

ALGORITMA PRINCIPAL COMPONENT ANALYSIS DALAM PEMROSESAN SINYAL ELECTROKARDIOGRAM

Husaini¹, Huzaeni², Fahmi³

^{1,2,3} Jurusan Teknologi Informasi dan Komputer Politeknik Negeri Lhokseumawe
Jln. B.Aceh Medan Km.280 Buketrata 24301 INDONESIA

¹husaini@pnl.ac.id, ²zaini_pnl@yahoo.co.id, ³fahmi_mfj@yahoo.com

Abstrak— Principal Component Analysis (PCA) merupakan salah satu teknik yang ada dalam statistic dan merupakan metode non parametric untuk mengekstraksi informasi-informasi yang bersesuaian dari sekumpulan data yang masih diragukan dan memerlukan proses untuk menghilangkan gangguan-gangguan yang ada. Data yang dimaksud salah satunya adalah sinyal elektrocardiogram (EKG). Sinyal EKG merupakan sinyal yang diperoleh dari rekaman aktifitas elektrik dari jantung. Rekaman sinyal EKG tidak saja digunakan untuk tujuan diagnosa, tapi juga disimpan sebagai referensi dalam mengklasifikasi EKG arrhythmia. Untuk mendapatkan hasil yang lebih baik maka data-data sinyal EKG akan direduksi dimensinya dengan tujuan untuk menghilangkan data-data yang tidak sesuai, tidak relevan dan data redundant sehingga dapat menghemat biaya komputasinya dan mencegah data-data yang over-fitting. Tulisan ini memaparkan tentang ide dasar dari PCA dalam mereduksi dimensi data-data dari sinyal EKG. Hasil yang ditampilkan adalah berupa proses-proses dalam algoritma PCA dan akurasi klasifikasi sinyal dengan metode KNN dan Naive Bayes.

Kata kunci : principal component analysis (PCA), sinyal EKG, reduksi dimensi

Abstract— *The Principal Component Analysis (PCA) is one of the existing techniques in statistics and a non parametric method for extracting the information from a collection of data that still in doubt and requires a process to remove any disturbances. The data in question one of them is the signal elektrocardiogram (ECG). ECG signals are signals obtained from recording electrical activity from the heart. ECG signal recording is not only used for diagnostic purposes, but is also stored as a reference in classifying ECG arrhythmias. To get better results then the ECG signal data will be reduced the dimension. The aim to removed data that are not appropriate, irrelevant and redundant data so as to save the cost of computing and prevent data over-fitting. This paper describes the basic idea of PCA in reducing the dimensions of data from ECG signals. The results shown are the processes in PCA algorithm and signal classification accuracy by KNN and Naive Bayes methods.*

Keywords— *Principal Component Analysis, ECG Signal, reduction dimentionality*

I. PENDAHULUAN

PCA merupakan salah satu metode yang paling umum digunakan dalam komunitas statistik dan data mining. PCA mempunyai fungsi sebagai salah satu alat yang diaplikasikan mulai dari ekstraksi informasi sampai dengan reduksi dimensi untuk visualisasi data. Jangkauan subjek data PCA menjadi meningkat karena PCA dianggap sebagai suatu aplikasi. Penerapan PCA berkembang mulai sosiologis dan data medis skala kecil dengan dimensi puluhan sampai dengan kumpulan data dengan dimensi yang melampaui ukuran puluhan ribu bahkan sampai berkembang menjadi dimensi dengan ukuran jutaan untuk data biomedis dan multimedia.

PCA dalam pengolahan data baik itu data numeric maupun data dalam bentuk sinyal merupakan suatu metode yang digunakan dalam mengidentifikasi pola dari data dengan dimensi yang cukup tinggi dan mentransformasikan fitur-fitur baru melalui proses ekstraksi fitur yang kemudian dapat direpresentasikan sebagai fitur-fitur yang telah direduksi dimensinya.

Reduksi dimensi sangat disarankan generalisasi yang lebih baik dan fitur yang lebih sedikit berdasarkan jumlah sinyal aslinya. Tujuan utama dari reduksi dimensi ini adalah untuk menjaga informasi maksimum dari data asli yang diberikan sebagai suatu kriteria optimasi (relevansi) [1], sehingga ada suatu solusi yang ditawarkan, untuk menyelesaikan masalah reduksi dimensi ini yaitu dengan seleksi fitur di tingkat klasifikasi yang lebih besar dari 50.

Permasalahan reduksi dimensi dapat didefinisikan sebagai berikut ini : Diasumsikan bahwa ada sekumpulan data

yang direpresentasikan dalam matriks X dengan ukuran $n \times D$ dimana n merupakan vector data X_i ($i \in \{1, 2, \dots, n\}$) D merupakan ukuran dimensinya. Selanjutnya diasumsikan juga bahwa kumpulan data mempunyai basis dimensi d (dimana $d < D$, dan sering juga $d \leq D$). Istilah basis dimensi dalam matematika dapat berarti bahwa titik dalam kumpulan data X terletak atau dekat dengan kedimensian d yang ditanamkan dalam ruang dimensi D . Teknik-teknik reduksi dimensi mentransformasikan sekumpulan data X dengan dimensi D ke dalam kumpulan data baru Y dengan dimensi d . [2].

