

# Teknik Reduksi Dimensi Menggunakan Komponen Utama Data Partisi Pada Pengklasifikasian Data Berdimensi Tinggi dengan Ukuran Sampel Kecil

Ronny Susetyoko, Elly Purwantini  
Politeknik Elektronika Negeri Surabaya  
rony@eepis-its.edu, elly@eepis-its.edu

## Abstrak

Dalam pengklasifikasian pola, jumlah data pembelajaran yang digunakan seringkali sangat terbatas, namun jumlah dimensi (variabel) sangat tinggi. Pada penelitian ini digunakan analisis komponen utama (Principal Component Analysis) untuk mereduksi dimensi. Dataset yang berdimensi tinggi dipartisi menjadi beberapa himpunan bagian (subset). Variabel baru dibentuk berdasarkan komponen utama pertama dari setiap subset. Hasil penelitian ini adalah klasifikasi variabel baru konsisten dengan data asli apabila nilai rata-rata total variasi berkisar antara 81,48% - 99,99%. Teknik ini sangat signifikan mengurangi waktu pembelajaran. Secara umum, performa hasil klasifikasi data learning sangat tinggi untuk beberapa kasus dan beberapa metode klasifikasi. Sedangkan hasil klasifikasi data non-learning berkisar 25,00% - 66,67%. Hal ini disebabkan oleh keragaman nilai koefisien variasi dataset yang sangat besar.

**Kata Kunci :** reduksi dimensi, komponen utama, total variasi, persamaan regresi

## 1. Pendahuluan

Klasifikasi pola merupakan permasalahan yang banyak digunakan dalam berbagai aplikasi. Dalam pengolahan citra maupun pengolahan sinyal seringkali pengelompokan maupun pengklasifikasian dilakukan hanya didahului dengan transformasi data saja tanpa melakukan ekstraksi fitur. Selain itu, sampel sebagai data pembelajaran yang digunakan seringkali sangat terbatas.

Menurut Borries [3], algoritma pengelompokan yang berkembang sekarang ini mempunyai performa yang relatif rendah ketika diaplikasikan pada kasus data berdimensi tinggi dengan ukuran sampel kecil. Performa yang rendah ini diindikasikan pada stabilitas hasil dan akurasi yang rendah pada data non-normal. Dalam

penelitiannya, Borries menggunakan statistik non parametrik yaitu uji rank (*rank test*) untuk mengetahui homogenitas distribusi dalam sebuah kelompok dari setiap variabel. Qiau dkk. [1], menggunakan Fisher's Linear Discriminant Analysis (FLDA) untuk mereduksi dimensi (ekstraksi fitur). Penelitian tersebut mengembangkan metode pemilihan variabel 'incorporating' secara otomatis dalam FLDA. FLDA dan nilai eigen digunakan sebagai regulator untuk memperoleh vektor-vektor *sparse linear discriminant* yaitu vektor-vektor diskriminan yang hanya mempunyai sejumlah kecil elemen-elemen tidak nol. Metode ini efektif untuk kasus data berdimensi tinggi dengan ukuran sampel kecil. Sedangkan Hoyle [4], menggunakan pemilihan dimensi secara otomatis yang didasarkan pada aproksimasi Laplace untuk pembuktian model yang bertujuan menentukan jumlah sinyal komponen utama dalam dataset.

Metode yang dikembangkan dalam penelitian ini digunakan untuk pengklasifikasian data berdimensi yang tinggi dengan ukuran sampel yang sangat kecil. Analisis komponen utama (*Principal Component Analysis / PCA*) digunakan untuk mereduksi dimensi. Dataset yang berdimensi tinggi dipartisi menjadi beberapa himpunan bagian (subset). Variabel baru yang terbentuk berdasarkan komponen utama pertama dari setiap subset digunakan untuk pengklasifikasian.

