

## REDUKSI DIMENSI INPUT PADA JARINGAN SYARAF PCA-RBF DENGAN SINGULAR VALUE DECOMPOSITION

**Abdul Hakim Maulana, Oni Soesanto, Thresye**

Program Studi Matematika  
Fakultas MIPA Universitas Lambung Mangkurat  
Email: [abdulhakimmaulana2691@gmail.com](mailto:abdulhakimmaulana2691@gmail.com)

### ABSTRAK

Jaringan syaraf tiruan adalah suatu sistem pemroses informasi yang memiliki karakteristik mirip dengan jaringan syaraf biologi. Jaringan syaraf tiruan dibedakan menjadi *single layer* dan *multiple layer*. Salah satu jaringan syaraf *multiple layer* adalah *Radial Basis Function* (RBF). RBF dikenal memiliki kecepatan komputasi yang tinggi. Namun, kinerja RBF berkurang ketika berkenaan ruang *input* dengan dimensi yang tinggi sehingga diperlukan penyederhanaan jaringan. Salah satu metode penyederhanaan RBF berkenaan dimensi ruang *input* adalah menggunakan *Principal Component Analysis* (PCA). Ketika jumlah variabel data lebih besar daripada jumlah pengamatan, kemampuan PCA menjadi kurang efektif maka diperlukan *Singular Value Decomposition* (SVD) untuk menyelesaikan masalah tersebut. Tujuan penelitian ini yakni untuk menerapkan proses *Singular Value Decomposition* (SVD) pada jaringan syaraf PCA-RBF. Penelitian ini membahas mengenai jaringan syaraf PCA-RBF. PCA berfungsi untuk mereduksi dimensi input pada RBF. Hasil reduksi dimensi input ini dikenal sebagai *principal component* (PC). Proses penentuan PC dilakukan menggunakan metode PCA yang dikombinasikan dengan SVD. Selanjutnya, PC digunakan sebagai input baru pada RBF dan dilakukan proses klustering pada PC dengan metode K-means untuk inisialisasi *center* RBF. Inisialisasi *center* merupakan langkah awal RBF dalam melakukan klasifikasi. Proses klasifikasi pada RBF terdiri dari dua proses yakni *training* dan *testing*. Hasil dari penelitian ini yaitu proses SVD pada PCA untuk mereduksi dimensi data input yang terdiri dari proses penentuan matriks singular kanan ( $V$ ) yakni menghitung matriks  $A^T A$ , mencari nilai eigen ( $\lambda$ ) dan vektor eigen dari matriks  $A^T A$ , melakukan Proses Gram-Schmidt dan normalisasi, serta proses pembentukan *Principal Component* (PC) yaitu dengan cara mengalikan matriks data *training* dengan matriks singular kanan ( $V$ ), sehingga PC digunakan sebagai input baru pada RBF. Pada penelitian ini diberikan contoh data klasifikasi yakni data *satelit Landsat*. Setelah dilakukan perulangan sebanyak 30 kali didapat rata-rata keberhasilan klasifikasi pada data training Landsat sebesar 79,889% dengan rata-rata errornya sebesar 20,111 %, sedangkan untuk data testing Landsat diperoleh rata-rata keberhasilan sebesar 93,333 % dengan persentase errornya sejumlah 6,667 %.

**Kata kunci:** Jaringan Syaraf Tiruan (JST), Jaringan Syaraf *Radial Basis Function* (RBF), *Principal Component Analysis* (PCA), *Singular Value Decomposition* (SVD)

### 1. PENDAHULUAN

Jaringan syaraf tiruan adalah suatu sistem pemrosesan informasi yang memiliki karakteristik mirip dengan jaringan syaraf biologi pada manusia. Berdasarkan arsitekturnya, jaringan syaraf tiruan dibedakan menjadi dua yaitu *single layer* dan *multiple layer*. Salah satu jaringan syaraf tiruan *multiple layer* adalah *Radial Basis Function* (RBF).

Jaringan saraf RBF sering digunakan untuk masalah klasifikasi. RBF dikenal memiliki kecepatan komputasi yang tinggi [1]. Kinerja RBF berkurang ketika berkenaan dengan ruang *input* berdimensi yang tinggi sehingga diperlukan penyederhanaan jaringan [4]. Salah satu penyederhanaan jaringan pada *Radial Basis Function* adalah *Principal Component Analysis* (PCA).

