

PERBANDINGAN HASIL KLASIFIKASI JENIS PENYAKIT *ERYTHEMATO-SQUAMOUS* MENGUNAKAN ANALISIS DISKRIMINAN FISHER DENGAN PEREDUKSIAN ANALISIS KOMPONEN UTAMA KERNEL LINEAR DAN ANALISIS KOMPONEN UTAMA KERNEL GAUSSIAN

Khoeria Rosa, Nurmaleni, dan Nur Inayah

Program Studi Matematika, Fakultas Sains dan Teknologi
Universitas Islam Negeri Syarif Hidayatullah Jakarta

khoeriarosa13@gmail.com

Abstract: This study discusses the classification of erythematous diseases using the Fisher Discriminant Analysis (FDA) with the linear Kernel Principal Component Analysis (PCA) and the Gaussian Kernel PCA to reduce the dimensionality. We used dermatology data of the examination results of 366 patients with 34 variables. Data will be divided into training data and test data with proportions of 80% and 20% respectively. The cumulative proportion in PCA reduction is 75% to 95% and misclassification was assessed using Apparent Rate Error (APER). The result shows that, based on cumulative proportions, there are 18 principal components selected to form the Fisher's discriminant function in the linear Kernel PCA and there are 19 principal components selected to form the Fisher's discriminant function on the Gaussian Kernel PCA. In addition, the linear Kernel PCA succeed to classify 98.63% training data and 60.27% test data. Whereas, the Kernel Gaussian PCA succeed to classify 98.98% training data and 82.19% test data. Therefore, FDA method with Gaussian Kernel PCA appropriates to clasify the erythematous diseases.

Keywords: cumulative proportion, dermatology data, Fisher's discriminant function, misclassification.

Abstrak: Penelitian ini membahas mengenai klasifikasi jenis penyakit erythematous menggunakan Analisis Diskriminan Fisher (ADF) dengan pereduksian Analisis Komponen Utama (AKU) Kernel linear dan AKU Kernel Gaussian. Kami menggunakan 366 data dermatologi dengan 34 peubah yang merupakan hasil pemeriksaan terhadap pasien yang menderita penyakit erythematous. Data dibagi menjadi data latih dan data uji dengan proporsi masing-masing 80% dan 20%. Proporsi kumulatif yang digunakan untuk pereduksian menggunakan AKU adalah 75% sampai dengan 95% dan kesalahan klasifikasi dinilai menggunakan *Apparent Rate Error* (APER). Hasil menunjukkan bahwa, berdasarkan proporsi kumulatif, terdapat 18 komponen utama terpilih untuk membentuk fungsi diskriminan Fisher pada AKU Kernel linier dan terdapat 19 komponen utama terpilih untuk membentuk fungsi diskriminan Fisher pada AKU Kernel Gaussian. Selain itu, AKU Kernel linear berhasil mengklasifikasikan dengan tepat 98,63% data latih dan 60,27% data uji. Sedangkan AKU Kernel Gaussian berhasil mengklasifikasikan dengan tepat 98,98% data latih dan 82,19% data uji. Oleh karena itu, untuk jenis penyakit erythematous klasifikasi lebih tepat menggunakan ADF dengan AKU Kernel Gaussian dibandingkan dengan AKU Kernel linier.

Kata kunci: proporsi kumulatif, data dermatologi, fungsi diskriminan Fisher, kesalahan klasifikasi.

PENDAHULUAN

Kurangnya kesadaran manusia dalam menjaga kesehatan diri dan lingkungan sekitarnya dapat menyebabkan timbulnya berbagai jenis penyakit, salah satunya adalah jenis penyakit kulit *erythemat-squamous*. Jenis penyakit kulit yang tergabung dalam kelompok *erythemat-squamous* adalah *psoriasis*, *seboreic dermatitis*, *lichen planus*, *pityriasis rosea*, *cronic dermatitis* dan *pityriasis rubra pilaris*. Salah satu masalah yang muncul dalam ilmu dermatologi adalah sulitnya mendiagnosa untuk membedakan penyakit-penyakit *erythemat-squamous* karena pada pengamatan awal penyakit *erythemat-squamous* memiliki ciri-ciri klinis *erythema* dan *scalling* dengan sedikit perbedaan.

Beberapa penelitian telah dilakukan menggunakan data dermatologi jenis penyakit *erythemat-squamous*, diantaranya adalah Kustiyo dan Tjandrasa [6]. Kustiyo dan Tjandrasa [6] menggunakan model neural network dengan inisialisasi pembobot awal menggunakan regresi logistik biner untuk memprediksi jenis penyakit *erythemat-squamous*. Selanjutnya, Kustiyo dkk. [7] berhasil memprediksi jenis penyakit *erythemat-squamous* menggunakan jaringan syaraf tiruan dengan k-means sebagai praproses.

Pada penelitian ini, akan dilakukan pengklasifian jenis-jenis penyakit *erythemat-squamous* menggunakan Analisis Diskriminan Fisher (ADF) dengan pereduksian Analisis Komponen Utama (AKU) Kernel linear dan Kernel Gaussian. Perhitungan kesalahan klasifikasi dievaluasi nilai Apparent Error Rates (APER).

LANDASAN TEORI

Fungsi Kernel

Fungsi kernel merupakan suatu fungsi k untuk setiap $(\mathbf{x}_i, \mathbf{x}_j) \in \mathbf{X}$ memenuhi

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle,$$

dengan ϕ merupakan pemetaan \mathbf{X} dari ruang asal ke hasil kali dalam ruang fitur F [13],

$$\phi: \mathbf{x} \mapsto \phi(\mathbf{x}) \in F. \quad (1)$$

Fungsi kernel umumnya dibagi menjadi dua kelompok:

- 1) Fungsi kernel stasioner.

Salah satu contoh fungsi kernel stasioner adalah kernel linear, yaitu

$$k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j. \quad (2)$$

- 2) Fungsi kernel homogen atau lebih dikenal dengan *Radial Basic Function* (RBF) [1].

Salah satu contoh fungsi kernel homogen atau RBF adalah fungsi kernel Gauss:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2h_0^2}\right) \quad (3)$$

dengan $h_0 > 0$ [13]. Hasil pemetaan dari ruang asal ke ruang fitur sangat ditentukan oleh pemilihan parameter h_0 dalam fungsi kernel Gauss. Selain itu, pemilihan parameter h_0 dapat membantu dalam menentukan hasil pereduksian menggunakan AKU Kernel.

Terdapat beberapa metode yang dapat digunakan untuk memilih parameter h_0 pada fungsi kernel Gauss: penelitian Rathi dkk. [10] dan Widjaja dkk. [14]. Untuk data dengan p peubah dan n observasi, untuk pemilihan parameter Rathi dkk. [10] menggunakan:

$$h_0^2 = \frac{c}{p} \sum_{i=1}^p \min_{j \neq i} \|\mathbf{x}_i - \mathbf{x}_j\|^2; j = 1, 2, \dots, p \quad (4)$$

dengan c merupakan parameter control [10]. Pada Widjaja dkk. [14], pemilihan nilai parameter h_0 terdapat pada interval:

$$\frac{\hat{\sigma}^2}{100} < h_0^2 < 100\hat{\sigma}^2 \text{ dengan } \hat{\sigma}^2 = \frac{n}{p} \sum_{i=1}^p s_i^2 \quad (5)$$

dengan s_i^2 adalah varians peubah ke- i [14]. Adapun penentuan parameter h_0 pada penelitian ini akan menggunakan algoritma yang terdapat pada Lobo [8].