## 2. Analisis Komponen Utama

Analisis komponen utama (*PCA*) adalah teknik yang digunakan untuk menyederhanakan suatu data, dengan cara mentransformasi linier sehingga terbentuk sistem koordinat baru dengan variansi maksimum. PCA dapat digunakan untuk mereduksi dimensi suatu data tanpa mengurangi karakteristik data tersebut secara signifikan. Metode ini mengubah dari sebagian besar variabel asli yang saling berkorelasi menjadi satu himpunan variabel baru yang lebih kecil dan saling bebas (tidak berkorelasi lagi). [6]

Komponen utama adalah kombinasi linier-kombinasi linier tertentu dari  $D$  dimensi acak  $\mathbf{x}_j (j \in \{1, 2, \dots, D\})$ .

Secara geometris kombinasi linier ini merupakan sistem koordinat baru yang didapat dari rotasi sistem semula. Koordinat baru tersebut merupakan arah dengan variabilitas maksimum dan memberikan kovariansi yang lebih sederhana. Analisis komponen utama lebih baik digunakan jika variabel-variabel asal saling berkorelasi Analisis komponen utama merupakan penyelesaian masalah eigen yang secara matematis ditulis dalam persamaan :

$$\mathbf{C}\mathbf{v} = \lambda\mathbf{v} \quad 1)$$

yang mana variabilitas suatu dataset yang dinyatakan dalam matriks kovariansi  $\mathbf{C}$  dapat digantikan oleh suatu scalar tertentu  $\lambda$  tanpa mengurangi variabilitas asal secara signifikan.

Diberikan dataset matrik  $\mathbf{X}$  berukuran  $(n \times D)$  yang terdiri dari  $n$  observasi  $\mathbf{x}_i (i \in \{1, 2, \dots, n\})$  dengan  $D$  dimensi. Algoritma dari analisis komponen utama adalah sebagai berikut :

- 1) Hitung vektor rata-rata  $\bar{x}_j (j \in \{1, 2, \dots, D\})$  dengan

$$\bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n} \quad 2)$$

- 2) Hitung matriks kovariansi  $\mathbf{C}$  atau  $\text{cov}(\mathbf{X})$  dengan

$$\mathbf{C} = \text{cov}(x_k, x_j) = \frac{\sum_{i=1}^n (x_{ik} - \bar{x}_k)(x_{ij} - \bar{x}_j)}{n-1} \quad 3)$$

- 3) Hitung nilai eigen  $\lambda$  dan vektor eigen  $\mathbf{v}$  yang memenuhi persamaan :

$$|\mathbf{C} - \lambda\mathbf{I}| = 0 \quad 4)$$

dan

$$(\mathbf{C} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0} \quad 5)$$

- 4) Vektor eigen- vektor eigen yang didapatkan merupakan komponen utama-komponen utama untuk membentuk variabel baru. Variabel-variabel baru merupakan perkalian antara vektor eigen  $\mathbf{v}$  dengan matriks  $\mathbf{X}_a$ , yaitu matriks  $\mathbf{X}$  yang telah dinormalisasi (adjusted) yang dihitung dengan rumus :

$$\mathbf{x}_{a_i} = \frac{(x_i - \bar{x})}{s_i} \quad 6)$$

- 5) Sedangkan variansi yang dapat dijelaskan oleh variabel baru ke- $i$  tergantung persentase kontribusi  $p_i$  dari masing-masing nilai eigen, yang dihitung dengan rumus :

$$p_i = \frac{\lambda_i}{\sum_{j=1}^D \lambda_j} \times 100\% \quad 7)$$

Sedangkan penentuan jumlah variabel baru yang digunakan tergantung persentase kontribusi kumulatif dari kumulatif nilai eigen yang telah diurutkan dari nilai yang terbesar. Nilai persentase kontribusi kumulatif sampai komponen ke -  $r$  dihitung dengan rumus :

$$pk_r = \frac{\sum_{j=1}^r \lambda_j}{\sum_{j=1}^D \lambda_j} \times 100\% \quad 8)$$

dengan  $\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_D$

### 3. Teknik Reduksi Dimensi

Diberikan dataset matrik  $\mathbf{X}$  berukuran  $(n \times D)$  yang terdiri dari  $n$  observasi  $\mathbf{x}_i (i \in \{1, 2, \dots, n\})$  dengan  $D$  dimensi. Teknik reduksi dimensi mengikuti algoritma sebagai berikut :

