

Kelompok 11 :

- **Safira Amalia P.R (LDA) 2101191008**
- **Tyas Oksi (PCA) 2101191014**
- **Agita Purwandani(FA) 2101191006**

**Analisis Perbandingan Metode PCA, LDA dan FA untuk Klasifikasi Menggunakan
*Neural Network***

Data berdimensi tinggi merupakan tantangan tersendiri. Dalam persoalan klasifikasi, tantangan tersebut muncul dari kenyataan bahwa dimensi yang lebih tinggi secara drastis meningkatkan ruang vektor sehingga data menjadi terlihat sangat jarang. Hal ini menjadi masalah karena pembelajaran statistik membutuhkan sampel yang padat untuk menghasilkan hasil pembelajaran yang baik secara statistik baik dan dapat diandalkan. Diberikan sebuah himpunan untuk pelatihan dengan ukuran tertentu, ada suatu ukuran maksimum untuk vektor ciri dimana ukuran yang lebih besar akan memperburuk klasifikasi daripada memperbaikinya. Reduksi dimensi bertujuan untuk mengatasi persoalan ini dengan melakukan pemetaan data dari ruang berdimensi tinggi ke ruang berdimensi rendah. Pemetaan tersebut dilakukan agar informasi yang relevan tetap muncul pada hasil pemetaan dan di saat yang sama metode-metode pembelajaran dapat diaplikasikan dengan sukses.

Principal component analysis (PCA) adalah pelopor metode *subspace linear* pada reduksi dimensi. PCA berusaha menemukan transformasi ortogonal yang memproyeksikan data ke suatu *subspace* yang meminimalkan korelasi hasil proyeksi. *Subspace* ini disebut sebagai *principal subspace*. PCA merupakan suatu metode *unsupervised* dimana label pada data pelatihan tidak digunakan untuk melakukan pembelajaran. Dalam hal klasifikasi, karenanya, proyeksi PCA boleh jadi tidak optimal. Proyeksi PCA selalu berusaha mempertahankan semua jenis variasi secara maksimal tanpa menghiraukan faktor-faktor yang memunculkan variasi tersebut. Seperti pada gambar 1.

	precision	recall	f1-score	support
B	0.97	0.98	0.98	290
M	0.97	0.95	0.96	165
accuracy			0.97	455
macro avg	0.97	0.96	0.97	455
weighted avg	0.97	0.97	0.97	455

Gambar 1. Hasil akurasi pelatihan

Data latih PCA merupakan data yang sudah ada lalu diolah berdasarkan fakta yang sudah terjadi, melakukan representasi rule klasifikasi dengan mengelompokkan berdasarkan (B) dan (M) didapatkan nilai dari masing-masing data awal latih sebanyak 455 memiliki fitur sebanyak 30, dengan PCA nilai fitur yang digunakan hanya 3, didapatkan precision B (0.97) dan M (0.97) accuracy yang didapat sebesar 0.97

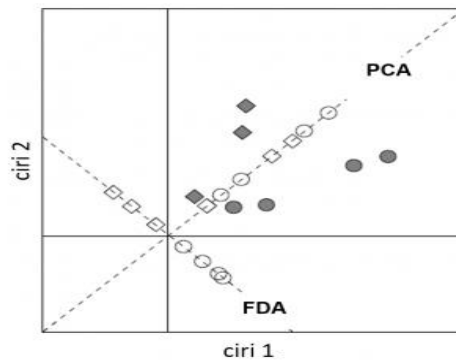
[[61 6] [3 44]]	precision	recall	f1-score	support
B	0.95	0.91	0.93	67
M	0.88	0.94	0.91	47
accuracy			0.92	114
macro avg	0.92	0.92	0.92	114
weighted avg	0.92	0.92	0.92	114

Gambar 2. Hasil Kurasi Pengujian PCA

Pada gambar 2 akan dilihat data uji PCA melalui prediksi x_{test} dan y_{test} , didapat matrik $\begin{bmatrix} 61 & 6 \\ 3 & 44 \end{bmatrix}$ didapatkan B dengan precision 0.95 dan M dengan nilai precision 0.88 sehingga akurasi yang didapat adalah 0.92 besarnya akurasi dipengaruhi oleh banyaknya database.

Fisher discriminant analysis (FDA) adalah metode *subspace* linear lain yang telah digunakan untuk mereduksi dimensi. Metode ini tergolong metode *supervised* yang bertujuan untuk menekan variasi dalam kelas dan di saat yang sama menguatkan variasi antarkelas. Gambar 3 menunjukkan keunggulan FDA dibanding PCA dalam persoalan klasifikasi. Seperti terlihat pada

gambar tersebut, proyeksi FDA menghasilkan data yang lebih mengelompok berdasarkan kelasnya.



Gambar 3. Keunggulan FDA dibanding PCA

Terdapat data 1, 2 dan 3 dengan precision yang paling tinggi dengan nilai 0.98 akan tetapi secara keseluruhan precision pada data yang lain nilainya baik yaitu diatas 0.80 akurasi dari seluruh data bernilai 0.96 sesuai dengan target pada pengambilan data latih, dapat dilihat pada gambar 4.

	precision	recall	f1-score	support
1	0.95	0.95	0.95	58
2	0.98	0.98	0.98	55
3	0.96	0.96	0.96	55
accuracy			0.96	168
macro avg	0.96	0.96	0.96	168
weighted avg	0.96	0.96	0.96	168

Gambar 4. Hasil Akurasi Pelatihan FDA

Data uji FDA dilakukan dengan menggunakan tiga database dengan precision terbaik 1.00 dan akurasi yang didapat sebesar 0.97 hal ini membuat metode FDA lebih baik dari PCA.