Analisis Komponen Utama Kernel (AKU Kernel)

AKU Kernel merupakan perluasan dari metode AKU dimana metode ini berguna untuk mereduksi dimensi peubah di ruang asal. AKU Kernel bekerja di ruang fitur berdimensi lebih tinggi dengan membentuk hasil kali dalam dari fungsi transformasi Φ [3]. Misalnya diberikan suatu himpunan dari vektor $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$. Matriks Gram (dilambangkan dengan \mathbf{G}) didefinisikan sebagai matriks berukuran $n \times n$ dengan elemen:

$$\mathbf{G}_{ij} = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle = k(\mathbf{x}_i, \mathbf{x}_j).$$

Dalam hal ini, matriks \mathbf{G} disebut juga sebagai matriks kernel \mathbf{K} . Matriks kernel \mathbf{K} adalah sebagai berikut [13]:

$$\mathbf{K} = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & k(\mathbf{x}_1, \mathbf{x}_2) & \cdots & k(\mathbf{x}_1, \mathbf{x}_n) \\ k(\mathbf{x}_2, \mathbf{x}_1) & k(\mathbf{x}_2, \mathbf{x}_2) & \cdots & k(\mathbf{x}_2, \mathbf{x}_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(\mathbf{x}_n, \mathbf{x}_1) & k(\mathbf{x}_n, \mathbf{x}_2) & \cdots & k(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix}.$$

Misal diberikan matriks data $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T$, dengan \mathbf{x}_i merupakan vektor objek data berdimensi p . Berdasarkan Persamaan (1), diperoleh matriks data: $\Phi = [\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \dots, \phi(\mathbf{x}_n)]^T$. Misal vektor rata-rata pada ruang fitur adalah $\bar{\phi} = \frac{1}{n} \sum_{i=1}^n \phi(\mathbf{x}_i)$, maka vektor objek data yang terkoreksi terhadap nilai tengah pada ruang fitur adalah $\tilde{\phi}(\mathbf{x}_i) = \phi(\mathbf{x}_i) - \bar{\phi}$.

Definisikan $\tilde{\mathbf{K}}$ sebagai matriks Kernel yang berukuran $n \times n$ dimana

$$\tilde{\mathbf{K}}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) - \frac{1}{n} \sum_{j=1}^n k(\mathbf{x}_i, \mathbf{x}_j) - \frac{1}{n} \sum_{i=1}^n k(\mathbf{x}_i, \mathbf{x}_j) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n k(\mathbf{x}_i, \mathbf{x}_j) \quad (6)$$

Asumsikan bahwa vektor rata-rata memiliki rata-rata $\sum_{i=1}^n \phi(\mathbf{x}_i) = 0$, maka matriks varians-kovarians dapat dihitung dengan

$$\mathbf{C} = \frac{1}{n-1} \sum_{i=1}^n \tilde{\phi}(\mathbf{x}_i) \tilde{\phi}(\mathbf{x}_i)^T. \quad (7)$$

Komponen utama kernel pada ruang fitur dicari dengan menyelesaikan persamaan eigen:

$$\mathbf{C}\mathbf{v} = \lambda\mathbf{v}. \quad (8)$$

Karena \mathbf{C} pada Persamaan (8) merupakan matriks semi-definit positif, maka $\lambda_m > 0$; $m = 1, 2, \dots, r$ untuk $r = \text{rank}(\mathbf{X})$ dan $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$. Definisikan hasil kali dalam pada ruang fitur, $\forall \mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^{PF}$ berlaku $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \mathbf{v}_1^T \mathbf{v}_2$, maka $\forall \lambda_m$ berlaku

$$\mathbf{v}_m = \sum_{i=1}^n \frac{\langle \tilde{\phi}(\mathbf{x}_i), \mathbf{v}_m \rangle}{\lambda_m(n-1)} \tilde{\phi}(\mathbf{x}_i).$$

Misal $\alpha_{mi} = \frac{\langle \tilde{\phi}(\mathbf{x}_i), \mathbf{v}_m \rangle}{\lambda_m(n-1)}$, sehingga diperoleh vektor eigen sebagai berikut

$$\mathbf{v}_m = \sum_{i=1}^n \alpha_{mi} \tilde{\phi}(\mathbf{x}_i). \quad (9)$$

Dengan menggunakan Persamaan (7) dan (9), Persamaan (7) pada ruang fitur, $\forall j = 1, \dots, n$, ekuivalen dengan $\langle \tilde{\phi}(\mathbf{x}_j), \mathbf{C}\mathbf{v}_m \rangle = \lambda_m \sum_{k=1}^n \alpha_{mk} \langle \tilde{\phi}(\mathbf{x}_j), \tilde{\phi}(\mathbf{x}_k) \rangle$. Persamaan eigen kernel dapat ditulis sebagai $\frac{1}{n-1} \tilde{\mathbf{K}} \alpha_{mk} = \lambda_m \alpha_{mk}$, dengan $k = 1, 2, \dots, n$ sehingga $\alpha_{mk} = (\alpha_{m1}, \alpha_{m2}, \dots, \alpha_{mn})^T$.

Dengan menggunakan hasil kali dalam antara vektor eigen yang telah dinormalisasi dengan vektor objek data pada ruang fitur, sehingga diperoleh komponen utama kernel.

Dengan menggunakan Persamaan (6) dan (9), karena matriks \mathbf{C} bersifat simetri, maka vektor eigen \mathbf{v}_m bersifat ortogonal. Vektor eigen \mathbf{v}_m yang ortonormal dapat ditulis sebagai $\tilde{\mathbf{v}}_m = \sum_{i=1}^n \tilde{\alpha}_{mi} \tilde{\phi}(\mathbf{x}_i)$, dengan $\tilde{\alpha}_{mi} = \frac{\alpha_{mi}}{((n-1)\lambda_m)^{1/2}}$. Menurut [1], koefisien proyeksi pada vektor eigen $\tilde{\mathbf{v}}_m$ untuk vektor uji \mathbf{x} adalah

$$\langle \tilde{\mathbf{v}}_m, \tilde{\phi}(\mathbf{x}) \rangle = \sum_{k=1}^n \tilde{\alpha}_{mk} \tilde{k}(\mathbf{x}_k, \mathbf{x}).$$

Sehingga komponen utama taklinearnya dapat ditulis sebagai berikut

$$\tilde{\Phi} \tilde{\mathbf{V}} = \tilde{\mathbf{K}} \tilde{\mathbf{A}}, \quad (10)$$

dengan $\tilde{\mathbf{V}}_r = (\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \dots, \tilde{\mathbf{v}}_r)$ dan $\tilde{\mathbf{A}}_r = (\tilde{\alpha}_1, \tilde{\alpha}_2, \dots, \tilde{\alpha}_r)$. Sedangkan komponen utama taklinearnya yang tidak terkoreksi terhadap nilai tengah dapat ditulis sebagai berikut

$$\Phi \tilde{\mathbf{V}} = \mathbf{K} \tilde{\mathbf{A}} \quad (11)$$

dengan $\mathbf{K} = k(\mathbf{x}_i, \mathbf{x}_j)$ [12].

Penelitian yang menggunakan metode AKU Kernel cukup banyak dilakukan, diantaranya Djakaria dkk. [2]. Djakaria dkk. [2] memvisualisasikan data iris menggunakan analisis komponen utama dan analisis komponen utama kernel. Pada penelitian tersebut, data iris divisualisasikan menggunakan metode AKU dan AKU Kernel. Adapun metode AKU Kernel pada penelitian tersebut digunakan untuk menyelesaikan permasalahan data non linear dan tidak dapat dipisahkan. Selain itu, visualisasi data berkelompok dengan AKU Kernel telah dilakukan oleh Lobo [8]. Metode klasifikasi yang digunakan dalam penelitian Lobo [8] adalah analisis diskriminan linear Fisher [8] dengan menggunakan dua komponen utama terpilih.

Analisis Diskriminan Fisher (ADF)

Analisis Diskriminan Fisher bertujuan untuk mengklasifikasikan populasi yang terpisah. Adapun dalam diskriminan ini tidak perlu diasumsikan bahwa populasi g multivariat normal. Namun diasumsikan bahwa matriks ragam peragam antar populasi sama yaitu $\Sigma_1 = \Sigma_2 = \dots = \Sigma_g = \Sigma$ dan berpangkat penuh.

Misalkan $\bar{\boldsymbol{\mu}}$ vektor rata-rata dari populasi gabungan dan \mathbf{B}_μ penjumlahan antar kelompok *cross product*, sehingga $\mathbf{B}_\mu = \sum_{i=1}^g (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})^T$ dengan $\bar{\boldsymbol{\mu}} = \frac{1}{g} \sum_{i=1}^g \boldsymbol{\mu}_i$.

Kombinasi linear untuk fungsi diskriminan Fisher adalah

$$Y = \mathbf{a}^T \mathbf{X}. \quad (12)$$

Pada umumnya, karena $\boldsymbol{\mu}_i$ dan Σ_i tidak diketahui. Misalkan sampel acak berukuran n_i dari populasi π_i , $i = 1, 2, \dots, g$. Himpunan data berukuran $n_i \times p$ dari populasi π_i oleh \mathbf{X}_i dan baris ke- j oleh \mathbf{x}_{ij}^T . Vektor rata-rata sampel adalah $\bar{\mathbf{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_{ij}$ dan matriks kovarian \mathbf{S}_i , $i = 1, 2, \dots, g$.