- 1) Lakukan partisi dimensi himpunan  $\mathbf{X}$  menjadi  $\ell$  himpunan bagian, sehingga masing-masing  $\mathbf{x}_i$  menjadi berukuran  $d$  dimensi dengan

$$d = \frac{D}{\ell} \quad 9)$$

- 2) Lakukan untuk setiap subset data dengan  $d$  dimensi sebagai berikut :

- i. Hitung matriks kovariansi masing-masing subset menurut persamaan 3).
- ii. Hitung nilai eigen dan vektor eigen masing-masing subset menurut persamaan 4) dan 5).
- iii. Ambil vektor eigen  $\mathbf{v}$  dengan nilai eigen terbesar, di mana  $\lambda^* = \max\{\lambda_1, \lambda_2, \dots, \lambda_2\}$ .
- iv. Lakukan perhitungan observasi baru  $\mathbf{x}^*$  berukuran  $(n \times 1)$  menggunakan

$$\mathbf{x}^* = (\mathbf{v}^T \mathbf{x}_a^T)^T \text{ atau } \mathbf{x}^* = (\mathbf{x}_a \mathbf{v})^T \quad 10)$$

dengan  $\mathbf{x}_{a_i} (i \in \{1, 2, \dots, n\})$  adalah data normalisasi dan  $v_j (j \in \{1, 2, \dots, d\})$  adalah vektor eigen yang mempunyai variansi maksimum.

- 3) Gabungkan observasi baru menjadi dataset baru  $\mathbf{X}^*$  berukuran  $(n \times \ell)$ .

### 4. Percobaan dan Hasil

Ada 3 dataset yang diujikan dalam penelitian ini, yaitu :

1. Dataset konvolusi sinyal electromyograph (EMG) pada sudut pergerakan siku  $45^\circ$ ,  $90^\circ$ , dan  $135^\circ$  yang digunakan sebagai input robot lengan [2].
2. Dataset sonar [7].
3. Dataset sintesis, yaitu data hasil pembangkitan variabel random Normal berdimensi  $D = 1500$  dan data pembelajaran  $n = 90$  observasi, dan 15 observasi untuk validasi, dengan aturan pembangkitan sebagai berikut :

Tabel 1. Pembangkitan Data Sintesis

Kelas	Dimensi ke 1 - 500	Dimensi ke 501 - 1000	Dimensi ke 1001 - 1500
1	N(5,0.5)	N(7,0.5)	N(3,0.5)
2	N(6,0.5)	N(4,0.5)	N(8,0.5)
3	N(10,0.5)	N(8,0.5)	N(6,0.5)

#### 4.1 Dataset EMG

Pada data EMG, jumlah data pembelajaran  $n = 30$  observasi, jumlah dimensi  $D = 2000$  variabel dan data untuk validasi (*non learning*) sebanyak 18 observasi. Sedangkan metode pengklasifikasian yang digunakan

adalah metode Kohonen Self Organizing Maps (K-SOM), Learning Vector Quantization Neural Network (LVQNN) dan Backpropagation Neural Network (BPNN). Dengan menggunakan software Matlab 7.1.0.246(R14) Service Pack 3 diperoleh hasil sebagai berikut :

a. Total Variasi

Pada Tabel 2 dapat dijelaskan bahwa untuk jumlah variabel baru,  $\ell$  antara 5 sampai dengan 1000, rata-rata total variasi komponen utama pertama masing-masing partisi berkisar 97,31% - 99,99%. Artinya, informasi yang hilang dari hasil reduksi dimensi berkisar antara 0,01% - 2,69%. Namun untuk jumlah variabel baru,  $\ell = 2$ , rata-rata-total variasi variabel baru hanya sebesar 81,48%.

