dengan regularisasi yaitu mengatur estimasi koefisien regresi dengan menambahkan aturan penalti yang dapat mengecilkan nilai estimasi menuju nilai sebenarnya. Regularisasi yang dilakukan menggunakan kalibrasi dengan cross validation. Klasifikasi untuk membentuk fungsi diskriminan seperti dalam regresi multivariat yang dilakukan dengan minimisasi fungsi tujuan yang melibatkan fungsi kerugian ϵ -insensitif dan penambahan penalti pada peubah prediktor [3].

Simpul pada Ruang \mathbb{R}^{k-1}

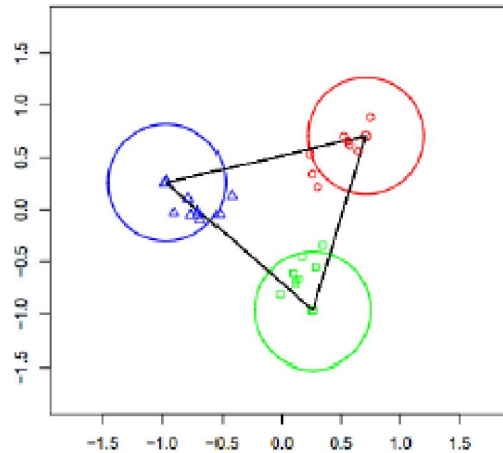
Simpul adalah suatu titik berjarak sama pada ruang simpleks yang digunakan untuk pengkodean indikator kelas dari masing-masing kelompok dimana masing-masing simpul menunjukkan kelompok yang berbeda. Agar setiap simpul memiliki jarak yang sama, VDA menggunakan simpul berjarak sama di ruang Euclidean \mathbb{R}^{k-1} untuk pengkodean indikator

kelas yang berbeda pada setiap kelompok, dimana k adalah banyaknya kelompok. Simpul berjarak sama di \mathbb{R}^{k-1} oleh Lange dan Wu didefinisikan sebagai berikut [6]:

$$v_j = \begin{cases} (k-1)^{-\frac{1}{2}} \mathbf{1} & \text{if } j = 1, \\ c\mathbf{1} + de_{j-1} & \text{if } 2 \leq j \leq k, \end{cases} \quad (1)$$

dengan $\mathbf{1}$: matriks berukuran $(1 \times n_j)$, e_j = vektor koordinat ke- j dalam \mathbb{R}^{k-1} (bernilai 1 pada saat ke- j dan 0 untuk lainnya), $c = -\frac{1+\sqrt{k}}{(k-1)^{\frac{3}{2}}}$ dan $d = \sqrt{\frac{k}{k-1}}$.

Dengan menggunakan Y dan X untuk masing-masing menunjukkan indikator kelompok (peubah respon) dan peubah penjelas dari masing-masing objek. Vektor Y sama dengan salah satu simpul dari simpleks tergantung kelompok mana yang dilabelkan, yaitu $Y_i = v_j$ [6]. Pada saat banyaknya kelompok 3 maka simpul berada pada \mathbb{R}^2 dengan koordinat simpul pada $v_1 = [0.707 \ 0.707]$, $v_2 = [0.258 \ -0.965]$ dan $v_3 = [-0.965 \ 0.258]$. Ilustrasi mengenai hal ini, digambarkan pada Gambar 1. Gambar ini menunjukkan ilustrasi simpul untuk tiga kelompok pada \mathbb{R}^2 .



Gambar 1. Ilustrasi Simpul untuk Tiga Kelas

Fungsi Kerugian

Fungsi kerugian (*loss function*) dinotasikan $L(y,x)$. Pada metode ini, fungsi kerugian yang digunakan adalah fungsi kerugian ε -insensitive $n^{-1} \sum_{i=1}^n L(y_i, x_i)$. Pada VDA, penerapan fungsi kerugian untuk kasus ini adalah $g(\mathbf{z}) = g(y_i - F(x_i))$, sedangkan jarak Euclidean ε -insensitive didefinisikan sebagai berikut [5]:

$$g(\mathbf{z}) = \|\mathbf{z}\|_{2,\varepsilon} = \max\{\|\mathbf{z}\|_2 - \varepsilon, 0\}, \quad (2)$$

dengan $\varepsilon = \frac{1}{2} \sqrt{\frac{2k+1}{k}}$, dan $\|\mathbf{z}\|_2$ adalah norm Euclidean dari vektor.

Fungsi Tujuan pada VDA

Fungsi tujuan merupakan nilai ekspektasi dari fungsi kerugian. Analisis diskriminan berusaha untuk meminimumkan nilai ekspektasi kerugian $E[L(Y,X)] = E\{E[L(Y,X)|X]\}$

agar nilai ekspektasi kerugian mendekati nilai populasi. Hal ini sulit dilakukan karena fungsi kerugian ϵ -insensitif yang non-trivial tidak dapat diturunkan sehingga pendugaan parameter pada VDA untuk proses klasifikasi dilakukan dengan cara meminimumkan $n^{-1} \sum_{i=1}^n L(y_i, f_i)$ ditambah aturan penalti untuk meregularisasi estimasi. Aturan penalti bergantung pada konstanta pemulus λ dan mengabaikan intersep konstanta b [3].