Definisikan sampel antara kelompok matriks \mathbf{B} yang mencakup ukuran sampel. Misalkan

$$\mathbf{B} = \sum_{i=1}^g n_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T \text{ dengan } \bar{\mathbf{x}} = \frac{\sum_{i=1}^g \sum_{j=1}^{n_i} \mathbf{x}_{ij}}{\sum_{i=1}^g n_i}.$$

Pendugaan dari Σ didasarkan pada sampel dalam matriks kelompok

$$\mathbf{W} = \sum_{i=1}^g (n_i - 1) \mathbf{S}_i = \sum_{i=1}^g \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^T,$$

dengan demikian $\mathbf{W}/(n_1 + n_2 + \dots + n_g - g) = \mathbf{S}_{gab}$ merupakan penduga dari Σ . Sebelum mempresentasikan sampel diskriminan, perhatikan bahwa \mathbf{W} merupakan konstanta ($n_1 +$

$n_2 + \dots + n_g - g$) dikali dengan \mathbf{S}_{gab} , sehingga \mathbf{a}^T memaksimumkan $\hat{\mathbf{a}}^T \mathbf{B} \hat{\mathbf{a}} / \hat{\mathbf{a}}^T \mathbf{S}_{gab} \hat{\mathbf{a}}$ juga memaksimumkan $\hat{\mathbf{a}}^T \mathbf{B} \hat{\mathbf{a}} / \hat{\mathbf{a}}^T \mathbf{W} \hat{\mathbf{a}}$. Selain itu, kita dapat mengoptimalkan $\hat{\mathbf{a}}$ dalam bentuk yang lebih umum sebagai vektor eigen $\hat{\mathbf{e}}_i$ dari $\mathbf{W}^{-1} \mathbf{B}$, karena jika $\mathbf{W}^{-1} \mathbf{B} \hat{\mathbf{e}} = \hat{\lambda} \hat{\mathbf{e}}$ maka $\mathbf{S}_{gab}^{-1} \mathbf{B} \hat{\mathbf{e}} = \hat{\lambda} (n_1 + n_2 + \dots + n_g - g) \hat{\mathbf{e}}$.

Misalkan $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_s > 0$ dengan $s \leq \min(g - 1, p)$ merupakan nilai eigen tak nol dari $\mathbf{W}^{-1} \mathbf{B}$ dan $\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_s$ yang sesuai dengan vektor eigen (sehingga $\hat{\mathbf{e}}^T \mathbf{S}_{gab} \hat{\mathbf{e}} = 1$). Maka vektor dari koefisien $\hat{\mathbf{a}}$ maksimum dengan rasio $\frac{\hat{\mathbf{a}}^T \mathbf{B} \hat{\mathbf{a}}}{\hat{\mathbf{a}}^T \mathbf{W} \hat{\mathbf{a}}} = \frac{\hat{\mathbf{a}}^T (\sum_{i=1}^g (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T) \hat{\mathbf{a}}}{\hat{\mathbf{a}}^T [\sum_{i=1}^g \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^T] \hat{\mathbf{a}}}$ diberikan oleh $\hat{\mathbf{a}}_1 = \hat{\mathbf{e}}_1$. Kombinasi linear $\hat{\mathbf{a}}_1^T \mathbf{x}$ disebut sampel pertama diskriminan. Pilihan $\hat{\mathbf{a}}_2 = \hat{\mathbf{e}}_2$ menghasilkan sampel kedua diskriminan, $\hat{\mathbf{a}}_2^T \mathbf{x}$, dan selanjutnya, kita dapatkan $\hat{\mathbf{a}}_k^T \mathbf{x} = \hat{\mathbf{e}}_k^T \mathbf{x}$, sampel ke- k diskriminan, $k \leq s$.

Adapun untuk pengklasifikasian observasi baru dapat dilakukan dengan mengklasifikasikan observasi baru berdasarkan persamaan diskriminan Fisher yaitu alokasikan \mathbf{x} ke kelompok π_k jika

$$\sum_{j=1}^r (\hat{y}_j - \bar{y}_{kj})^2 = \sum_{j=1}^r [\hat{\mathbf{a}}_j^T (\mathbf{x} - \bar{\mathbf{x}}_k)]^2 \leq \sum_{j=1}^r [\hat{\mathbf{a}}_j^T (\mathbf{x} - \bar{\mathbf{x}}_i)]^2, \quad (13)$$

dengan $\bar{y}_{kj} = \hat{\mathbf{a}}_j^T \bar{\mathbf{x}}_k$, untuk semua $i \neq k$ dan $r \leq s$ [4].

Evaluasi Hasil Pengklasifikasian

Terdapat suatu nilai ukuran dari prosedur klasifikasi yang tidak bergantung pada populasinya dan dapat dihitung untuk prosedur klasifikasi apapun. Ukuran ini disebut *Apparent Error Rate* (APER). Nilai APER menyatakan nilai proporsi sampel yang salah diklasifikasikan oleh fungsi klasifikasi [4]. Penentuan kesalahan pengklasifikasian dapat diketahui melalui tabel 1.

Tabel 1 Notasi untuk Banyaknya Observasi Hasil Pengklasifikasian.

Kelompok Asli	Kelompok Prediksi		Jumlah Observasi
	1	2	
1	n_{1c}	$n_{1M} = n_1 - n_{1c}$	n_1
2	$n_{2M} = n_2 - n_{2c}$	n_{2c}	n_2
Total	n_{1+}	n_{2+}	n_{++}

Berdasarkan Tabell, nilai APER dihitung menggunakan [4]:

$$\text{APER} = \frac{n_{1M} + n_{2M}}{n_1 + n_2}. \quad (14)$$

METODOLOGI

Sumber Data

Data yang digunakan dalam penelitian ini berasal dari Merz dan Murphy (1996). Data tersebut adalah data dermatologi yang merupakan hasil pemeriksaan terhadap 366 pasien yang menderita penyakit erythematous-squamous [9] yang terdiri atas 34 peubah. Karena nilai standar deviasi dari data jenis penyakit erythematous-squamous memiliki keragaman yang berbeda diantara ciri-ciri klinis dan histopatologis lainnya. Oleh karena itu, pada penelitian ini dilakukan standarisasi data karena terdapat peubah data yang memiliki nilai keragaman yang bersifat mendominasi.

Peubah Data

Peubah respon (Y) pada penelitian ini adalah kelompok jenis penyakit erythemato-squamous yaitu psoriasis, dermatitis seborea, lichen planus, pityriasis rosea, dermatitis kronik, dan pityriasis rubra pilaris. Untuk tujuan penelitian, data dermatologi untuk 366 pasien dibagi menjadi 2 yaitu data latih dan data uji dengan persentase masing-masing 80% dan 20% (Tabel 2). Sedangkan peubah prediktor (X) merupakan pemeriksaan klinis dan histopatologis yang berjumlah 34 ciri-ciri klinis dan histopatologisnya (Tabel 3 dan 4).

Tabel 2. Kelompok Jenis Penyakit Erythemato-Squamous.

No Kelompok	Kelompok	Jumlah	Data Latih	Data Uji
1	<i>Psoriasis</i>	112	96	16
2	<i>Dermatitis Seborea</i>	61	42	19
3	<i>Lichen Planus</i>	72	59	13
4	<i>Pityriasis Rosea</i>	49	40	9
5	<i>Dermatitis Kronik</i>	52	40	12
6	<i>Pityriasis Rubra Pilaris</i>	20	16	4

Tabel 3. Daftar Peubah Berdasarkan Pemeriksaan Klinis.