Tabel 2. Rata-rata Total Variasi

Jumlah variabel baru ( $\ell$ )	Rata-rata Total Variasi, $\bar{\lambda}_1$ (%)	Simpangan baku Total Variasi, $s_{\lambda_1}$ (%)	Waktu komputasi, t (detik)
asli (2000)	100,00	-	-
1000	99,99	0,00	3,72
500	99,99	0,01	0,86
250	99,99	0,04	0,39
200	99,98	0,06	0,41
100	99,94	0,20	0,31
50	99,83	0,40	0,11
20	99,35	1,21	0,06
10	98,88	0,67	0,08
5	97,31	1,47	0,06
2	81,48	19,28	0,22

b. Waktu Komputasi vs Jumlah Variabel Baru

Sedangkan hubungan antara waktu komputasi (partisi data dan perhitungan komponen utama) dengan jumlah variabel baru hasil partisi dapat dinyatakan dalam persamaan regresi (Tabel 3) sebagai berikut :

Tabel 3. Model Regresi Waktu Komputasi

The regression equation is WakKom = 0,0032 JumVar				
Predictor	Coef	SE Coef	T	P
Noconstant				
JumVar	0,0031993	0,0002980	10,74	0,000
S = 0,3482		R <sup>2</sup> = 92,76%		

Dari persamaan regresi tersebut dapat dijelaskan bahwa setiap komputasi variabel baru membutuhkan waktu rata-rata selama 0,0032 detik.

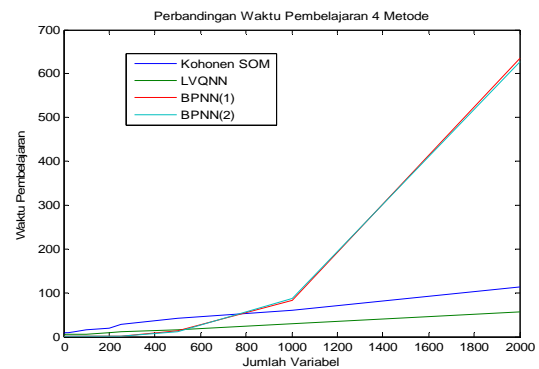
c. Waktu Pembelajaran

Dalam Tabel 4., waktu pembelajaran metode K-SOM lebih lama dibandingkan dengan metode LVQNN yaitu sekitar dua kalinya. Sedangkan waktu pembelajaran metode BPNN(1) tidak jauh berbeda dengan metode BPNN(2).

Tabel 4. Perbandingan Waktu Pembelajaran

Jumlah variabel baru ( $\ell$ )	Waktu Pembelajaran Metode Klasifikasi (detik)			
	K-SOM	LVQNN	BPNN (1)	BPNN (2)
asli (2000)	114,36	55,36	635,13	627,58
1000	59,81	29,38	83,67	86,28
500	41,97	14,47	12,84	10,89
250	27,59	10,25	1,69	1,94
200	19,33	8,22	1,23	1,27
100	15,48	6,02	1,08	0,50
50	11,77	5,53	0,64	0,36
20	10,08	4,81	0,92	0,34
10	9,13	4,50	0,19	0,33
5	8,38	4,42	0,30	0,31
2	7,81	2,94	0,28	0,30

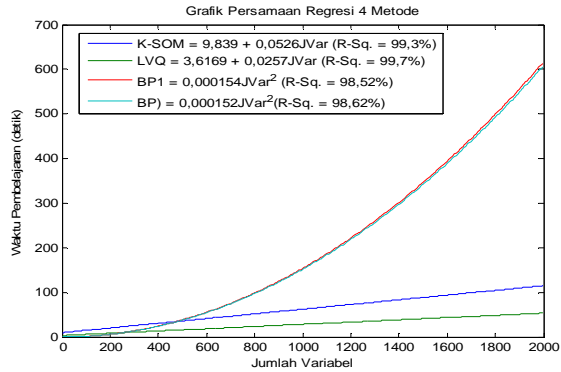
Dari Gambar 1. dapat dijelaskan bahwa pada metode K-SOM dan LVQNN, jumlah variabel mempunyai hubungan linier dengan waktu pembelajaran. Sedangkan pada metode BPNN(1) maupun BPNN(2), jumlah variabel mempunyai hubungan tidak linier dengan waktu pembelajaran.