Penalti Euclidean pada VDA (VDA_E)

Penalti Euclidean didefinisikan sebagai $P(A) = \sum_{j=1}^k \|a_j\|_2$. Penambahan penalti pada tahap pembentukan fungsi diskriminan disarankan baik pada saat menggunakan semua parameter dalam kelompok maupun menyeleksi parameter dalam kelompok. Dalam klasifikasi multikategori, parameter slope untuk dimensi R^{k-1} membentuk kelompok sebenarnya. Norm Euclidean $\lambda_E \|a_l\|_2$ adalah penalti yang ideal karena mempertahankan konveksitas [4]. Pada metode VDA, penalti Euclidean efektif karena fitur dari penalti melakukan seleksi parameter untuk pemilihan model. Selain itu penalti Euclidean secara terus menerus dapat diturunkan meskipun a_l non trivial [5]. Setelah dilakukan identifikasi terhadap kelas indikator dengan simpul, fungsi tujuan terdiri dari fungsi kerugian dan penalti Euclidean, didefinisikan sebagai berikut [5]:

$$R(\theta) = \frac{1}{n} \sum_{i=1}^n g(\mathbf{y}_i - \mathbf{A}x_i - \mathbf{b}) + \lambda_E \sum_{j=1}^k \|a_j\|_2.$$

Penalti Lasso pada VDA (VDA_L)

Penalti lasso didefinisikan sebagai $P(A) = \sum_{j=1}^{k-1} \sum_{l=1}^p |a_{jl}|$. Sama halnya dengan penalti Euclid, Metode VDA dengan penalti Lasso juga efektif karena fitur dari penalti melakukan seleksi parameter untuk pemilihan model. Setelah dilakukan identifikasi terhadap kelas indikator dengan simpul, fungsi tujuan terdiri dari fungsi kerugian dan penalti lasso, didefinisikan sebagai berikut [5]:

$$R(\theta) = \frac{1}{n} \sum_{i=1}^n g(\mathbf{y}_i - \mathbf{A}x_i - \mathbf{b}) + \lambda_E \sum_{j=1}^k \|a_j\|_2.$$

Penalti Ridge pada VDA (VDA_R)

Penalti Ridge didefinisikan sebagai $P(A) = \sum_{j=1}^k \|a_j\|_2$ VDA linier mengasumsikan model regresi linier $f(x) = Ax + b$, dimana $A = (a_{jm})$ adalah matriks berukuran $(k-1) \times p$ dari slope dan $b = (b_j)$ adalah vektor berukuran $(k-1)$ dari kolom intersep. Untuk menghindari overfitting, dilakukan penerapan penalti pada parameter slope a_{jm} . Penambahan penalti Ridge digunakan pada saat seleksi peubah dengan cara memperkecil koefisien parameter yang berkorelasi mendekati nilai nol untuk peubah-peubah yang berkorelasi [6].

Setelah dilakukan identifikasi terhadap kelas indikator dengan simpul, fungsi tujuan terdiri dari fungsi kerugian dan penalti Ridge, didefinisikan sebagai berikut [6]:

$$R(\theta) = \frac{1}{n} \sum_{i=1}^n g(\mathbf{y}_i - \mathbf{A}x_i - \mathbf{b}) + \lambda \sum_{j=1}^{k-1} \sum_{l=1}^p a_{jl}^2, \quad (3)$$

dengan $n = n_1 + n_2 + n_3 + \dots + n_j$, $\theta = (A, b)$ menunjukkan himpunan semua parameter yang tidak diketahui, dan $\lambda \geq 0$ adalah parameter pemulus.

Penalti Ridge didefinisikan sebagai a_j adalah baris ke- j dari matriks $A_{(k-1) \times p}$ [2]. Pada fungsi kerugian Persamaan (3) nilai estimasi \hat{A} dan estimasi \hat{b}_{convex} (tidak memiliki solusi unik), oleh karena itu solusi unik dapat diestimasi dengan titik yang minimum. Ketika $\lambda > 0$, nilai minimum tunggal untuk fungsi tujuan $R(\theta)$ dijamin karena konveksitasnya yang ketat. Nilai estimasi \hat{A} dan estimasi \hat{B} diperoleh, setelah itu dapat menetapkan kasus baru ke simpul terdekat, dan dapat diprediksi oleh $\hat{Y} = \operatorname{argmin}_{j=1,2,\dots,k} \|v_j - \hat{A}x - \hat{b}\|_2$ untuk mendapatkan kelompok kategori yang sesuai [6].

Prediksi dengan metode VDA_R memiliki keunggulan dalam perhitungan statistika dan keunggulan dalam kecepatan perhitungan komputasi menggunakan algoritma terbaik untuk analisis diskriminan, akan tetapi VDA_R memiliki keterbatasan. Meskipun VDA_R mengecilkan parameter menuju 0 untuk parameter tertentu, VDA_R tidak dapat melakukan seleksi parameter pada model VDA_R yang dibentuk kecuali jika peneliti memiliki aturan sendiri untuk menghilangkan parameter dengan batas besaran parameter yang ditentukan sendiri [5].