Peubah	Ciri-ciri klinis yang diamati	Nilai
X_1	<i>Erythema</i>	0, 1, 2, 3
X_2	<i>Scaling</i>	0, 1, 2, 3
X_3	<i>Definite borders</i>	0, 1, 2, 3
X_4	<i>Itching</i>	0, 1, 2, 3
X_5	<i>Koebner phenomenon</i>	0, 1, 2, 3
X_6	<i>Polygonal papules</i>	0, 1, 2, 3
X_7	<i>Follicular papules</i>	0, 1, 2, 3
X_8	<i>Oral mucosal involvement</i>	0, 1, 2, 3
X_9	<i>Knee and elbow involvement</i>	0, 1, 2, 3
X_{10}	<i>Scalp involvement</i>	0, 1, 2, 3
X_{11}	<i>Family history</i>	0 atau 1
X_{34}	<i>Age</i>	≥ 0 tahun

Metode Pengolahan Data

Pada pengolahan data, peneliti mengolah data jenis penyakit erythemato-squamous dengan menggunakan software Minitab 16, Matlab R2014a dan R i386 3.4.1. Berikut adalah langkah-langkah dalam pengolahan data:

1. Melihat plot pencar dan korelasi antar peubah untuk data dermatologi jenis penyakit erythemato-squamous,
2. Melakukan standarisasi data pada data dermatologi jenis penyakit erythemato-squamous,
3. Membagi data dermatologi jenis penyakit erythemato-squamous yang telah distandarisasi menjadi data dan data uji (Tabel 2),
4. Analisis Komponen Utama Kernel linear pada metode klasifikasi ADF.

Pada tahap ini, peneliti mereduksi data menggunakan AKU Kernel dengan fungsi kernel linear pada metode klasifikasi ADF. Adapun tahapan dalam pereduksian data menggunakan AKU Kernel linear pada metode klasifikasi ADF adalah sebagai berikut.

- a. Mereduksi peubah data latih menggunakan AKU Kernel dengan fungsi kernel linear pada Persamaan (2) menggunakan algoritma pada penelitian Lobo [8] sehingga diperoleh komponen utama dengan menggunakan Persamaan (10),
- b. Melihat proporsi kumulatif antara 75% sampai dengan 95% untuk memilih komponen utama terpilih yang dapat meminimumkan kesalahan klasifikasi,
- c. Komponen utama terpilih diklasifikasikan menggunakan metode klasifikasi ADF,

Tabel 4. Daftar Peubah Berdasarkan Pemeriksaan Histopatologis.

Peubah	Ciri-ciri Histopatologis yang diamati	Nilai
X_{12}	<i>Melanin incontinence</i>	0, 1, 2, 3
X_{13}	<i>Eosinophils in the infiltrate</i>	0, 1, 2, 3
X_{14}	<i>PNL infiltrate</i>	0, 1, 2, 3
X_{15}	<i>Fibrosis of the papillary dermis</i>	0, 1, 2, 3
X_{16}	<i>Exocytosis</i>	0, 1, 2, 3
X_{17}	<i>Acanthosis</i>	0, 1, 2, 3
X_{18}	<i>Hyperkeratosis</i>	0, 1, 2, 3
X_{19}	<i>Parakeratosis</i>	0, 1, 2, 3
X_{20}	<i>Clubbing of the rete ridges</i>	0, 1, 2, 3
X_{21}	<i>Elongation of the rete ridges</i>	0, 1, 2, 3
X_{22}	<i>Thinning of the suprapapillary epidermis</i>	0, 1, 2, 3
X_{23}	<i>Spongiform pustule</i>	0, 1, 2, 3
X_{24}	<i>Munro microabcess</i>	0, 1, 2, 3
X_{25}	<i>Focal hypergranulosis</i>	0, 1, 2, 3
X_{26}	<i>Disappearance of the granular layer</i>	0, 1, 2, 3
X_{27}	<i>Vacuolisation and damage of basal layer</i>	0, 1, 2, 3
X_{28}	<i>Spongiosis</i>	0, 1, 2, 3
X_{29}	<i>Saw-tooth appearance of retes</i>	0, 1, 2, 3
X_{30}	<i>Follicular horn plug</i>	0, 1, 2, 3
X_{31}	<i>Perifollicular parakeratosis</i>	0, 1, 2, 3
X_{32}	<i>Inflammatory mononuclear infiltrate</i>	0, 1, 2, 3
X_{33}	<i>Band-like infiltrate</i>	0, 1, 2, 3

- d. Menghitung total proporsi kesalahan klasifikasi pada data latih menggunakan APER pada Persamaan (14),
 - e. Membentuk fungsi diskriminan untuk komponen utama terpilih,
 - f. Melakukan transformasi data untuk menghasilkan fungsi diskriminan hasil pereduksian menggunakan AKU Kernel linear pada metode klasifikasi ADF,
 - g. Melakukan pengklasifikasian untuk objek baru pada data uji dengan menggunakan fungsi diskriminan yang terbentuk dari hasil pereduksian menggunakan AKU Kernel linear pada metode klasifikasi ADF pada Persamaan (13),
 - h. Menghitung total proporsi kesalahan klasifikasi pada data uji menggunakan APER pada Persamaan (14),
5. AKU Kernel Gaussian pada metode klasifikasi ADF
- Pada tahap ini, peneliti mereduksi data menggunakan AKU Kernel dengan fungsi kernel Gaussian pada metode klasifikasi ADF. Adapun tahapan dalam pereduksian data menggunakan AKU Kernel Gaussian pada metode klasifikasi ADF adalah sebagai berikut:
- a. Menghitung pendugaan parameter fungsi kernel Gaussian pada data latih menggunakan Persamaan (4) dan (5) untuk memperoleh hasil pereduksian dengan AKU Kernel Gaussian pada metode klasifikasi ADF dengan menggunakan algoritma pada Lobo [8] yang memiliki total proporsi kesalahan klasifikasi minimum,
 - b. Mereduksi peubah data latih menggunakan AKU Kernel dengan fungsi kernel Gaussian pada Persamaan (3) dengan melihat nilai parameter σ tertentu yang menghasilkan kesalahan klasifikasi minimum sehingga diperoleh komponen utama dengan menggunakan Persamaan (10) [8],
 - c. Melihat proporsi kumulatif antara 75% sampai dengan 95% untuk memilih komponen utama terpilih yang dapat meminimumkan kesalahan klasifikasi,
 - d. Mengklasifikasikan komponen utama terpilih menggunakan metode klasifikasi ADF,

- e. Menghitung total proporsi kesalahan klasifikasi pada data latih menggunakan APER pada Persamaan (14),
- f. Membentuk fungsi diskriminan untuk komponen utama terpilih,
- g. Melakukan transformasi data untuk menghasilkan fungsi diskriminan hasil pereduksian menggunakan AKU Kernel Gaussian pada metode klasifikasi ADF,
- h. Melakukan pengklasifikasian untuk objek baru pada data uji dengan menggunakan fungsi diskriminan yang terbentuk dari hasil pereduksian menggunakan AKU Kernel Gaussian pada metode klasifikasi ADF pada Persamaan (13),
- i. Menghitung total proporsi kesalahan klasifikasi pada data uji menggunakan APER pada Persamaan (14).

Adapun algoritma dalam pendugaan parameter fungsi kernel Gaussian pada AKU Kernel Gaussian adalah sebagai berikut [8]:

- a. Menghitung jarak antarobjek pada data latih menggunakan Persamaan (3),
- b. Mendefinisikan interval dari jarak terdekat hingga jarak terjauh pada data latih menggunakan Persamaan (4),
- c. Mereduksi peubah menggunakan AKU Kernel Gaussian untuk setiap nilai σ pada interval yang telah didefinisikan sehingga diperoleh komponen utama dengan menggunakan Persamaan (10),
- d. Untuk setiap objek dari komponen utama terpilih, diklasifikasikan menggunakan metode klasifikasi ADF,
- e. Menghitung total proporsi kesalahan klasifikasi data latih menggunakan APER dengan menggunakan Persamaan (14),
- f. Memilih σ yang meminimumkan kesalahan klasifikasi pada komponen utama terpilih.

Melakukan perbandingan hasil klasifikasi antara metode AKU Kernel linear pada metode klasifikasi ADF dengan AKU Kernel Gaussian pada metode klasifikasi ADF dengan melihat nilai APER keduanya.