Gambar 1. Perbandingan Waktu Pembelajaran 4 Metode

Berdasarkan informasi Tabel 4, dilakukan analisis regresi untuk melihat pola hubungan (pengaruh) kenaikan jumlah variabel baru terhadap waktu pembelajaran. Persamaan regresi terbaik untuk keempat metode pembelajaran tersebut disajikan dalam Gambar 2., yang mana dapat diinterpretasikan bahwa setiap penambahan satu variabel baru, waktu yang dibutuhkan untuk pembelajaran pada :

- Metode K-SOM cenderung naik sebesar 0,0526 detik (hubungan linier). Sedangkan pada metode LVQNN cenderung naik sebesar 0,0257 detik (hubungan linier)
- Metode BPNN(1) cenderung naik sebesar 0,000154 kali kuadrat jumlah variabel baru (detik). Agak berbeda dengan metode LVQNN, pada metode BPNN(1) jumlah variabel berpengaruh secara kuadratik terhadap waktu pembelajaran. Demikian juga untuk metode BPNN(2), cenderung naik sebesar 0,000152 kali kuadrat jumlah variabel baru (detik).



Gambar 2. Grafik Persamaan Regresi 4 Metode

#### d. Hasil Klasifikasi

Untuk mengetahui efektifitas dari metode reduksi dimensi dalam data EMG, dilakukan perbandingan persentase hasil klasifikasi yang valid dari variabel asli dan variabel hasil reduksi (masing-masing 10 perulangan).

Tabel 5. Persentase Hasil Klasifikasi Data *Learning*

Jumlah variabel baru ( $\ell$ )	Hasil klasifikasi yang valid (%)			
	K-SOM	LVQNN	BPNN (1)	BPNN (2)
asli (2000)	40,00	43,33	96,67	93,33
1000	40,00	60,00	100,00	93,33
500	40,00	60,00	100,00	96,67
250	40,00	60,00	96,67	96,67
200	40,00	60,00	93,33	90,00
100	40,00	60,00	100,00	100,00
50	40,00	56,67	100,00	100,00
20	40,00	60,00	100,00	100,00
10	40,00	46,67	100,00	100,00
5	40,00	36,67	100,00	100,00
2	40,00	33,33	100,00	100,00

Dari Tabel 5., pada metode K-SOM tidak ada perbedaan persentase hasil klasifikasi yang valid antara variabel asli (2000) dengan variabel hasil reduksi yaitu 40%. Pada metode LVQNN, variabel hasil reduksi lebih efektif dibandingkan variabel asli yang mana persentase hasil klasifikasi yang valid pada variabel asli hanya sebesar 43,33%. Sedangkan persentase hasil klasifikasi yang valid pada variabel hasil reduksi dengan  $\ell = 500, 250, 200, 100, 20$  sebesar 60%. Pada metode BPNN(1) maupun BPNN(2), variabel hasil reduksi juga relatif lebih efektif dibandingkan variabel asli yang mana persentase hasil klasifikasi yang valid pada variabel asli sebesar 96,67% untuk BPNN(1) dan 93,33% untuk BPNN(2). Sedangkan persentase hasil klasifikasi yang valid pada variabel hasil reduksi dengan  $\ell = 100, 50, 20, 10, 5, 2$  adalah sebesar 100% (untuk kedua metode BPNN).

Sedangkan pada Tabel 6., hasil klasifikasi pada data *non-learning* untuk metode LVQNN lebih bagus dibandingkan dengan metode yang lain yaitu dengan

modus 61,11%. Sedangkan metode K-SOM dan BPNN masing-masing hanya sebesar 38,89% dan 33,33%. Namun tidak ada perbedaan hasil klasifikasi antara variabel asli dengan variabel hasil reduksi.

Tabel 6. Persentase Hasil Klasifikasi Data *Non-Learning*

Jumlah variabel baru ( $\ell$ )	Hasil klasifikasi yang valid (%)			
	Kohonen SOM	LVQNN	BPNN (1)	BPNN (2)
asli (2000)	38,89	61,11	33,33	33,33
1000	38,89	61,11	33,33	33,33
500	38,89	61,11	33,33	33,33
250	38,89	61,11	33,33	33,33
200	38,89	61,11	33,33	33,33
100	38,89	61,11	33,33	33,33
50	38,89	61,11	33,33	33,33
20	38,89	61,11	33,33	33,33
10	38,89	61,11	33,33	33,33
5	38,89	55,56	33,33	33,33
2	38,89	61,11	33,33	33,33

Rendahnya hasil klasifikasi pada data *non-learning* disebabkan karena nilai koefisien variasi yang sangat besar yaitu antara -715,98 – 207,25 yang mengindikasikan bahwa data tidak berdistribusi Normal dan mempunyai variansi yang sangat besar.