*Principal Component Analysis* merupakan metode dalam statistika yang digunakan untuk mereduksi dimensi ruang data. Terdapat dua proses penyelesaian pada PCA klasik yaitu dengan dekomposisi nilai *eigen* dari matriks kovarian atau dengan menguraikan matriks data asli menggunakan *Singular Value Decomposition* (SVD) [3]. PCA klasik tidak berguna secara penuh ketika jumlah variabel lebih besar daripada jumlah pengamatan, sehingga SVD diterapkan untuk menyelesaikan masalah penyederhanaan jaringan pada PCA-RBF [3].

*Singular Value Decomposition* (SVD) diperkenalkan dan dikembangkan oleh lima matematikawan yaitu E. Beltrami, C. Jordan, J.J Sylvester, E. Schmidt, dan H.Weyl [5]. SVD adalah suatu metode faktorisasi khusus untuk matriks berukuran  $m \times n$ , dengan  $m$  adalah jumlah baris dan  $n$  adalah jumlah kolom. Misalkan  $\mathbf{A}$  adalah matriks berukuran  $m \times n$ , maka sebuah faktorisasi dapat berbentuk  $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$  [2]. Metode SVD memiliki tiga buah komponen matriks yang menjadi penyusun utamanya yaitu matriks  $\mathbf{U}$ , matriks  $\mathbf{V}$  dan matriks  $\mathbf{\Sigma}$ . Matriks  $\mathbf{U}$  adalah matriks yang terbentuk dari vektor eigen matriks  $\mathbf{A}\mathbf{A}^T$ , sehingga disebut matriks singular kiri sedangkan matriks  $\mathbf{V}$  adalah matriks yang terbentuk dari vektor eigen matriks  $\mathbf{A}^T\mathbf{A}$  sehingga disebut matriks singular kanan. Penentuan matriks  $\mathbf{U}$  dan matriks  $\mathbf{V}$  pada SVD masing-masing disusun berdasarkan nilai eigen yang bersesuaian. Selanjutnya adalah matriks  $\mathbf{\Sigma}$ . Matriks  $\mathbf{\Sigma} \in \mathbf{R}^{m \times n}$  adalah matriks diagonal yang mempunyai bentuk  $\mathbf{\Sigma} = \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ , dimana  $\mathbf{D}$  adalah matriks diagonal  $r \times r$  untuk  $r$  lebih kecil dari  $m$  dan  $n$ . Diagonal  $\mathbf{D}$  berisikan nilai-nilai *singular* dari  $\mathbf{A}$ . Nilai *singular* dari  $\mathbf{A}$  merupakan akar kuadrat dari nilai *eigen*  $\mathbf{A}^T\mathbf{A}$  yang dilambangkan dengan  $\sigma_1, \dots, \sigma_n$ , dan disusun dalam urutan menurun sesuai ketentuan  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq 0$ , dimana  $r = \min(m, n)$ . Dengan kata lain,  $\sigma_i = \sqrt{\lambda_i}$  untuk  $1 \leq i \leq n$  [2].

Metode ini akan disimulasikan pada data klasifikasi dengan karakteristiknya *multivariate*. Data yang digunakan pada penelitian ini adalah data *satelit landsat* yang merupakan hasil pencitraan potret daratan dari satelit. Dalam penelitian ini *Singular Value Decomposition* (SVD) digunakan untuk mereduksi dimensi input pada jaringan syaraf PCA-RBF.

## 2. METODOLOGI PENELITIAN

Proses SVD pada PCA untuk mereduksi dimensi data input RBF dilakukan dalam dua tahapan. Tahap pertama dimulai dengan menentukan matriks *singular* kanan ( $\mathbf{V}$ ) dari  $\mathbf{A}$ , kemudian dilanjutkan pada tahap kedua yakni membentuk *Principal Component* (PC). Menentukan matriks singular kanan ( $\mathbf{V}$ ) dapat dilakukan dengan langkah sebagai berikut:

- 1) Menghitung matriks  $\mathbf{A}^T\mathbf{A}$
- 2) Mencari nilai eigen ( $\lambda$ ) dan vektor eigen dari matriks  $\mathbf{A}^T\mathbf{A}$
- 3) Melakukan Proses *Gram-Schmidt* dan normalisasi