HASIL DAN PEMBAHASAN

Asumsi Data

Untuk menunjukkan bahwa data yang digunakan merupakan data non linear dan memiliki korelasi antar peubah, dapat dilakukan dengan cara melihat plot pencar dan nilai korelasi antar peubah menggunakan korelasi pearson. Untuk plot pencar antar peubah data dermatologi jenis penyakit erythemato-squamous, karena peubah data yang digunakan pada penelitian ini banyak yang memiliki nilai dengan skala ordinal, sehingga dapat dilihat dengan jelas bahwa setiap peubah tidak dapat dipisah oleh garis linear antar kelompoknya. Selain itu, untuk nilai korelasi antar peubah pada data dermatologi jenis penyakit erythemato-squamous mencapai 0.942 yaitu peubah X_{12} dengan X_{27} dimana interpretasi untuk nilai korelasi tersebut menurut Sugiyono (2007) merupakan korelasi yang sangat kuat. Karena terdapat cukup banyak peubah yang memiliki nilai korelasi sangat kuat, sehingga terjadi gejala multikolinearitas. Multikolinearitas dapat terjadi karena peubah data yang digunakan bergantung linear dengan peubah lainnya. Oleh sebab itu, tidak dapat dihitung determinan matriksnya sehingga tidak memiliki invers dan disebut dengan matriks singular. Dengan demikian, data yang digunakan dalam penelitian ini merupakan data nonlinear dan memiliki korelasi antar peubah sehingga memenuhi asumsi yang dibutuhkan.

AKU Kernel Linear pada Metode ADF

Pereduksian untuk data dermatologi jenis penyakit erythemato-squamous dengan menggunakan metode AKU Kernel linear pada metode klasifikasi ADF dilakukan dengan membagi data menjadi data latih 80% yaitu sebanyak 293 objek dan data uji 20% yaitu

sebanyak 73 objek. Dengan menggunakan algoritma [8] pada *software* Matlab R2014a, menghasilkan nilai eigen dan proporsi kumulatif untuk masing-masing komponen utama. Karena belum terdapat patokan baku besar proporsi kumulatif yang dapat digunakan untuk memilih komponen utama terpilih [5] dalam melakukan pereduksian dengan menggunakan AKU Kernel linear pada metode klasifikasi ADF, peneliti melihat proporsi kumulatif antara 75% sampai dengan 95%. Untuk hasil pereduksian dengan menggunakan AKU Kernel linear pada metode klasifikasi ADF dengan melihat proporsi kumulatif antara 75% sampai dengan 95% pada data latih diperoleh hasil kesalahan klasifikasi seperti pada Tabel 5. Dengan melihat hasil kesalahan klasifikasi pada Tabel 5, peneliti memilih proporsi kumulatif 92% untuk melakukan perbandingan dengan AKU Kernel Gaussian. Oleh karena itu, terpilih 18 KU yang akan dilakukan pengklasifikasian dengan menggunakan metode klasifikasi ADF.

Adapun hasil klasifikasi untuk 18 komponen utama terpilih pada pereduksian data latih dengan menggunakan metode AKU Kernel linear pada metode klasifikasi ADF tertera pada Tabel 6. Dari Tabel 6 dapat dilihat bahwa nilai APER untuk metode AKU Kernel linear pada metode klasifikasi ADF untuk data latih diperoleh kesalahan klasifikasi sebesar 0.0136519 dengan 4 objek yang salah klasifikasi dari 293 objek. Hal tersebut dapat dilihat bahwa ketepatan klasifikasi untuk data latih adalah sebesar 0.9863481 dengan 289 objek yang tepat klasifikasi dari 293 objek.

Tabel 5. Kesalahan Klasifikasi AKU Kernel Linear dengan Proporsi Kumulatif antara 75% sampai dengan 95% pada Metode Klasifikasi ADF

Komponen Utama (KU)	Proporsi Kumulatif (%)	Kesalahan Klasifikasi
9	75.4713	0.030717
10	77.8443	0.03413
11	80.1284	0.027304
12	82.2126	0.023891
13	84.1825	0.023891
14	86.0847	0.023891
15	87.7931	0.027304
16	89.4788	0.027304
17	90.9214	0.017065
18	92.1179	0.013652
19	93.2116	0.013652
20	94.1624	0.013652

Tabel 6. Hasil Klasifikasi AKU Kernel Linear pada Metode Klasifikasi ADF untuk Data Latih

Kelompok Asli	Kelompok Prediksi						Total
	1	2	3	4	5	6	
1	96	0	0	0	0	0	96
2	0	39	0	3	0	0	42
3	0	0	59	0	0	0	59
4	0	1	0	39	0	0	40
5	0	0	0	0	40	0	40
6	0	0	0	0	0	16	16
Total	96	40	59	42	40	16	293

Adapun dengan menggunakan *library* MASS pada *software* R i386 3.4.1 [11], diperoleh fungsi diskriminan Fisher untuk 18 KU terpilih. Dengan melakukan transformasi pada persamaan fungsi diskriminan Fisher untuk komponen utama terpilih, sehingga diperoleh

model persamaan diskriminan Fisher hasil pereduksian AKU Kernel linear pada metode klasifikasi ADF adalah sebagai berikut:

$$\begin{aligned}
 \hat{Y}_1 &= -0.00134X_1 - 0.00703X_2 + 0.00715X_3 + 0.00879X_4 - 0.00705X_5 \\
 &\quad + 0.00613X_6 + 0.013786X_7 + 0.01846X_8 + 0.00813X_9 + 0.00164X_{10} \\
 &\quad - 0.00187X_{11} - 0.01708X_{12} - 0.01589X_{13} - 0.00909X_{14} + 0.01057X_{15} \\
 &\quad + 0.00116X_{16} + 0.011643X_{17} - 0.01995X_{18} + 0.01617X_{19} - 0.00213X_{20} \\
 &\quad + 0.019399X_{21} - 0.03443X_{22} + 0.04968X_{23} + 0.01099X_{24} + 0.00057X_{25} \\
 &\quad - 0.005099X_{26} - 0.02703X_{27} - 0.01522X_{28} + 0.00035X_{29} + 0.02658X_{30} \\
 &\quad + 0.04427X_{31} - 0.00795X_{32} - 0.01337X_{33} - 0.098022X_{34} \\
 \hat{Y}_2 &= 0.00173X_1 + 0.00035X_2 + 0.00118X_3 + 0.00681X_4 - 0.00264X_5 \\
 &\quad + 0.00214X_6 - 0.00533X_7 + 0.00674X_8 + 0.00122X_9 - 0.00161X_{10} \\
 &\quad - 0.01008X_{11} - 0.00378X_{12} - 0.007126X_{13} + 0.0012X_{14} + 0.01005X_{15} \\
 &\quad + 0.00911X_{16} - 0.00331X_{17} - 0.002894X_{18} + 0.02479X_{19} + 0.00193X_{20} \\
 &\quad + 0.01621X_{21} - 0.00317X_{22} - 0.016171X_{23} + 0.01318X_{24} + 0.02494X_{25} \\
 &\quad - 0.00703X_{26} - 0.011697X_{27} + 0.04187X_{28} - 0.00973X_{29} + 0.06143X_{30} \\
 &\quad + 0.03892X_{31} - 0.01603X_{32} + 0.01348X_{33} + 0.05445X_{34} \\
 \hat{Y}_3 &= -0.00256X_1 - 0.00373X_2 + 0.00039X_3 + 0.00242X_4 + 0.0006X_5 \\
 &\quad - 0.00287X_6 + 0.00054X_7 + 0.001391X_8 + 0.00061X_9 + 0.00541X_{10} \\
 &\quad + 0.00531X_{11} - 0.00097X_{12} + 0.00712X_{13} - 0.00355X_{14} - 0.00395X_{15} \\
 &\quad + 0.00201X_{16} - 0.00178X_{17} + 0.00046X_{18} - 0.01516X_{19} + 0.01304X_{20} \\
 &\quad - 0.01687X_{21} - 0.00112X_{22} + 0.02348X_{23} - 0.002399X_{24} - 0.00544X_{25} \\
 &\quad + 0.006398X_{26} + 0.00463X_{27} - 0.02677X_{28} + 0.00787X_{29} - 0.01221X_{30} \\
 &\quad + 0.00085X_{31} + 0.02049X_{32} - 0.0022X_{33} - 0.094692X_{34} \\
 \hat{Y}_4 &= 0.00008X_1 + 0.00025X_2 + 0.00083X_3 + 0.00296X_4 + 0.0072X_5 \\
 &\quad - 0.00323X_6 - 0.00053X_7 + 0.00507X_8 + 0.00922X_9 - 0.01136X_{10} \\
 &\quad + 0.00602X_{11} + 0.02042X_{12} - 0.005996X_{13} - 0.0087X_{14} + 0.01157X_{15} \\
 &\quad - 0.00195X_{16} - 0.00177X_{17} + 0.00215X_{18} - 0.00833X_{19} - 0.02188X_{20} \\
 &\quad + 0.01333X_{21} - 0.0352995X_{22} - 0.0035X_{23} + 0.02162X_{24} - 0.06086X_{25} \\
 &\quad - 0.002996X_{26} - 0.02273X_{27} + 0.02551X_{28} - 0.01416X_{29} - 0.00235X_{30} \\
 &\quad + 0.02711X_{31} + 0.00075X_{32} - 0.04015X_{33} + 0.00086X_{34} \\
 \hat{Y}_5 &= 0.00182X_1 - 0.00078X_2 - 0.00164X_3 - 0.00193X_4 - 0.00131X_5 \\
 &\quad + 0.00435X_6 + 0.00237X_7 + 0.00366X_8 - 0.00116X_9 + 0.00411X_{10} \\
 &\quad - 0.00063X_{11} - 0.00021X_{12} + 0.00128X_{13} - 0.00922X_{14} + 0.00696X_{15} \\
 &\quad + 0.00338X_{16} + 0.00017X_{17} - 0.009795X_{18} - 0.01403X_{19} - 0.00864X_{20} \\
 &\quad - 0.00408X_{21} - 0.00685X_{22} + 0.01249X_{23} - 0.01267X_{24} + 0.01076X_{25} \\
 &\quad - 0.00376X_{26} + 0.0129X_{27} - 0.022198X_{28} - 0.01272X_{29} - 0.002595X_{30} \\
 &\quad + 0.00159X_{31} - 0.00526X_{32} - 0.00351X_{33} - 0.01547X_{34}
 \end{aligned}$$