#### 4.2 Dataset Sonar

Pada dataset sonar, jumlah data pembelajaran  $n = 160$  observasi, jumlah dimensi  $D = 60$  variabel dan data untuk validasi (*non learning*) sebanyak 48 observasi, diperoleh hasil sebagai berikut :

##### a. Total Variasi

Tabel 7. Rata-rata Total Variasi

Jumlah variabel baru ( $\ell$ )	Rata-rata Total Variasi, $\lambda_1$ (%)	Simp. baku Total Variasi, $s_{\lambda_1}$ (%)	Waktu komputasi, t (detik)
asli (60)	100,00	0,00	0,11
30	88,54	6,96	0,05
20	81,74	9,30	0,06
15	76,34	10,40	0,03
10	70,56	13,42	0,02
5	62,98	22,73	0,02
3	57,90	16,66	0,00
2	39,66	0,88	0,03

Pada Tabel 7. dapat dijelaskan bahwa, untuk jumlah variabel baru,  $\ell = 10, 15, 20, 30$ , rata-rata total variasi komponen utama pertama sebagai variabel baru dari masing-masing partisi masih berkisar 70,56% - 88,54%. Untuk  $\ell = 2, 3, 5$ , rata-rata total variasi komponen utama pertama sebagai variabel baru dari masing-masing partisi

sangat kecil yaitu berkisar 39,66% - 62,98%. Simpangan baku total variasi juga cenderung membesar. Artinya, informasi yang hilang dari hasil reduksi dimensi sangatlah signifikan. Selanjutnya diambil sampel matriks koefisien korelasi dari variabel ke-1 sampai dengan variabel ke-6 disajikan pada Tabel 8. Dari tabel tersebut, matriks koefisien korelasi variabel ke-1 sampai dengan variabel ke-6 signifikan secara statistik, namun secara umum besarnya koefisien korelasi tidak mendekati 1. Akibatnya total variasi yang dapat dijelaskan berkurang drastis atau persentase informasi data yang hilang sangat besar.

Tabel 8. Matriks Koefisien Korelasi ( $\rho$ ) dari V1 s/d V6

$\rho$	V1	V2	V3	V4	V5
V2	0,736 (0.00)				
V3	0,572 (0.00)	0,780 (0.00)			
V4	0,491 (0.00)	0,607 (0.00)	0,782 (0.00)		
V5	0,345 (0.00)	0,420 (0.00)	0,546 (0.00)	0,727 (0.00)	
V6	0,239 (0.00)	0,332 (0.00)	0,346 (0.00)	0,353 (0.00)	0,597 (0.00)

#### b. Hasil Klasifikasi

Untuk mengetahui efektifitas dari metode reduksi dimensi dalam dataset sonar, dilakukan perbandingan persentase hasil klasifikasi yang benar dari variabel asli dan variabel hasil reduksi (masing-masing 10 perulangan).

Tabel 9. Persentase Hasil Klasifikasi Data *Learning*

Jumlah variabel baru ( $\ell$ )	Hasil klasifikasi yang valid (%)			
	Kohonen SOM	LVQNN	BPNN (1)	BPNN (2)
60 (asli)	27,50	41,25	100,00	100,00
30	27,50	41,25	100,00	100,00
20	41,88	40,00	100,00	100,00
15	33,13	39,38	100,00	100,00
10	41,88	36,88	100,00	100,00
5	38,75	34,38	100,00	100,00
3	26,25	41,88	100,00	100,00
2	37,50	40,63	83,75	83,75

Dari Tabel 9., pada metode K-SOM, persentase hasil klasifikasi yang benar variabel hasil reduksi relatif lebih besar (antara 27,50% - 41,88%) dibandingkan dengan variabel asli (27,50%). Pada metode LVQNN, variabel hasil reduksi tidak jauh berbeda persentase hasil klasifikasi yang valid dengan variabel asli yaitu antara 34,38% - 41,25%. Pada metode BPNN(1) maupun BPNN(2), variabel hasil reduksi mempunyai persentase klasifikasi yang valid yang sama dengan variabel asli

yaitu sebesar 100%, kecuali untuk  $\ell = 2$  hanya sebesar 83,75%.