Setelah matriks *singular* kanan ( $\mathbf{V}$ ) didapatkan maka dilanjutkan ke tahap berikutnya yaitu membentuk *Principal Component* (PC) yang nantinya akan digunakan sebagai input pada RBF. PC yang terbentuk merupakan kombinasi linier dari matriks  $\mathbf{A}$  dengan vektor-vektor *singular* kanan yang termuat dalam  $\mathbf{V}$ . PC dinyatakan dengan  $K_1, K_2, \dots, K_p$  dan  $\mathbf{V} = [v_1, v_2, \dots, v_n]$  maka:

$$K_1 = \mathbf{A}\mathbf{v}_1 = \mathbf{A}v_{11} + \mathbf{A}v_{21} + \dots + \mathbf{A}v_{n1}$$
$$K_2 = \mathbf{A}\mathbf{v}_2 = \mathbf{A}v_{12} + \mathbf{A}v_{22} + \dots + \mathbf{A}v_{n2}$$

⋮

$$K_p = \mathbf{A}\mathbf{v}_p = \mathbf{A}v_{1p} + \mathbf{A}v_{2p} + \dots + \mathbf{A}v_{np}$$

Sehingga  $K$  dapat dinyatakan:

$$K_i = \mathbf{A} \cdot \mathbf{v}_i \quad \text{dengan } i = 1, 2, \dots, p$$

Proses reduksi pada PCA dilakukan dengan memilih PC berdasarkan total persentase kumulatif yang dihasilkan. PC yang dipilih inilah yang merupakan hasil reduksi dan dapat dinyatakan dalam matriks  $\mathbf{K}_{\text{PCA}}$ , yaitu:

$$\mathbf{K}_{\text{PCA}} = [K_1, K_2, \dots, K_p]$$

dengan  $p < n$  dimana  $p$  adalah banyak PC yang dipilih dan  $n$  adalah banyaknya variabel pada data.  $\mathbf{K}_{\text{PCA}}$  inilah yang akan digunakan sebagai input pada RBF, baik pada tahap training maupun tahap testing.

Setelah memperoleh PC dari proses SVD dan PCA, selanjutnya melakukan klasifikasi menggunakan metode RBF. Penyelesaian klasifikasi pada RBF terdiri dari dua proses yaitu *training* dan *testing*.

Proses training RBF meliputi langkah awal yakni inialisasi *center* dengan K-means, selanjutnya menghitung nilai *spread* ( $\ddot{\sigma}$ ) menggunakan persamaan (1) sebagai berikut:

$$\ddot{\sigma} = \frac{d_{\max}}{\sqrt{K}} \quad \dots(1)$$

dengan  $d_{\max}$  adalah jarak Euclid maksimal dari data *training* dan  $K$  adalah jumlah total data *training*. Nilai *spread* ( $\ddot{\sigma}$ ) dan bobot *center* yang telah diketahui kemudian digunakan untuk menghitung matriks *Gaussian* dengan menggunakan persamaan (2) sebagai berikut:

$$\phi_{ij}(\|x_i - c_j\|) = \exp\left(-\frac{\|x_i - c_j\|^2}{2\ddot{\sigma}^2}\right) \quad i = 1, 2, \dots, p; \quad j = 1, 2, \dots, m \quad \dots(2)$$

Data dari *input layer* akan diteruskan ke *hidden layer* yang berupa matriks *Gaussian* atau matriks  $\boldsymbol{\phi}(\mathbf{r})$  dimana  $r = \|x_i - c_j\|$ . Setelah didapatkan matriks *Gaussian*, selanjutnya menghitung nilai bobot output  $\mathbf{W}$  menggunakan persamaan (3) sebagai berikut:

$$\mathbf{W} = \boldsymbol{\phi}(\mathbf{r})^+ \mathbf{t} \quad \dots(3)$$

Tahapan terakhir pada training RBF adalah menghitung nilai output menggunakan persamaan (4) sebagai berikut:

$$y_{ij} = \sum_{i=1}^n \phi_{ij}(\|x_i - c_j\|) w_{ij} \quad \dots(4)$$

Proses pada testing RBF yakni meliputi langkah menghitung matriks *Gaussian* dengan menggunakan persamaan (2)

$$\phi_{ij}(\|x_i - c_j\|) = \exp\left(-\frac{\|x_i - c_j\|^2}{2\ddot{\sigma}^2}\right)$$