Untuk mengevaluasi model dari fungsi diskriminan yang terbentuk, peneliti mengambil contoh 1 orang pasien dari data uji 73 objek pasien. Misalkan klasifikasi dilakukan pada data uji untuk pasien nomor 1, dengan kelompok asli ialah kelompok 2. Adapun untuk nilai fungsi diskriminan Fisher yang sudah di substitusikan dengan data pasien nomor 1 adalah $\hat{y}_1 = -0,0233$, $\hat{y}_2 = -0,0852$, $\hat{y}_3 = -0,1315$, $\hat{y}_4 = 0,0584$, dan $\hat{y}_5 = -0,0555$. Dengan demikian nilai \bar{y}_{kj} untuk $k = 1, 2, \dots, 6$ dan $j = 1, 2, \dots, 5$ adalah sebagai berikut:

$$\bar{y}_{kj} = \begin{bmatrix} 0.0275 & -0.0652 & 0.0271 & -0.0257 & -0.038 \\ -0.0335 & -0.001 & -0.0255 & 0.0653 & 0.0028 \\ -0.1268 & 0.0371 & -0.001 & -0.2002 & 0.0325 \\ -0.0792 & -0.0137 & -0.035 & 0.0942 & -0.0066 \\ 0.0595 & -0.0443 & 0.0031 & 0.1085 & 0.044 \\ 0.4563 & 0.1915 & 0.1143 & 0.171 & 0.027 \end{bmatrix}$$

Dengan menggunakan Persamaan (12) diperoleh hasil berikut:

$$\begin{aligned} \sum_{j=1}^5 (\hat{y}_j - \bar{y}_{1j})^2 &= 0.05773382, \\ \sum_{j=1}^5 (\hat{y}_j - \bar{y}_{2j})^2 &= 0.02221698, \\ \sum_{j=1}^5 (\hat{y}_j - \bar{y}_{3j})^2 &= 0.10467407, \\ \sum_{j=1}^5 (\hat{y}_j - \bar{y}_{4j})^2 &= 0.02589112, \\ \sum_{j=1}^5 (\hat{y}_j - \bar{y}_{5j})^2 &= 0.05415351, \\ \sum_{j=1}^5 (\hat{y}_j - \bar{y}_{6j})^2 &= 0.3212185. \end{aligned}$$

Berdasarkan hasil perhitungan di atas, dapat dilihat bahwa data uji untuk pasien nomor 1 termasuk ke dalam kelompok 2. Hal tersebut sesuai dengan klasifikasi asli. Oleh sebab itu, model fungsi diskriminan hasil pereduksian dengan menggunakan AKU Kernel linear pada metode klasifikasi ADF ini mengklasifikasikan dengan tepat untuk data uji pada pasien nomor 1.

Adapun hasil klasifikasi untuk data uji jenis penyakit erythematous-squamous dengan menggunakan model fungsi diskriminan hasil pereduksian dengan menggunakan AKU Kernel linear pada metode klasifikasi ADF terdapat pada Tabel 7. Dengan demikian, berdasarkan Tabel 7 dapat dilihat bahwa nilai APER untuk metode AKU Kernel linear pada metode klasifikasi ADF untuk data uji diperoleh kesalahan klasifikasi sebesar 0.397260274 dengan 29 objek yang salah klasifikasi dari 73 objek. Hal tersebut dapat dilihat bahwa ketepatan klasifikasi untuk data uji adalah sebesar 0.602739726 dengan 44 objek yang tepat klasifikasi dari 73 objek.

Tabel 7. Hasil Klasifikasi Data Uji dengan menggunakan Pereduksian AKU Kernel Linear pada metode klasifikasi ADF

Kelompok Asli	Kelompok Prediksi						Total
	1	2	3	4	5	6	
1	9	3	0	3	1	0	16
2	1	7	0	7	4	0	19
3	0	1	11	0	1	0	13
4	0	3	0	6	0	0	9
5	0	3	0	2	7	0	12
6	0	0	0	0	0	4	4
Total	10	17	11	18	13	4	73

AKU Kernel Gaussian pada Metode ADF

Pereduksian untuk data dermatologi jenis penyakit erythematous-squamous dengan menggunakan metode AKU Kernel Gaussian pada metode klasifikasi ADF dilakukan dengan membagi data menjadi data latih 80% yaitu sebanyak 293 objek dan data uji 20% yaitu sebanyak 73 objek. Dengan menggunakan algoritma [8] pada perangkat lunak Matlab R2014a, berdasarkan jarak antar objeknya diperoleh interval untuk parameter σ pada data latih adalah $1.4224 \leq \sigma \leq 15.2872$. Interval untuk parameter σ tersebut kemudian dipartisi sebanyak 200 partisi. Dengan menggunakan partisi tersebut menghasilkan parameter σ dan nilai proporsi kumulatif yang berbeda untuk masing-masing KU yang digunakan. Adapun nilai parameter σ dan nilai proporsi kumulatif dapat dilihat pada Tabel 8.

Pada tabel 8, karena belum terdapat patokan baku besar proporsi kumulatif yang dapat digunakan untuk memilih komponen utama terpilih [5] dalam melakukan pereduksian dengan menggunakan AKU Kernel Gaussian pada metode klasifikasi ADF, peneliti melihat proporsi kumulatif antara 75% sampai dengan 95%. Titik pembuat total proporsi kesalahan klasifikasi terkecil untuk proporsi kumulatif antara 75% sampai dengan 95% terdapat pada proporsi

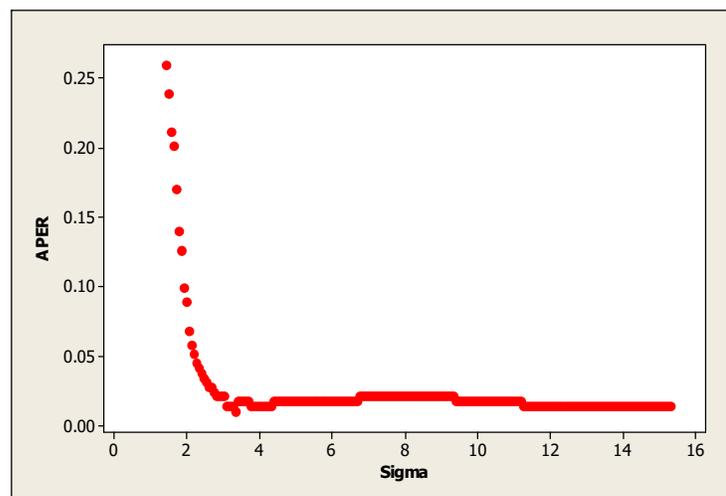
kumulatif 80% dengan $\sigma = 3.3635$. Oleh karena itu, terpilih 19 KU yang akan dilakukan pengklasifikasian dengan menggunakan metode klasifikasi ADF. Tren nilai parameter σ terhadap total proporsi kesalahan klasifikasi untuk 19 KU pada data latih jenis penyakit erythemato-squamous diberikan pada Gambar 1. Adapun hasil klasifikasi untuk 19 komponen utama terpilih pada pereduksian data latih dengan menggunakan metode ADF dengan AKU Kernel Gaussian dapat dilihat pada Tabel 9.