Sedangkan hasil klasifikasi pada data *non-learning* (Tabel 10) untuk metode KSOM tidak konsisten, yaitu berkisar antara 10,42% - 60,42%. Untuk metode LVQNN berkisar antara 31,61% - 36,38%. Sedangkan metode BPNN(1) dan BPNN(2) semuanya sama yaitu sebesar 35,42%, sehingga tidak ada perbedaan hasil klasifikasi antara variabel asli dengan variabel hasil reduksi.

Tabel 10. Persentase Hasil Klasifikasi Data *Non-Learning*

Jumlah variabel baru ( $\ell$ )	Hasil klasifikasi yang valid (%)			
	Kohonen SOM	LVQNN	BPNN (1)	BPNN (2)
60 (asli)	25,00	33,58	35,42	35,42
30	25,00	34,45	35,42	35,42
20	27,08	36,38	35,42	35,42
15	54,16	35,30	35,42	35,42
10	29,17	32,27	35,42	35,42
5	60,42	32,05	35,42	35,42
3	41,67	31,61	35,42	35,42
2	10,42	32,20	35,42	35,42

#### 4.3 Dataset Sintetis

Pada dataset sintetis, jumlah data pembelajaran  $n = 90$  observasi, jumlah dimensi  $D = 1500$  variabel dan data untuk validasi (*non learning*) sebanyak 15 observasi, diperoleh hasil sebagai berikut :

##### a. Total Variasi

Tabel 11. Rata-rata Total Variasi

Jumlah variabel baru ( $\ell$ )	Rata-rata Total Variasi, $\bar{\lambda}_1$ (%)	Simp. baku Total Variasi, $s_{\lambda_1}$ (%)	Waktu komputasi, t (detik)
asli (1500)	100,00	-	0,53
500	99,77	0,87	0,08
300	99,79	0,05	0,05
150	99,77	0,05	0,05
100	99,36	2,82	0,03
50	99,19	2,76	0,11
30	99,75	0,06	0,09
20	98,35	4,30	0,14
15	99,74	0,06	0,17
10	95,75	8,51	0,19
5	94,17	7,65	0,16
3	99,74	0,07	0,20
2	85,82	1,34	0,25

Pada Tabel 11. dapat dijelaskan bahwa rata-rata total variasi komponen utama pertama sebagai variabel baru dari masing-masing partisi sangat besar yaitu berkisar 94,17% - 99,79%, kecuali untuk  $\ell = 2$ , rata-rata total

variasi hanya sebesar 85,82%. Simpangan baku total variasi untuk  $\ell = 5$  dan  $\ell = 10$  lebih besar jika dibandingkan dengan  $\ell$  yang lain, yaitu masing-masing sebesar 7,65% dan 8,51%. Artinya, variabilitas total variasi variabel baru hasil reduksi pada  $\ell$  tersebut sangatlah besar.

**b. Hasil Klasifikasi**

Dari Tabel 12., pada metode K-SOM, persentase hasil klasifikasi yang valid pada variabel asli maupun variabel hasil reduksi pada data *learning* semuanya 100%. Pada metode LVQNN, variabel hasil reduksi untuk  $\ell = 50, 300, 500$  persentase hasil klasifikasi yang benar adalah 100%, selain itu hanya berkisar 33,33% - 60,00%.