Setelah diperoleh matriks *Gaussian* untuk data *testing*, selanjutnya dilakukan perhitungan nilai output untuk data *testing* menggunakan persamaan (4):

$$y_{ij} = \sum_{i=1}^n \phi_{ij}(\|x_i - c_j\|) w_{ij}$$

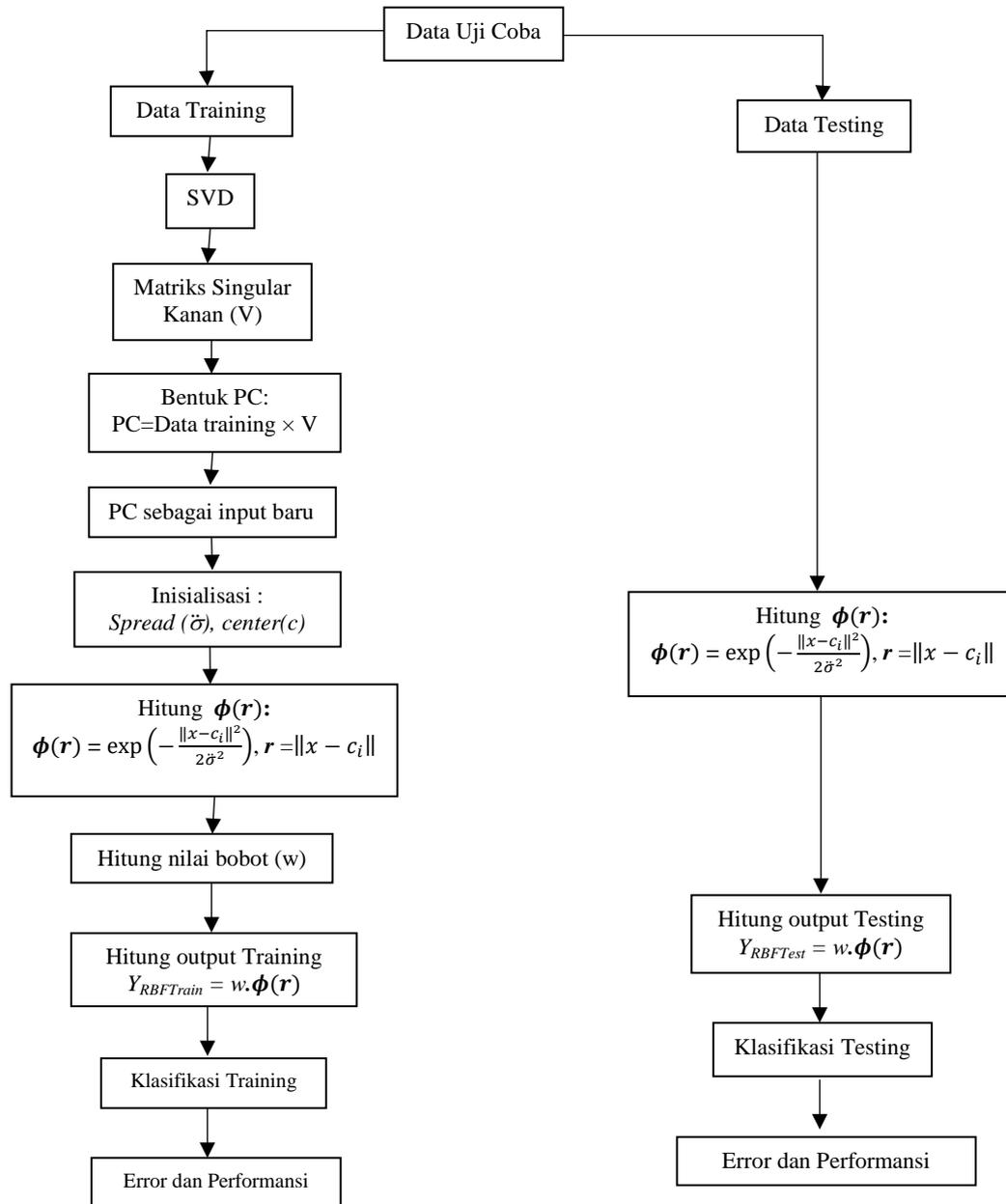
Pada penelitian ini diberikan data klasifikasi yang disimulasikan menggunakan metode PCA-RBF yang dikombinasikan dengan SVD. Data di akses pada fasilitas unduh website *Machine Learning Repository* (<http://archive.ics.uci.edu/ml/index.html>). Data yang digunakan pada penelitian ini

adalah data *satelit landsat*. Data *satelit landsat* merupakan data hasil pencitraan potret daratan dari satelit. Satu *frame* citra *Landsat* terdiri dari empat gambar digital yang dilakukan pengambilan potret secara bersamaan pada *spectral bands* yang berbeda. Citra yang dihasilkan *satelit Landsat* berupa citra *grayscale* berukuran  $3 \times 3$  piksel dengan format 8-bit *binary word* pada setiap piksel nya. Setiap baris pada data *satelit Landsat* mengandung nilai-nilai piksel dalam empat *spectral band* dari masing-masing  $(3 \times 3)$  piksel citra yang dihasilkan. Nilai piksel tersebut dikonversi ke dalam derajat keabuan pada citra *grayscale*.