Algoritma Majorize Minimize (MM)

Algoritma MM dapat digunakan untuk meminimisasi fungsi kerugian dengan cara optimasi iterasi [3]. Minimalisasi Persamaan (3) dapat diimplementasikan dengan menggabungkan algoritma MM dan estimasi ridge terkecil Algoritma MM. Mayorisasi pada M pertama dilakukan dengan cara memperoleh fungsi pengganti $g(\theta|\theta^m)$ untuk fungsi tujuan $f(\theta)$ pada titik tertentu di iterasi θ^m . Fungsi ini dikatakan mayorisasi fungsi $f(\theta)$ pada θ^m apabila

$$\begin{aligned} f(\theta^m) &= g(\theta|\theta^m), \\ f(\theta) &\leq g(\theta|\theta^m), \quad \text{untuk } \theta \neq \theta^m. \end{aligned}$$

Untuk masalah minimisasi, M kedua berarti 'minimisasi' Ketika fungsi pengganti dioptimalkan, fungsi tujuan akan didorong agar meningkat atau menurun sesuai kebutuhan. Pada VDA algoritma MM dibutuhkan untuk meminimumkan fungsi pengganti $g(\theta|\theta^m)$ daripada $f(\theta)$. Jika θ^{m+1} dinotasikan yang meminimumkan $g(\theta|\theta^m)$ sehingga $f(\theta^{m+1}) \leq f(\theta^m)$, fungsi ini diperoleh dari

$$\begin{aligned} f(\theta^{m+1}) &= g(\theta^{m+1}|\theta^m) + f(\theta^{m+1}) - g(\theta^{m+1}|\theta^m), \\ &\leq g(\theta^m|\theta^m) + f(\theta^m) - g(\theta^m|\theta^m) = f(\theta^m). \end{aligned}$$

Mengakibatkan $(\theta^{m+1}|\theta^m) \leq g(\theta^m|\theta^m)$. Penurunan dari hasil algoritma MM ini akan menghasilkan nilai yang stabil. Untuk tujuan komputasi fungsi ridge adalah yang idel untuk digunakan. Algoritma MM bertujuan untuk mencapai konvergen minimum global untuk fungsi tujuan yang non-trivial sehingga dapat menghasilkan solusi unik [5].

Mayorisasi Fungsi Kerugian dan Fungsi Tujuan

Mayorisasi fungsi kerugian $f(x) = \|x\|_\varepsilon$ yang dihasilkan dari persamaan Cauchy-Schwarz adalah sebagai berikut [3]:

$$g(x|x^m) = \begin{cases} \frac{1}{2\|x^m\|} \|x^2\| + \frac{1}{2} \|x^m\| - \varepsilon & \text{untuk } \|x^m\| \geq 2\varepsilon, \\ \frac{1}{4(\varepsilon - \|x^m\|)} \|x - x^m\|^2 & \text{untuk } \|x^m\| < \varepsilon, \\ \frac{1}{4(\varepsilon - \|z\|)} \|x - z\|^2 & \text{untuk } \varepsilon < \|x^m\| < 2\varepsilon, \end{cases} \quad (4)$$

dengan $z = cx^m$ dan $c = \frac{2\varepsilon}{\|x^m\|} - 1$.

Fungsi kerugian ε -insensitive pada Persamaan (2) dapat digunakan pada semua simpul kecuali dua kinks pada $\pm \varepsilon$, Mayorisasi fungsi kerugian ε -insensitive oleh Lange dan Wu sebagai berikut [3]:

$$\begin{aligned} g(\theta) &= \frac{1}{n} \sum_{i=1}^n g(y_i - f(x_i)), \\ &\leq \frac{1}{n} \sum_{i=1}^n w_i \|r_i - s_i\|_2^2 + \lambda \sum_{j=1}^{k-1} \|a_j\|^2 + f, \\ &= \sum_{j=1}^{k-1} \left[\frac{1}{n} \sum_{i=1}^n w_i (r_{ij} - s_{ij})^2 + \lambda \|a_j\|^2 \right] + f, \end{aligned} \quad (5)$$

dengan residual $r_i = y_i - f(x_i)$ dan bobot

$$w_i = \begin{cases} \frac{1}{2\|r_i^{(m)}\|} & \text{jika } \|r_i^{(m)}\| \geq 2\varepsilon, \\ \frac{1}{4(\varepsilon - \|r_i^{(m)}\|)} & \text{jika } \|r_i^{(m)}\| < \varepsilon, \\ \frac{1}{4(\|r_i^{(m)}\| - \varepsilon)} & \text{jika } \varepsilon < \|r_i^{(m)}\| < 2\varepsilon, \end{cases}$$

dengan

$$s_i = \begin{cases} 0 & \text{jika } \|r_i^{(m)}\| \geq 2\varepsilon, \\ r_i^{(m)} & \text{jika } \|r_i^{(m)}\| < \varepsilon, \\ \left(\frac{2\varepsilon}{\|r_i^{(m)}\|} - 1 \right) r_i^{(m)} & \text{jika } \varepsilon < \|r_i^{(m)}\| < 2\varepsilon. \end{cases}$$

Konstan pada f tergantung pada residual pada iterasi m tetapi tidak pada nilai dari \hat{A} dan \hat{b} . Seperti yang ditunjukkan oleh Lange dan Wu, Majorization Persamaan (5) bertujuan untuk memisahkan parameter pada setiap dimensi \mathbb{R}^{k-1} [6]. Langkah minimalisasi Algoritma MM, mereduksi $(k-1)$ ridge terkecil dengan ukuran n daripada satu lebih besar $(k-1) \times n$.