	precision	recall	f1-score	support
1	0.92	1.00	0.96	12
2	1.00	1.00	1.00	15
3	1.00	0.93	0.97	15
accuracy			0.98	42
macro avg	0.97	0.98	0.98	42
weighted avg	0.98	0.98	0.98	42

Gambar 5. Hasil Akurasi Pengujian FDA

Proyeksi FDA ternyata merupakan hal yang sama dengan solusi *maximum-likelihood* untuk “*reduced rank*” *Linear Discriminant Analysis* (LDA). LDA sendiri merupakan pengklasifikasi generatif yang mengasumsikan bahwa data di setiap kelas terdistribusi secara normal dengan struktur kovarian intra-kelas yang identik. Dalam reduced rank LDA, model generatif tersebut dibatasi agar semua mean dari kelas-kelas yang ada terletak pada suatu *subspace linear* berdimensi rendah yang sama.

	precision	recall	f1-score	support
B	0.98	0.99	0.99	290
M	0.99	0.97	0.98	165
accuracy			0.98	455
macro avg	0.99	0.98	0.98	455
weighted avg	0.98	0.98	0.98	455

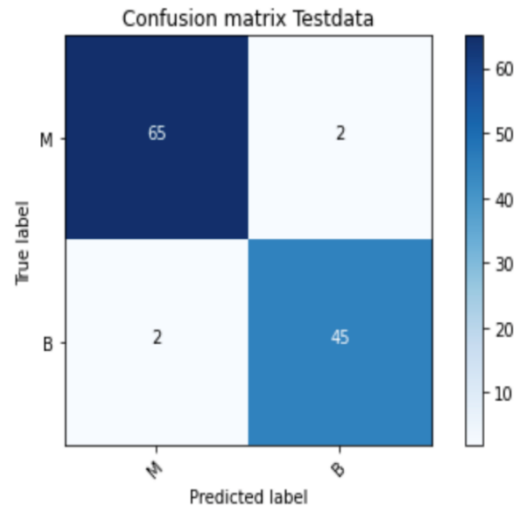
Gambar 6. Hasil Akurasi Pelatihan LDA

Berdasarkan gambar 6, diproses untuk mengetahui precision dari nilai B dan M dengan precision tertinggi bernilai 0.99 dan akurasi yang didapat adalah 0.99, selanjutnya dilakukan pengujian seperti pada gambar 7.

	precision	recall	f1-score	support
B	0.97	0.97	0.97	67
M	0.96	0.96	0.96	47
accuracy			0.96	114
macro avg	0.96	0.96	0.96	114
weighted avg	0.96	0.96	0.96	114

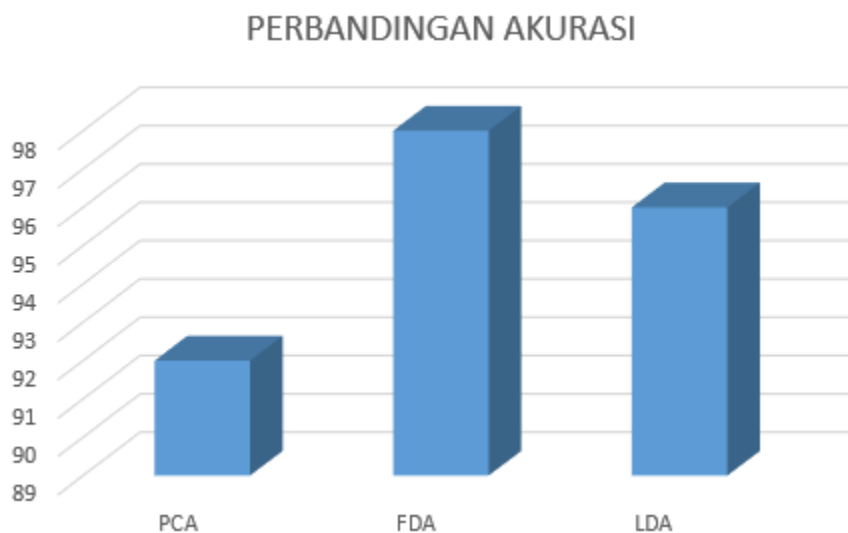
Gambar 7. Hasil Akurasi Pelatihan LDA

Pengujian data LDA diambil dari beberapa database yang telah ditentukan dari awal, sehingga didapatkan nilai B sebesar 0.97 sedangkan nilai M sebesar 0.96 dari gambar 7 dapat dilihat akurasi dengan menggunakan metode LDA didapatkan hasil 0.96 dimana hasil ini lebih besar dari pada menggunakan metode PCA.



Gambar 8. *Confusion matrix Testdata*

Data awal memiliki 30 fitur, dengan menggunakan metode LDA dari 30 fitur menjadi satu fitur LDA seperti yang dilihat pada gambar 8, dimana label B dideteksi menjadi label M yaitu dua buah, dan label M dideteksi menjadi label B adalah dua buah, hal tersebut terjadi dikarenakan fitur pada label B dan M ada yang sama.



Gambar 9. Perbandingan Akurasi

Mengacu pada gambar 9, Metode klasifikasi diatas dapat disimpulkan untuk akurasi terbaik adalah FDA metode ini tergolong metode *supervised* yang bertujuan untuk menekan variasi dalam kelas dan di saat yang sama menguatkan variasi antarkelas, selanjutnya adalah LDA karena dapat pengklasifikasikan secara generatif yang mengasumsikan bahwa data di setiap kelas terdistribusi secara normal dengan struktur kovarian intra-kelas yang identic, lalu untuk akurasi terendah diantara ketiga metode tersebut adalah PCA.