Data yang tersedia yakni sebanyak 4.435 untuk data *training* dan sebanyak 2.000 data untuk data *testing*. Jumlah atribut/variabel untuk data satelit Landsat adalah 36 buah ( 4 buah *spectral band*  $\times$   $(3 \times 3)$  piksel ukuran citra). Jumlah kelas yang akan digunakan yakni 6 kelas : tanah merah, tanaman kapas, tanah abu-abu, tanah abu-abu basah, tanah dengan vegetasi jerami, dan tanah abu-abu sangat basah.

Dalam penelitian ini data pengukuran yang diambil terdiri dari 36 variabel. Pada tahap *training*, diambil 30 data pengamatan dari 4.435 data *training* yang tersedia, sedangkan untuk tahap *testing* diambil 30 data pengamatan dari 2.000 data *testing* yang tersedia.

Gambar 1 menunjukkan proses jaringan syaraf PCA-RBF yang dikombinasikan dengan SVD.



Gambar 1. Proses learning PCA-RBF dengan SVD

### 3. HASIL DAN PEMBAHASAN

*Center* merupakan pusat dari data. Setiap kelas pada data *training* akan dicari nilai bobot *centernya*. Pada penelitian ini, inisialisasi bobot *center* awal dilakukan menggunakan metode *K-means clustering*.

#### 3.1 Inisialisasi center

Langkah awal pada proses RBF adalah penentuan *center*. Secara umum *K-Mean* klustering merupakan metode untuk mengelompokkan *n* data berdasarkan

atributnya menjadi  $k$  buah grup, dimana  $k < n$ . Langkah-langkah dalam algoritma *K-Means* klustering adalah :

1. Menentukan jumlah kelas,  $k$  dan nilai *center*. Dalam menentukan nilai *center* untuk awal iterasi dilakukan secara acak. Sedangkan jika menentukan nilai *center* yang merupakan tahap dari iterasi, maka digunakan rumus sebagai berikut :

$$c_{ij} = \frac{1}{N_i} \sum_{k=1}^{N_i} x_{kj}$$

dengan  $c_{ij}$  adalah titik *center* dari kelas ke- $i$  untuk variabel ke- $j$ ,  $N_i$  adalah jumlah data yang menjadi anggota kelas ke- $i$  dan  $x_{kj}$  adalah nilai data ke- $k$  yang ada di dalam kelas tersebut untuk variabel ke- $j$ .

2. Menghitung jarak antara titik tiap data dengan titik *center* menggunakan jarak *Euclidean*

Pengelompokan data berdasarkan jarak *Euclidean* yang minimum antara data dengan *center* awal. Proses diatas lakukan perulangan hingga nilai *center* yang dihasilkan tetap dan anggota kelas tidak berpindah ke kelas yang lain.

### 3.2 Simulasi pada Data

Pada penelitian ini diberikan data klasifikasi yang disimulasikan menggunakan metode PCA-RBF dengan SVD yaitu data *satelit Landsat*. Simulasi dilakukan sebanyak 30 kali perulangan pada proses *training* dan juga *testing*. Hasil perulangan klasifikasi diberikan pada tabel 1. Pada tabel tersebut terlihat pergerakan nilai error klasifikasi tidak terlalu signifikan, baik pada tahap *training* maupun pada tahap *testing*.