Apparent Error Rate (APER)

Apparent Error Rate adalah salah satu metode untuk menentukan nilai kesalahan klasifikasi yang biasa digunakan pada analisis diskriminan. Nilai APER didefinisikan sebagai fraksi objek pada data uji yang salah klasifikasi dengan fungsi klasifikasi pada data uji.

Nilai APER kemudian dikenali sebagai proporsi item dalam rangkaian pelatihan yang salah klasifikasi yang didefinisikan sebagai berikut [1]:

$$APER = \frac{n_{1m} + n_{2m}}{n_1 + n_2}$$

dengan n_1 : banyaknya objek pada kelompok 1, n_{1m} : banyaknya objek yang salah klasifikasi pada kelompok 1, n_2 : banyaknya objek pada kelompok 2, dan n_{2m} : banyaknya objek yang salah klasifikasi pada kelompok 2.

METODOLOGI PENELITIAN

Sumber Data

Data yang digunakan dalam penelitian ini adalah data dermatologi yang berasal dari Merz dan Murphy (1996) yang merupakan hasil pemeriksaan terhadap 366 pasien yang menderita penyakit *erythemato-squamous* [4]. Distribusi 366 data tersedia pada Tabel 1. Untuk tujuan penelitian, data dermatologi 366 pasien dibagi menjadi 2 yaitu data latih dan data uji dengan persentase masing-masing 70% dan 30%.

Tabel 1. Distribusi Kelas Jenis Penyakit Erythemato-Squamous

| Kode Kelompok | Kelompok | Banyak data | Banyak data latih | Banyak data uji |
|---------------|--------------------------|-------------|-------------------|-----------------|
| 1 | Psoriasis | 112 | 84 | 28 |
| 2 | Seboreic dermatitis | 61 | 37 | 24 |
| 3 | Lichen planus | 72 | 53 | 19 |
| 4 | Pityriasis rosea | 49 | 33 | 16 |
| 5 | Cronic dermatitis | 52 | 36 | 16 |
| 6 | Pityriasis rubra pilaris | 20 | 13 | 7 |

Data Peubah

Data yang digunakan berasal dari pemeriksaan klinis dan histopatologis. Tahap pertama, dilakukan pemeriksaan klinis yang terdiri dari 12 ciri-ciri pada pasien. Setelah itu sampel kulit diambil untuk mengevaluasi 22 ciri-ciri histopatologis tersebut. Nilai dari ciri-ciri histopatologis ditentukan dengan analisis sampel kulit menggunakan mikroskop [4]. Peubah yang digunakan ditetapkan berdasarkan pemeriksaan klinis dan pemeriksaan histopatologis. Daftar peubah berdasarkan pemeriksaan klinis tersedia pada Tabel 2 dan daftar peubah berdasarkan pemeriksaan histopatologis tersedia pada Tabel 3.

Tabel 2. Daftar Peubah Berdasarkan Pemeriksaan Klinis

| Peubah | Ciri klinis yang diamati | Nilai |
|--------|-----------------------------------|----------------|
| X1 | <i>Erythema</i> | 0,1,2,3 |
| X2 | <i>Scaling</i> | 0,1,2,3 |
| X3 | <i>Definite borders</i> | 0,1,2,3 |
| X4 | <i>Itching</i> | 0,1,2,3 |
| X5 | <i>Koebner phenomenon</i> | 0,1,2,3 |
| X6 | <i>Polygonal papules</i> | 0,1,2,3 |
| X7 | <i>Follicular papules</i> | 0,1,2,3 |
| X8 | <i>Oral mucosal involvement</i> | 0,1,2,3 |
| X9 | <i>Knee and elbow involvement</i> | 0,1,2,3 |
| X10 | <i>Scalp involvement</i> | 0,1,2,3 |
| X11 | <i>Family history</i> | 0 atau 1 |
| X34 | <i>Age</i> | ≥ 0 tahun |

Secara keseluruhan terdapat 34 ciri-ciri klinis dan histopatologis yang digunakan dalam penelitian menentukan jenis penyakit erthemato-squamous. Peubah family history memiliki nilai 1 untuk pasien yang memiliki riwayat penyakit yang didiagnosis tersebut dan nilai 0 untuk pasien yang tidak memiliki riwayat penyakit tersebut. Peubah age adalah umur pasien. Sedangkan ciri-ciri klinis maupun histopatologis lain memiliki nilai 0 sampai dengan 3, dimana nilai 0 menunjukkan tidak terdapatnya ciri-ciri tersebut dan nilai 1 sampai 3 menyatakan terdapatnya ciri-ciri tersebut. Nilai 3 menunjukkan ciri-ciri terbesar yang ada pada pasien, sedangkan nilai 1 dan 2 adalah ciri-ciri diantaranya [4].