Beberapa metode telah diimplementasikan untuk reduksi dimensi ini diantaranya adalah penggunaan teknik *large number of nonlinear* ([3,4]), *Principle Component Analysis (PCA)* [5], *Independent Component Analysis (ICA)*, dan *multidimensional scaling (MDS)* [6].

Reduksi yang dilakukan pada penelitian ini adalah reduksi dimensi untuk sinyal EKG. Data sinyal EKG dengan dimensi tinggi cenderung memiliki klasifikasi lambat sehingga pada penelitian dilakukan percobaan untuk mengurangi beban dengan mereduksi dimensi, hal ini penting dilakukan untuk mempercepat proses klasifikasi.

Reduksi dimensi pada EKG merupakan transformasi data dimensi tinggi kedalam representasi yang mempunyai nilai dan berarti. Idealnya representasi berkurang harus memiliki dimensi yang sesuai dengan dimensi intrinsik data. Dimensi intrinsik data adalah jumlah parameter minimum kebutuhan dari sifat objek yang diamati dari data. Reduksi dimensi dapat dicapai dengan dua model yaitu Fitur seleksi dan Fitur Ekstraksi

II. METODOLOGI PENELITIAN

Penelitian ini akan dibagi dalam tiga tahap yaitu pertama mengimplementasikan PCA untuk fitur ekstraksi, kedua penggunaan algoritma *compressed sensing* untuk seleksi fitur dan ketiga klasifikasi dengan menggunakan KNN classifier dan Naive Bayes. Perlakuan ketiga tahapan tersebut akan dilakukan pada sinyal ECG. Data sinyal ECG akan diambil dari www.physionet.org selanjutnya dilakukan proses *denosing* untuk menghilangkan *noise* yang terdapat pada sinyal EKG. Detail pelaksanaan penelitian ini adalah :

a. **Preprocessing**

Preprocessing ini dilakukan untuk membersihkan sinyal ECG dari gangguan-gangguan, karena Sinyal ECG dari permukaan tubuh sering terkontaminasi oleh berbagai macam suara seperti gangguan aliran listrik, kebisingan elektromiografi (EMG), artefak gerak elektroda, *power line interface*, *base line wandering* dan sebagainya. Suara ini membawa hambatan bagi diagnosis penyakit kardiovaskular. Untuk menghilangkan suara-suara di atas dalam sinyal EKG, maka akan menggunakan FFT filtering,

b. **Ekstraksi fitur dengan PCA**

Ekstraksi Fitur merupakan proses mengubah input data kedalam sekumpulan fitur. Ekstraksi fitur bertujuan untuk mengekstrak fitur dengan memproyeksikan data asli dengan dimensi tinggi ke ruang dimensi yang lebih rendah melalui transformasi aljabar. Metode yang akan digunakan dalam penelitian ini untuk mengekstrak fitur dari EKG setelah proses preprocessing adalah PCA. Penggunaan PCA untuk proses ekstraksi dimaksudkan untuk menentukan kombinasi linear dari variabel input yang memaksimalkan variannya. Hasil ini nantinya akan digunakan untuk tahapan selanjutnya yaitu untuk *compressive sensing*.

c. **Ekstraksi seleksi dengan *compressive sensing***

Ekstraksi fitur dilakukan untuk mereduksi dimensi data dengan memilih hanya sebagian dari fitur (variabel prediktor) untuk mendapatkan modelnya. Pemilihan tersebut dilakukan karena banyak data fitur yang redundan tidak banyak memberikan informasi yang berarti dalam reduksi dimensi. Eksperimen yang akan diujikan pada untuk reduksi dimensi adalah pada sinyal ECG dengan ukuran 1200*256 dan hasil yang akan dilihat adalah untuk uji reduksi dengan dimensi 10, 20, 30, 40 dan 256..

d. **Klasifikasi**

Klasifikasi dilakukan untuk melihat performa dan akurasi dari eksperiment yang dilakukan dengan variasi dimensi yang telah ditentukan pada tahap ekstraksi seleksi. Metode klasifikasi yang digunakan adalah *Naive Bayes Classifier* dan KNN.

III. HASIL DAN PEMBAHASAN

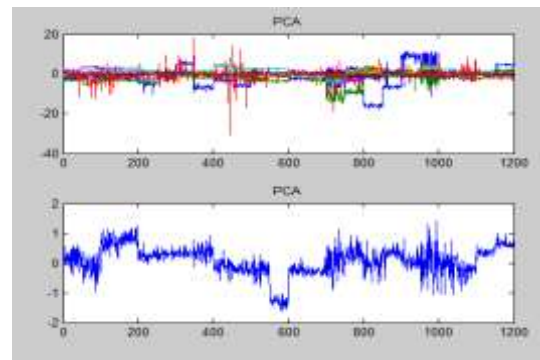
Percobaan yang dilakukan untuk reduksi dimensi ini adalah untuk sinyal ECG dengan ukuran 1200*256. Reduksi

dimensi yang dilakukan dengan PCA menggunakan dimensi 10,20 and 30.