**Tabel 1.** Hasil klasifikasi menggunakan PCA-RBF dengan SVD

Perulangan	PCA-RBF dengan SVD					
	Training			Testing		
	Error(%)	Perfom (%)	Cpu time (detik)	Error (%)	Performansi (%)	Cpu time (detik)
1	20.0000	80.0000	0.5000	3.3333	96.6667	0.2344
2	20.0000	80.0000	0.0625	6.6667	93.3333	0.2031
⋮	⋮	⋮	⋮	⋮	⋮	⋮
29	20.0000	80.0000	0.0313	6.6667	93.3333	0.2656
30	20.0000	80.0000	0.0781	6.6667	93.3333	0.2813
<b>Rataan</b>	<b>20.111</b>	<b>79.889</b>	<b>0.049</b>	<b>6.667</b>	<b>93.333</b>	<b>0.243</b>
<b>Maks</b>	<b>23.3333</b>	<b>83.3333</b>	<b>0.5000</b>	<b>10.0000</b>	<b>96.6667</b>	<b>0.3750</b>
<b>Min</b>	<b>16.6667</b>	<b>76.6667</b>	<b>0.0156</b>	<b>3.3333</b>	<b>90.0000</b>	<b>0.2031</b>

Dari tabel di atas terlihat bahwa rata-rata error klasifikasi menggunakan PCA-RBF dengan SVD pada 30 kali perulangan adalah sebesar 20,111 % untuk data *training* dan 6,667 % untuk data *testing*. Sedangkan error maksimum terletak pada 23,3333 % untuk data *training* dan 10 % untuk data *testing*. Kemudian error minimumnya terletak pada 16,6667 % untuk data *training* serta 3,333 % untuk data *testing*.

Kemudian hasil performansi, diperoleh rata-rata sebesar 79,889% untuk data *training* dan 93,333 % untuk data *testing*. Performansi maksimum terletak pada 83,3333% untuk data *training* dan 96,6667% untuk data *testing*. Sedangkan performansi minimumnya terlihat pada tabel, sebesar 76,6667% untuk data *training* serta 90 % untuk data *testing*.

#### 4. KESIMPULAN

Berdasarkan hasil penelitian tentang *Singular Value Decomposition* (SVD) Pada Jaringan Syaraf PCA-RBF diperoleh kesimpulan sebagai berikut :

1. Proses SVD pada PCA untuk mereduksi dimensi data input RBF dilakukan dalam dua tahapan. Tahap pertama dimulai dengan menentukan matriks *singular* kanan ( $\mathbf{V}$ ) dari  $\mathbf{A}$ , kemudian dilanjutkan pada tahap kedua yakni membentuk *Principal Component* (PC). PC merupakan hasil reduksi SVD dan PCA terhadap data asli. PC inilah yang digunakan sebagai input baru pada RBF dalam proses klasifikasi. Penyelesaian klasifikasi pada RBF terdiri dari dua proses yaitu *training* dan *testing*. Proses training RBF meliputi: inialisasi *center* dengan K-means, perhitungan nilai *spread*, perhitungan nilai matriks *Gaussian*  $\phi(\mathbf{r})$  data *training*, penentuan bobot *output*, perhitungan nilai *output training* serta melakukan klasifikasi. Sedangkan proses *testing* RBF meliputi: perhitungan matriks *Gaussian*  $\phi(\mathbf{r})$  data *testing*, perhitungan nilai *output testing* serta mengklasifikasikannya.
2. Performansi/kemampuan metode PCA-RBF dengan SVD untuk klasifikasi pada data *Landsat satelit* setelah dilakukan percobaan sebanyak 30 kali, pada data *training* diperoleh rata-rata keberhasilan sebesar 79,889% dan rata-rata error sebesar 20,111 %, sedangkan pada data *testing* diperoleh rata-rata keberhasilan sebesar 93,333 % dan rata-rata error sebesar 6,667 %.

#### DAFTAR PUSTAKA

- [1] Gupta, M.M., L. Jin, & N. Homma. 2003. *Static and Dinamic Neural Networks : From Fundamentals to Advanced Theory*. John Willey & Sons, Inc., New York.
- [2] Lay, D.C. 2006. *Linear Algebra and Its Applications Third Edition Update*. University of Maryland, College Park.
- [3] Lim, D.H. 2013. Principal Component Analysis using Singular Value Decomposition of Microarray Data. *International Journal of Mathematical, Computational, Physical and Quantum Engineering*. Vol:7 No:9.
- [4] Lu, W.Z., W.J. Wang, X.K. Wang, S.H. Yan, & J.C. Lam. 2004. Potential assessment of a neural network model with PCA/RBF approach for forecasting pollutant trends in Mong Kok urban air, Hong Kong. *Environmental Research*. 96 (2004) 79–87.
- [5] Stewart, G.W. 1993. On The Early History of the Singular Value Decomposition. *Society for Industrial and Applied Mathematics*. Vol: 35, No. 4, pp. 551-566.