Tabel 3. Daftar Peubah Hasil Pemeriksaan Histopatologis

| Peubah | Ciri klinis yang diamati | Nilai |
|--------|---|---------|
| X12 | <i>Melanin incontinence</i> | 0,1,2,3 |
| X13 | <i>Eosinophils in the infiltrate</i> | 0,1,2,3 |
| X14 | <i>PNL infiltrate</i> | 0,1,2,3 |
| X15 | <i>Fibrosis of the papillary dermis</i> | 0,1,2,3 |
| X16 | <i>Exocytosis</i> | 0,1,2,3 |
| X17 | <i>Acanthosis</i> | 0,1,2,3 |
| X18 | <i>Hyperkeratosis</i> | 0,1,2,3 |
| X19 | <i>Parakeratosis</i> | 0,1,2,3 |
| X20 | <i>Clubbing of the rete ridges</i> | 0,1,2,3 |
| X21 | <i>Elongation of the rete ridges</i> | 0,1,2,3 |
| X22 | <i>Thinning of the suprapapillary epidermis</i> | 0,1,2,3 |
| X23 | <i>Spongiform pustule</i> | 0,1,2,3 |
| X24 | <i>Munro microabcess</i> | 0,1,2,3 |
| X25 | <i>Focal hypergranulosis</i> | 0,1,2,3 |
| X26 | <i>Disappearance of the granular layer</i> | 0,1,2,3 |
| X27 | <i>Vacuolisation and damage of basal layer</i> | 0,1,2,3 |
| X28 | <i>Spongiosis</i> | 0,1,2,3 |
| X29 | <i>Saw-tooth appearance of retes</i> | 0,1,2,3 |
| X30 | <i>Follicular horn plug</i> | 0,1,2,3 |
| X31 | <i>Perifollicular parakeratosis</i> | 0,1,2,3 |
| X32 | <i>Inflammatory mononuclear infiltrate</i> | 0,1,2,3 |
| X33 | <i>Band-like infiltrate</i> | 0,1,2,3 |

Metode Analisis

Langkah analisis yang dilakukan penulis untuk mengklasifikasikan jenis penyakit *erythemat-squamous* adalah sebagai berikut :

- I. Penerapan data kasus untuk metode VDA dengan penalti Euclidean, penalti Lasso dan penalti Ridge dengan tahapan analisis sebagai berikut:
 1. Membagi data dermatologi menjadi dua, yaitu data latih dan data uji dengan proporsi 70% dan 30% dari seluruh data (366 data), data latih digunakan untuk membentuk fungsi diskriminan dan data uji digunakan untuk mengevaluasi kesalahan pengklasifikasian.
 2. Membentuk 3 fungsi diskriminan untuk memprediksi data baru masing-masing menggunakan metode VDA dengan penalti Euclidean, penalti Lasso dan penalti Ridge.
 3. Menghitung nilai APER pada data latih menggunakan metode VDA dengan penalti Euclidean, penalti Lasso dan penalti Ridge.
 4. Melakukan pengklasifikasian objek baru (data uji) pada fungsi diskriminan dengan penalti Euclidean, penalti Lasso dan penalti Ridge yang telah dibentuk,
 5. Menghitung masing-masing nilai APER dengan data uji pada metode VDA dengan penalti Euclidean, penalti Lasso dan penalti Ridge.
 6. Membandingkan hasil ketepatan klasifikasi jenis penyakit *erythemat-squamous* metode VDA dengan 3 penalti yaitu penalti Euclidean, penalti Lasso dan penalti Ridge.
- II. Pembentukan fungsi untuk metode VDA [3]
 1. Menetapkan iterasi awal $m = 0$, dan inialisasi $A^{(0)} = 0, b^{(0)} = 0$,
 2. Menentukan simpul dari masing-masing kelompok, definisika $y_i = v_j$ dimana v_j telah didefinisikan pada Persamaan (1),

3. Majorisasi fungsi kerugian regular sebagaimana ditunjukkan dalam Persamaan (5) dengan residual ke- i $r_i^{(m)} = y_i - f(x_i)$,
4. Minimumkan fungsi pengganti dengan *weighted least square* untuk menduga $A^{(m+1)}$ dan $b^{(m+1)}$ dengan menyelesaikan set persamaan linier pada $(k-1)$,
5. Jika $\|A^{(m+1)} - A^{(m)}\| < \gamma$, $|R(\theta^{(m+1)}) - R(\theta^{(m)})| < \gamma$ dengan $\gamma = 10^{-4}$ keduanya memenuhi maka iterasi berhenti,
6. Jika tidak, ulangi langkah iii sampai v sampai memenuhi dan fungsi diskriminan didapatkan.

Pada penelitian ini pembentukan fungsi diskriminan metode VDA dengan penalti Euclidean, penalti Lasso dan penalti Ridge serta penentuan peubah penjelas yang berpengaruh terhadap peubah respon dengan metode VDA menggunakan bantuan software R.