Tabel 12. Persentase Hasil Klasifikasi Data *Learning*

Jumlah variabel baru ( $\ell$ )	Hasil klasifikasi yang valid (%)			
	Kohonen SOM	LVQNN	BPNN (1)	BPNN (2)
1500 (asli)	100,00	100,00	95,56	100,00
500	100,00	100,00	100,00	100,00
300	100,00	100,00	100,00	100,00
150	100,00	46,67	100,00	100,00
100	100,00	33,33	100,00	100,00
50	100,00	100,00	100,00	100,00
30	100,00	33,33	66,67	100,00
20	100,00	33,33	100,00	100,00
15	100,00	33,33	66,67	100,00
10	100,00	33,33	100,00	100,00
5	100,00	33,33	33,33	33,33
3	100,00	33,33	33,33	100,00
2	100,00	60,00	100,00	66,67

Pada metode BPNN(1) maupun BPNN(2) hampir semua variabel baru hasil reduksi mempunyai persentase hasil klasifikasi sebesar 100%, kecuali pada  $\ell = 3, 5, 15, 30$  untuk BPNN(1) dan  $\ell = 2, 5$  untuk BPNN(2).

Tabel 13. Persentase Hasil Klasifikasi Data *Non-Learning*

Jumlah variabel baru ( $\ell$ )	Hasil klasifikasi yang valid (%)			
	Kohonen SOM	LVQNN	BPNN (1)	BPNN (2)
1500 (asli)	33,33	33,33	33,33	33,33
500	33,33	33,33	33,33	33,33
300	33,33	33,33	33,33	33,33
150	33,33	33,33	33,33	33,33
100	33,33	33,33	33,33	33,33
50	33,33	33,33	33,33	33,33
30	33,33	33,33	33,33	33,33
20	33,33	33,33	33,33	33,33
15	33,33	33,33	33,33	33,33
10	33,33	66,67	33,33	33,33
5	33,33	66,67	33,33	33,33
3	33,33	33,33	33,33	33,33
2	33,33	33,33	33,33	33,33

Sedangkan hasil klasifikasi pada data *non-learning* pada Tabel 13., untuk semua metode hasil klasifikasi sebesar 33,33%, kecuali pada metode LVQNN untuk  $\ell = 5$  dan  $\ell = 10$  sebesar 66,67%. Nilai simpangan baku total variasi pada  $\ell$  tersebut sangatlah besar dibandingkan dengan yang lain.

**5. Kesimpulan**

Dari percobaan dan hasil penelitian ini dapat disimpulkan sebagai berikut :

1. Penggunaan komponen utama pertama data partisi sebagai variabel baru mempunyai hasil klasifikasi yang konsisten dengan data asli apabila nilai rata-rata total variasi berkisar 81,48% - 99,99%.
2. Teknik reduksi dimensi menggunakan komponen utama data partisi sangat signifikan mengurangi waktu pembelajaran.
3. Secara umum, performa hasil klasifikasi data *learning* dengan teknik reduksi dimensi ini sangat tinggi untuk beberapa metode dan beberapa kasus. Sedangkan performa hasil klasifikasi data *non-learning* sangat rendah, yaitu berkisar 25,00% - 66,67%. Hal ini disebabkan oleh keragaman nilai koefisien variasi dataset yang sangat besar, yang mana mengindikasikan bahwa data tidak berdistribusi Normal dan mempunyai variansi yang sangat besar.

**6. Referensi**

[1] Z. Qiao, L. Zhou and J.Z. Huang, "Sparse Linear Discriminant Analysis with Applications to High Dimensional Low Sample Size Data", IAENG International Journal of Applied Mathematics, 39:1,IJAM\_39\_1\_06, 2009.

[2] M. Susianti, "Identifikasi Sinyal Electromyograph (EMG) Pada Gerak Ekstensi-Fleksi Siku Dengan Metode Konvolusi dan Jaringan Syaraf Tiruan Untuk Input Robot Lengan, Proyek Akhir PENS, 2009.

[3] G.F.V. Borries, "Partition Clustering of High Dimensional Low Sample Size Data Based On P-Value", Kansas State University, Manhattan, Kansas, 2008

[4] D.C. Hoyle, "Automatic PCA Dimension Selection for High Dimensional Data and Small Size", Journal of Machine Learning Research, 2008 (2733-2759).

[5] L.I. Smith, "A Tutorial on Principal Components Analysis, 2002.

[6] R.A. Johnson & D.W. Wichern, "Applied Multivariate Statistical Analysis", Prentice Hall, 1988.

[7] \_\_\_\_\_, <http://www.cs.sfu.ca/~wangk/ucidata/dataset>