Tabel 8. Nilai Parameter σ dan Nilai Proporsi Kumulatif 75% sampai dengan 95% Data Latih

KU	σ	Proporsi Kumulatif (%)	Kesalahan Klasifikasi
1	6.0671	21.8314	0.2320819
2	2.8782	25.8271	0.1399317
⋮	⋮	⋮	⋮
17	15.2872	90.2518	0.01365188
18	3.9181	80.9217	0.01365188
19	3.3635	79.8992	0.01023891
20	15.2872	93.5905	0.01365188
⋮	⋮	⋮	⋮
32	3.9874	98.1791	0.01023891
33	4.1954	99.1562	0.01023891
34	4.6113	100	0.01023891

Tabel 9. Hasil Klasifikasi AKU Kernel Gaussian pada Metode Klasifikasi ADF untuk Data Latih

Kelompok Asli	Kelompok Prediksi						Total
	1	2	3	4	5	6	
1	96	0	0	0	0	0	96
2	0	40	0	2	0	0	42
3	0	0	59	0	0	0	59
4	0	1	0	39	0	0	40
5	0	0	0	0	40	0	40
6	0	0	0	0	0	16	16
Total	96	41	59	41	40	16	293



Gambar 1. Tren Nilai σ – APER Data Latih Jenis Penyakit Erythemato-Squamous dengan AKU Kernel Gaussian pada Metode Klasifikasi ADF.

Pada Tabel 9 dapat dilihat nilai APER metode AKU Kernel Gaussian pada metode klasifikasi ADF untuk data latih diperoleh kesalahan klasifikasi sebesar 0,010239 dengan 3 objek yang salah klasifikasi dari 293 objek. Hal tersebut dapat dilihat bahwa ketepatan klasifikasi untuk data latih adalah sebesar 0.989761 dengan 290 objek yang tepat klasifikasi dari 293 objek.

Adapun dengan menggunakan *library* MASS pada *software* R i386 3.4.1 [11], diperoleh fungsi diskriminan Fisher untuk 19 komponen utama terpilih. Dengan melakukan transformasi pada persamaan fungsi diskriminan untuk komponen utama terpilih, sehingga diperoleh model persamaan diskriminan Fisher hasil pereduksian AKU Kernel Gaussian pada metode klasifikasi ADF adalah sebagai berikut:

$$\begin{aligned} \hat{Y}_1 &= 0.01745X_1 - 0.19109X_2 + 0.707647X_3 - 0.06791X_4 - 0.0833X_5 \\ &\quad - 0.23854X_6 - 0.18726X_7 - 0.31657X_8 - 0.27425X_9 + 0.278288X_{10} \\ &\quad - 0.04005X_{11} + 0.046276X_{12} - 0.52767X_{13} + 0.91299X_{14} - 0.20666X_{15} \\ &\quad + 0.152651X_{16} - 1.20517X_{17} + 1.396525X_{18} - 0.12353X_{19} + 0.292965X_{20} \\ &\quad - 0.80288X_{21} + 0.215225X_{22} + 0.285392X_{23} - 0.47523X_{24} - 0.84749X_{25} \\ &\quad + 0.249238X_{26} - 0.41716X_{27} + 1.087076X_{28} - 1.36938X_{29} + 1.18688X_{30} \\ &\quad + 0.454044X_{31} + 0.996813X_{32} + 0.327295X_{33} + 1.527486X_{34} \\ \hat{Y}_2 &= 0.139524X_1 - 0.32442X_2 - 0.60443X_3 - 0.44301X_4 - 0.07795X_5 \\ &\quad - 0.45262X_6 + 0.118675X_7 - 0.1557X_8 - 0.10986X_9 + 0.314426X_{10} \\ &\quad - 0.03563X_{11} + 0.41522X_{12} - 1.27527X_{13} - 0.85196X_{14} - 0.39801X_{15} \\ &\quad - 0.02416X_{16} - 0.7217X_{17} - 0.1175X_{18} - 1.15745X_{19} + 0.093412X_{20} \\ &\quad + 0.435648X_{21} - 0.16115X_{22} + 0.117499X_{23} - 0.82241X_{24} + 0.42222X_{25} \\ &\quad + 0.701539X_{26} + 0.396691X_{27} + 0.228967X_{28} + 0.968164X_{29} - 1.1056X_{30} \\ &\quad + 0.200288X_{31} - 0.22546X_{32} - 0.13597X_{33} - 0.24155X_{34} \\ \hat{Y}_3 &= -0.26772X_1 + 0.298204X_2 - 0.14942X_3 - 0.22984X_4 - 0.4867X_5 \\ &\quad - 0.62665X_6 + 0.277091X_7 - 0.18888X_8 + 0.089398X_9 + 0.501597X_{10} \\ &\quad - 0.23855X_{11} - 1.3028X_{12} - 0.56897X_{13} + 0.640852X_{14} - 0.32125X_{15} \\ &\quad + 0.085088X_{16} - 0.11986X_{17} - 0.60904X_{18} - 0.98668X_{19} - 0.00559X_{20} \\ &\quad - 0.13602X_{21} - 0.10995X_{22} + 1.094289X_{23} + 0.503911X_{24} - 0.09999X_{25} \\ &\quad + 0.146812X_{26} - 0.09141X_{27} - 0.48165X_{28} - 0.20337X_{29} + 1.411783X_{30} \\ &\quad - 0.42002X_{31} + 0.018513X_{32} - 0.78569X_{33} + 0.457051X_{34} \\ \hat{Y}_4 &= -0.03168X_1 - 0.12694X_2 - 0.45547X_3 + 0.318922X_4 - 0.18434X_5 \\ &\quad + 0.455144X_6 + 0.485496X_7 + 0.265089X_8 + 0.313697X_9 - 1.15017X_{10} \\ &\quad - 1.02442X_{11} - 0.02528X_{12} + 0.45288X_{13} - 0.7703X_{14} + 0.262742X_{15} \\ &\quad + 0.777817X_{16} - 1.05733X_{17} - 0.91705X_{18} - 1.421X_{19} + 0.331876X_{20} \\ &\quad + 0.321487X_{21} + 1.273112X_{22} - 1.88156X_{23} + 0.244415X_{24} + 1.507888X_{25} \\ &\quad - 0.83123X_{26} - 0.66161X_{27} + 2.356311X_{28} + 0.409757X_{29} - 1.13649X_{30} \\ &\quad + 0.097859X_{31} - 1.50728X_{32} + 2.057696X_{33} - 0.75247X_{34} \\ \hat{Y}_5 &= 0.09636X_1 + 0.132927X_2 - 0.02147X_3 + 0.181931X_4 - 0.05076X_5 \\ &\quad + 0.224148X_6 + 0.348565X_7 - 0.07023X_8 + 0.458377X_9 - 0.43436X_{10} \\ &\quad + 0.523037X_{11} - 0.05291X_{12} - 0.1157X_{13} - 0.71793X_{14} - 0.22699X_{15} \\ &\quad - 0.72306X_{16} + 1.164027X_{17} - 0.09189X_{18} - 0.26531X_{19} + 0.360003X_{20} \\ &\quad - 0.01283X_{21} - 0.18007X_{22} - 0.09311X_{23} - 0.29984X_{24} - 1.08791X_{25} \\ &\quad + 0.015362X_{26} - 0.35349X_{27} + 0.233149X_{28} - 0.45439X_{29} + 1.371978X_{30} \\ &\quad + 0.012649X_{31} + 0.488722X_{32} + 0.087497X_{33} - 0.5398X_{34} \end{aligned}$$

Untuk mengevaluasi model dari fungsi diskriminan yang terbentuk, peneliti mengambil contoh 1 orang pasien dari data uji 73 objek pasien. Misalkan klasifikasi dilakukan pada data uji untuk pasien nomor 1, dengan kelompok asli ialah kelompok 2. Adapun untuk nilai fungsi diskriminan Fisher yang sudah di substitusikan dengan data pasien nomor 1 adalah $\hat{y}_1 =$