HASIL DAN PEMBAHASAN

Deskripsi Data

Hasil deskripsi statistik antara peubah respon (Y) dengan peubah prediktor (X) pada 6 kelompok jenis penyakit erythemato-squamous adalah rata-rata ciri yang sangat terlihat pada kelompok penyakit *psoriasis* adalah ciri pada peubah X1, X2, X3, X9, X10, X20, X21, X22 dan X32. Rata-rata ciri yang sangat terlihat pada kelompok penyakit *seboreic dermatitis* adalah ciri pada peubah X1, X2, X4, X17, X16, X28 dan X32. Rata-rata ciri yang sangat terlihat pada kelompok penyakit *lichen planus* adalah ciri pada peubah X1, X2, X3, X4, X6, X8, X12, X16, X17, X19, X25, X27, X29, X32 dan X33. Rata-rata ciri yang sangat terlihat pada kelompok penyakit pada kelompok penyakit *pityriasis rosea* adalah ciri pada peubah X1, X2, X16, X28 dan X32. Rata-rata ciri yang sangat terlihat pada kelompok penyakit *chronic dermatitis* adalah ciri pada peubah X1, X4, X15, X17, X21 dan X32. Rata-rata ciri yang sangat terlihat pada kelompok penyakit *pityriasis rubra pilaris* adalah ciri pada peubah X1, X2, X7, X9, X16, X17, X30, X31 dan X32.

Simpul Berjarak Sama di \mathbb{R}^{k-1}

Dengan perhitungan menggunakan persamaan (1) simpul v_j berjarak sama di \mathbb{R}^{k-1} untuk 6 kelompok tersedia pada Tabel 4.

Tabel 4. Simpul berjarak sama di \mathbb{R}^{k-1} untuk 6 kelompok

| v_1 | v_2 | v_3 | v_4 | v_5 | v_6 |
|-------|---------|---------|---------|---------|---------|
| 0,497 | 0,7865 | -0,3085 | -0,3085 | -0,3085 | -0,3085 |
| 0,497 | -0,3085 | 0,7865 | -0,3085 | -0,3085 | -0,3085 |
| 0,497 | -0,3085 | -0,3085 | 0,7865 | -0,3085 | -0,3085 |
| 0,497 | -0,3085 | -0,3085 | -0,3085 | 0,7865 | -0,3085 |
| 0,497 | -0,3085 | -0,3085 | -0,3085 | -0,3085 | 0,7865 |

Pembentukan Model VDA dengan Penalti Euclidean

Metode VDA dengan penalti Euclidean membentuk 5 fungsi diskriminan yaitu Y_1, Y_2, Y_3, Y_4, Y_5 untuk membedakan 6 kelompok berdasarkan banyaknya objek 256 data latih pasien dengan $\lambda = 0,00390625$, Ketepatan klasifikasi model diskriminan data latih yang terdiri dari 256 data pasien menggunakan metode VDA dengan penalti Euclidean pada setiap kelompok adalah banyaknya objek yang diklasifikasikan secara tepat oleh ke 5 fungsi

diskriminan menggunakan metode VDA dengan penalti Euclidean adalah sebanyak 253 data dari 256 data, artinya hanya terdapat 3 objek yang pengklasifikasiannya tidak tepat, Artinya metode VDA dengan penalti Euclidean pada data latih melakukan klasifikasi secara tepat sebesar 0,9921875 dengan APER sebesar 0,0078125,

Setelah fungsi diskriminan terbentuk, selanjutnya dilakukan evaluasi pada 110 data uji, Misalkan klasifikasi dilakukan pada data uji pasien nomer 1 secara manual, dengan klasifikasi sebenarnya pasien nomer 1 data uji termasuk ke dalam pasien yang didiagnosa memiliki penyakit *seboreic dermatitis* (kelompok 2),

Nilai masing-masing data peubah pasien nomer 1 disubstitusikan dalam 5 fungsi diskriminan yang telah dibentuk, sehingga diperoleh nilai $\hat{Y}_1 = 0,153227$, $\hat{Y}_2 = -0,06976$, $\hat{Y}_3 = 0,03603$, $\hat{Y}_4 = -0,04451$, $\hat{Y}_5 = -0,22965$, Berdasarkan Tabel 4,2 nilai jarak terkecil antara objek dengan simpul untuk 6 kelompok, data uji pasien nomer 1 diklasifikasikan ke dalam kelompok 2, Hal ini sesuai dengan klasifikasi sebenarnya, hasil klasifikasi metode VDA mengklasifikasikan dengan tepat data uji pasien 1,

Tabel 5. Jarak Antar Objek dengan Simpul menggunakan Metode VDA dengan Penalti Euclidean pada Data Pasien 1

| Kel, | $\ v_j - \bar{c}x_i - \bar{b}\ $ | Kel, | $\ v_j - \bar{c}x_i - \bar{b}\ $ |
|------|----------------------------------|------|----------------------------------|
| 1 | 1,330911 | 4 | 0,986556 |
| 2 | 0,761542 | 5 | 0,972556 |
| 3 | 1,336048 | 6 | 1,209922 |

Ketepatan klasifikasi model diskriminan data uji menggunakan metode VDA dengan penalti Euclidean pada setiap kelompok menunjukkan bahwa banyaknya objek yang diklasifikasikan secara tepat oleh ke 5 fungsi diskriminan menggunakan metode VDA dengan penalti Euclidean adalah sebanyak 104 data dari 110 data, artinya terdapat 6 objek yang diklasifikasikan salah atau tidak sama dengan klasifikasi sebenarnya, Metode VDA dengan penalti Euclidean pada data uji menghasilkan klasifikasi secara tepat sebesar 0,9272727 dengan APER sebesar 0,0727272,

Berdasarkan pembentukan fungsi diskriminan menggunakan metode VDA dengan penalti Euclidean dengan koefisien parameter yang tidak sama dengan 0 diperoleh 27 peubah penjelas yaitu 12 ciri-ciri klinis dan 15 ciri-ciri histopatologis yang berpengaruh terhadap peubah respon (Y) kelompok jenis penyakit erythemato-squamous, antara lain adalah peubah "X1", "X2", "X3", "X4", "X5", "X6", "X7", "X8", "X9", "X10", "X11", "X12", "X13", "X14", "X15", "X16", "X17", "X19", "X20", "X21", "X25", "X26", "X27", "X28", "X30", "X32", "X34", Peubah penjelas yang tidak berpengaruh adalah 7 ciri-ciri histopatologisnya yaitu "X18", "X22", "X23", "X24", "X29", "X31", "X33" masing-masing adalah ciri-ciri *hyperkeratosis*, *Thinning of the suprapapillary epidermis*, *Spongiform pustule*, *Munro microabcess*, *Saw-tooth appearance of retes*, *Perifollicular parakeratosis*, dan *Inflammatory mononuclear infiltrate*,

Pembentukan Model VDA dengan Penalti Lasso

Metode VDA dengan penalti Lasso membentuk 5 fungsi diskriminan untuk membedakan 6 kelompok berdasarkan banyaknya objek 256 data pasien dengan $\lambda = 0,02$, Dugaan koefisien yang dibentuk oleh VDA dengan penalti Lasso membentuk fungsi diskriminan Y_1, Y_2, Y_3, Y_4, Y_5 , Ketepatan klasifikasi model diskriminan data latih yang terdiri dari 256 data pasien menggunakan metode VDA dengan penalti Lasso menunjukkan bahwa banyaknya objek yang diklasifikasikan secara tepat oleh ke 5 fungsi diskriminan adalah

sebanyak 254 data dari 256 data, artinya hanya terdapat 2 objek yang pengklasifikasiannya kurang tepat dengan kata lain berbeda dengan klasifikasi sebenarnya, Artinya metode VDA dengan penalti Lasso pada data latih melakukan klasifikasi secara tepat sebesar 0,96484375 dengan APER sebesar 0,03515625.

Ketepatan klasifikasi model diskriminan data uji menggunakan metode VDA dengan penalti Lasso pada setiap kelompok menunjukkan bahwa banyaknya objek yang diklasifikasikan secara tepat oleh ke 5 fungsi diskriminan menggunakan metode VDA dengan penalti Lasso adalah sebanyak 102 data dari 110 data, artinya hanya terdapat 8 objek yang diklasifikasikan salah atau tidak sama dengan klasifikasi sebenarnya, Metode VDA dengan penalti Lasso pada data uji melakukan klasifikasi secara tepat sebesar 0,9454545 dengan APER sebesar 0,0545454.

Berdasarkan pembentukan fungsi diskriminan menggunakan metode VDA dengan penalti Lasso dengan koefisien parameter yang tidak sama dengan 0 diperoleh 25 peubah penjelas yaitu 11 ciri-ciri klinis dan 14 ciri-ciri histopatologis yang berpengaruh terhadap peubah respon (Y) kelompok jenis penyakit *erythemato-squamous*, antara lain adalah peubah "X1", "X2", "X3", "X4", "X5", "X6", "X7", "X8", "X9", "X10", "X11", "X12", "X13", "X14", "X15", "X16", "X17", "X20", "X21", "X22", "X25", "X27", "X28", "X29", Peubah penjelas yang tidak berpengaruh adalah 9 ciri-ciri histopatologisnya yaitu "X18", "X19", "X23", "X24", "X26", "X30", "X31", "X32", "X33", "X34" masing-masing adalah ciri-ciri *hyperkeratosis*, *Parakeratosis*, *Spongiform pustule*, *Munro microabcess*, *Disappearance of the granular layer*, *Follicular horn plug*, *Perifollicular parakeratosis*, *Inflammatory mononuclear infiltrate*, *Band like infiltrate*, dan *Age*.

Pembentukan Model VDA dengan Penalti Ridge

Metode VDA dengan penalti Ridge membentuk 5 fungsi diskriminan untuk membedakan 6 kelompok berdasarkan banyaknya objek 256 data pasien dengan $\lambda = 0,05$, Dugaan koefisien yang dibentuk oleh VDA dengan penalti Ridge membentuk fungsi diskriminan Y_1, Y_2, Y_3, Y_4, Y_5 .