7.7903, $\hat{y}_2 = -1.4662$, $\hat{y}_3 = 3.4971$, $\hat{y}_4 = -5.1372$ dan $\hat{y}_5 = -2.0469$. Dengan demikian nilai \bar{y}_{kj} untuk $k = 1, 2, \dots, 6$ dan $j = 1, 2, \dots, 5$ adalah sebagai berikut:

$$\bar{y}_{kj} = \begin{bmatrix} 1.0081 & -1.0754 & 3.3682 & -8.2989 & 0.6811 \\ 1.8198 & -0.5268 & 1.7122 & 3.0185 & -1.0607 \\ -4.2327 & 2.8465 & -7.3273 & 9.1313 & -3.4201 \\ 2.5189 & 1.6168 & 0.6591 & 2.3812 & -0.3392 \\ -2.1618 & 1.0315 & 0.616 & 0.0468 & 1.5278 \\ 4.4166 & 0.0885 & 3.1192 & -0.7847 & 6.5777 \end{bmatrix}$$

Dengan menggunakan persamaan (12) diperoleh hasil sebagai berikut:

$$\sum_{j=1}^5 (\hat{y}_j - \bar{y}_{1j})^2 = 63.6059$$

$$\sum_{j=1}^5 (\hat{y}_j - \bar{y}_{2j})^2 = 107.203$$

$$\sum_{j=1}^5 (\hat{y}_j - \bar{y}_{3j})^2 = 485.795$$

$$\sum_{j=1}^5 (\hat{y}_j - \bar{y}_{4j})^2 = 104.789$$

$$\sum_{j=1}^5 (\hat{y}_j - \bar{y}_{5j})^2 = 153.236$$

$$\sum_{j=1}^5 (\hat{y}_j - \bar{y}_{6j})^2 = 107.270$$

Dengan melihat hasil perhitungan diatas, sehingga dapat dilihat bahwa data uji untuk pasien nomor 1 termasuk ke dalam kelompok 1. Hal tersebut tidak sesuai dengan klasifikasi asli. Oleh sebab itu, model fungsi diskriminan hasil pereduksian dengan menggunakan AKU Kernel linear pada metode klasifikasi ADF ini mengklasifikasikan tidak tepat untuk data uji pada pasien nomer 1. Adapun hasil klasifikasi untuk data uji jenis penyakit erythemato-squamous dengan menggunakan model fungsi diskriminan hasil pereduksian dengan menggunakan metode ADF dengan AKU Kernel Gaussian dapat dilihat pada Tabel 10. Pada Tabel 10 dapat dilihat bahwa nilai APER untuk metode AKU Kernel Gaussian pada metode klasifikasi ADF untuk data uji diperoleh kesalahan klasifikasi sebesar 0.178082 dengan 13 objek yang salah klasifikasi dari 73 objek. Hal tersebut dapat dilihat bahwa ketepatan klasifikasi untuk data uji adalah sebesar 0.821918 dengan 60 objek yang tepat klasifikasi dari 73 objek.

Tabel 10. Hasil Klasifikasi Data Uji dengan menggunakan Pereduksian AKU Kernel Gaussian pada Metode Klasifikasi ADF

Kelompok Asli	Kelompok Prediksi						Total
	1	2	3	4	5	6	
1	13	1	0	0	2	0	16
2	1	12	0	4	2	0	19
3	0	0	13	0	0	0	13
4	0	2	0	7	0	0	9
5	0	0	0	0	12	0	12
6	1	0	0	0	0	3	4
Total	15	15	13	11	16	3	73

KESIMPULAN

Penelitian ini telah berhasil mengklasifikasikan jenis penyakit erythemato-squamous dengan menggunakan metode ADF dengan pereduksian menggunakan AKU Kernel linear dan AKU Kernel Gaussian. Berdasarkan proporsi kumulatif, terdapat 18 komponen utama terpilih pada AKU Kernel linier untuk membentuk fungsi diskriminan Fisher dan terdapat 19 komponen utama terpilih pada AKU Kernel Gaussian untuk membentuk fungsi diskriminan Fisher. Metode ADF dengan AKU Kernel linear berhasil mengklasifikasikan data latih dengan ketepatan 98,63% dan berhasil mengklasifikasikan data uji dengan ketepatan 60,27%.

Sedangkan Metode ADF dengan AKU Kernel Gaussian berhasil mengklasifikasikan data latih dengan ketepatan 98,98% dan berhasil mengklasifikasikan data uji dengan ketepatan 82,19%. AKU Kernel Gaussian berhasil mengklasifikasikan dengan tepat 98,98% data latih dan 82,19% data uji. Oleh karena itu, untuk jenis penyakit erythemat-squamous klasifikasi lebih tepat menggunakan ADF dengan AKU Kernel Gaussian dibandingkan dengan AKU Kernel linier. Oleh karena itu, dapat disimpulkan bahwa klasifikasi data dermatologi jenis penyakit erythemat-squamous lebih tepat menggunakan metode ADF dengan AKU Kernel Gaussian dibandingkan dengan AKU Kernel linear. Selain itu, fungsi diskriminan yang telah diperoleh dari metode klasifikasi ADF dengan pereduksian menggunakan AKU Kernel Gaussian ini dapat dijadikan alternatif untuk mendiagnosa jenis penyakit erythemat-squamous.

REFERENSI

- [1] Bishop, C. M., 2006, *Pattern Recognition and Machine Learning*, Cambridge: Springer.
- [2] Djakaria, I., Guritno, S., dan Kartiko, S.H., 2010, Visualisasi Data Iris menggunakan Analisis Komponen Utama dan Analisis Komponen Utama Kernel, *Jurnal Ilmu Dasar*. Vol. 11, No. 1, hal. 31 – 38.
- [3] Izenman, A. J and Shen, Y., 2007, *Outlier Detection Using the Smallest Kernel Principal Components*, Philadelphia (US): Temple University.
- [4] Johnson, R. A. dan Wichern, D.W., 2007, *Applied Multivariate Statistical Analysis Sixth Edition*. New Jersey (US): Pearson Prentice Hall.
- [5] Jolliffe, I. T., 2002, *Principal Component Analysis, Second Edition*, New York: Springer-Verlag.
- [6] Kustiyo, A. dan Tjandrasa, H., 2004, Model Neural Network dengan Inisialisasi Pembobot Awal menggunakan Regresi Logistik Biner untuk Memprediksi Jenis Penyakit Erythemat-Squamous, *Jurnal Ilmiah-Ilmu Komputer*. Vol. 2 No.2, hal. 14-25.
- [7] Kustiyo, A., Hermadi, I. dan Fajriyati, M., 2007, Prediksi Jenis Penyakit Erythemat-Squamous menggunakan Jaringan Syaraf Tiruan dengan K-Means sebagai Praproses, *Jurnal Ilmu Matematika dan Terapan*, Vol. 1, No.1, hal. 51-59. ISSN: 1978-7227.
- [8] Lobo, S. A., 2015, *Visualisasi Data Berkelompok dengan Analisis Komponen Utama Kernel*, Tesis, Sekolah Pascasarjana Institut Pertanian Bogor, Indonesia.
- [9] Merz, C. and 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.
- [10] Rathi, Y., Dambreville S. dan Tannenbaum A., 2006, Statistical Shape Analysis Using Kernel PCA, *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*, SPIE-IS&T/ Vol. 6064 60641B-1.
- [11] Ripley, B., Venables B. and Hornik K., 2018, *Support Functions and Datasets for Venables and Ripley's MASS R Package Version 7.3-23* [<https://cran.r-project.org/web/packages/MASS/MASS.pdf>]. Repository CRAN.
- [12] Schölkopf, B., Smola, A. dan Muller, K. R., 1998, Nonlinear Component Analysis as a Kernel Eigenvalue Problem, *Neural Computation*, 10(5): pp.1299-1399.
- [13] Taylor, J. S. dan Cristianini, N., 2004, *Kernel Methods for Pattern Analysis*. New York: Cambridge University Press.
- [14] Widjaja, D., Varon, C., Dorado, A. C., Suykens, J. A. K., dan Van Huffel, S., 2012. Application of Kernel Principal Component Analysis for Single Lead ECG-Derived Respiration, *IEEE Transactions on Biomedical Engineering*, Vol 59, No. 4, April 2012: 1169-76. doi: 10.1109/TBME.2012.2186448.