Ketepatan klasifikasi model diskriminan data latih yang terdiri dari 256 data pasien menggunakan metode VDA dengan penalti Ridge pada setiap kelompok menunjukkan bahwa banyaknya objek yang diklasifikasikan secara tepat oleh ke 5 fungsi diskriminan menggunakan metode VDA dengan penalti ridge adalah sebanyak 254 data dari 256 data, artinya hanya terdapat 2 objek yang pengklasifikasiannya tidak tepat dengan kata lain berbeda dengan klasifikasi sebenarnya, Artinya metode VDA dengan penalti ridge pada data latih melakukan klasifikasi secara tepat sebesar 0,9921875 dengan APER sebesar 0,0078125.

Banyaknya objek yang diklasifikasikan secara tepat oleh ke 5 fungsi diskriminan menggunakan metode VDA dengan penalti ridge adalah sebanyak 107 data dari 110 data, artinya terdapat 3 objek yang diklasifikasikan salah atau tidak sama dengan klasifikasi sebenarnya, Artinya metode VDA dengan penalti Ridge pada data uji juga melakukan klasifikasi secara tepat sebesar 0,9727273 dengan APER 0,0272727.

Berdasarkan pembentukan fungsi diskriminan menggunakan metode VDA dengan penalti Ridge dengan koefisien parameter yang tidak sama dengan 0 untuk semua peubah penjelas yaitu 12 ciri-ciri klinis dan 22 ciri-ciri histopatologis yang berpengaruh terhadap peubah respon (Y) kelompok jenis penyakit *erythemato-squamous*, Pembentukan fungsi diskriminan menggunakan metode VDA dengan penalti Ridge tidak melakukan seleksi peubah tetapi hanya memperkecil koefisien parameter yang berkorelasi mendekati nilai nol.

KESIMPULAN

Berikut adalah kesimpulan dari pengklasifikasian jenis penyakit erythematous-squamous menggunakan metode VDA:

| | Hasil Ketepatan Klasifikasi | | |
|------------|-----------------------------|---------------|-------------------|
| | PenaltiEuclidean | Penalti Lasso | PenaltiRidge |
| Lamda | 0,00390625 | 0,02 | 0,05 |
| Data Latih | 0,9921875 | 0,96484375 | 0,9921875 |
| Data Uji | 0,9272727 | 0,9454545 | 0,97272727 |
| | Nilai APER | | |
| | PenaltiEuclidean | Penalti Lasso | Penalti Ridge |
| Data Latih | 0,0078125 | 0,03515625 | 0,0078125 |
| Data Uji | 0,0727272 | 0,0545454 | 0,0272727 |
| Rata-rata | 0,0402698 | 0,04485085 | 0,0175426 |

Berdasarkan hasil klasifikasi tersebut, metode VDA dengan penalti Ridge lebih tepat digunakan karena memiliki nilai APER terkecil penalty lainnya. Selain itu, VDA dengan pinalti Euclidian berhasil mengklasifikasikan dengan tepat 104 data dari 110 data training dengan 27 peubah penjelas yang terdiri dari 12 ciri klinis dan 15 ciri hispatologis. VDA dengan pinalti Lasso berhasil mengklasifikasikan dengan tepat 102 data dari 110 data training dengan 25 peubah penjelas yang terdiri dari 11 ciri klinis dan 14 ciri hispatologis.

Pada penelitian ini, metode VDA digunakan untuk kasus multikategori linier, namun ada kemungkinan bahwa klasifikasi penyakit ini dapat dilakukan menggunakan menggunakan multikategori non-linier. Salah satu metode VDA untuk multikategori non-linier adalah VDA berbasis RKHS sehingga dapat dibandingkan hasil pengklasifikasian menggunakan kedua metode tersebut.

REFERENSI

- [1] Johnson, R.A. dan Wichern, D.W., 2008, *Applied Multivariate Statistical Analysis*, 6th Edition, New Jersey (US): Pearson Prentice Hall.
- [2] Kustiyo, A. dan Tjandrasa, H., 2004, Model Neural Network dengan Inisialisasi Pembobot Awal Menggunakan Regresi Logistik Biner untuk Memprediksi Jenis Penyakit Erythematous-squamous, *Jurnal Ilmiah Ilmu Komputer*, Vol. 2 No.2, hal. 14-25.
- [3] Lange, K. dan Wu, T.T., 2008, An MM Algorithm for Multicategory Vertex Discriminant Analysis, *Journal of Computational Graphical Statistics* 17(3), pp. 527-544,
- [4] Merz, C. dan Murphy, P.M., 1996, UCI Repository of Machine Learning Databases [<http://www.ics.uci.edu/pub/machine-learningdatabases>] University of California, Irvine, Department of Information and Computer Science,
- [5] Wu, T.T. dan Lange, K, 2010, Multicategory Vertex Discriminant Analysis For High-Dimensional Data, *The Annals of Applied Statistics* 4(4): 1698-1721,
- [6] Wu, T.T. dan Wu, Y., 2012, Nonlinier Vertex Discriminant Analysis with Reproducing Kernels, *Stat Anal Data Min*: 5(2) pp. 167-176.