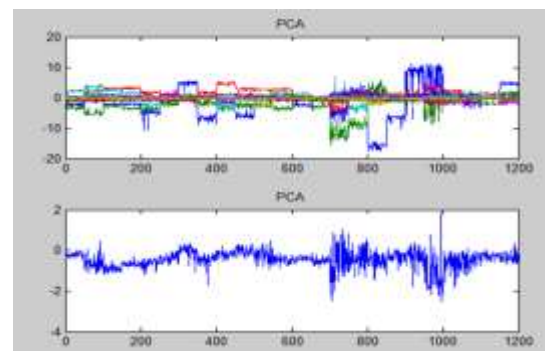
Berikut beberapa langkah implementasi PCA untuk analisa pemrosesan sinyal EK :

- Langkah 1. Mengambil data EKG dari www.physionet.org,
- Langkah 2. Menghitung rata-rata dari data EKG
- Langkah 3. Menghitung selisih data EKG dengan tiap data lainnya
- Langkah 4. Menghitung matriks kovarians dari langkah 3.
- Langkah 5. Menghitung vector eigen dan nilai eigen dari matriks kovarian.
- Langkah 6. Mengurutkan nilai eigen dan vector eigen dengan urutan berdasarkan nilai eigen paling tinggi.
- Langkah 7. Menerapkan PCA pada data sinyal EKG untuk vektorvektor fiturnya.
- Langkah 8. Mendapatkan nilai data ECG yang baru dengan rumus : $\text{DataBaru}=[\text{Baris Vektor fitur}]T^*$

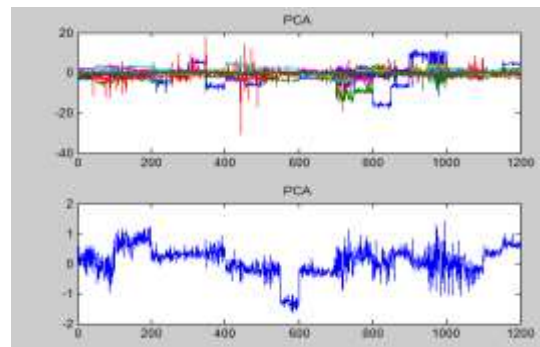
Hasil percobaan untuk masing-masing dimensi seperti yang ditunjukkan pada gambar 1-3 berikut ini :



Gamabr. 1. Reduksi dengan dimensi 10



Gambar. 2. Reduksi dengan dimensi 20



Gambar. 3. Reduksi dengan dimensi 30
K-Nearest Neighbour(K-NN)

K-NN merupakan algoritma yang sederhana yang berbasis pada perhitungan jarak. K-NN biasanya digunakan dalam estimasi atau prediksi secara statistic. K-NN sudah mulai diperkenalkan sejak tahun 1970 sebagai suatu teknik non parametrik.

Sama juga dengan beberapa metode klasifikasi yang lain. Kumpulan data untuk K-NN juga terdiri dari data training dan data uji.

Langkah-langkah untuk menghitung K-NN dengan data yang digunakan kuantitatif adalah :

1. Menentukan nilai K=jumlah tetangga terdekat .
2. Menghitunglah jarak antara data prediksi dengan semua sampel yang ada dengan menggunakan beberapa metode seperti euclidian , manhattan, dan minkoluski.

$$\text{Euclidian : } d(x_i y_i) = \sqrt{\sum_{i=1}^k (x_i - y_i)^2}$$

$$\text{Manhattan : } d(x_i y_i) = \sum_{i=1}^k |x_i - y_i|$$

$$\text{Minkoluski : } d(x_i y_i) = \sum_{i=1}^k ((|x_i - y_i|)^q)^{1/q}$$

3. Mengurutkan semua data yang telah dihitung dan hitung Nearest Neighbournya berdasarkan nilai k-nya.
4. Mengumpulkan urutan data tersebut sebagai target datanya.

Percobaan untuk KNN yang dilakukan dalam penelitian ini menggunakan nilai K=1, 3, 5, 7, 9.

Hasil akurasi yang diperoleh dari penggunaan dua metode untuk klasifikasi yaitu KNN dan Naive Bayes seperti table 1 dan tabel 2 berikut ini :

Tabel 1. Akurasi dengan Metode KNN dan Naive Bayes

	Accuracy (%)			
	Dimensional			
	256	10	20	30
Naïve Bayes	52.67	16.58	18.08	17.17
KNN	98.33	50.75	50.92	51.08

Tabel 2. Akurasi KNN dengan berbagai variasi nilai N

	Accuracy (%)			
	Dimensional			
	256	10	20	30

K=1	98.33	50.75	50.92	51.08
K=3	98.33	49.92	50.25	50.58
K=5	98.33	48.17	48.25	48.33
K=7	98	47.33	47.58	47.75
k=9	97.83	46.75	46.42	46.5

IV. KESIMPULAN

Reduksi dimensi yang dilakukan untuk sinyal EKG menggunakan algoritma PCA dengan berbagai variasi dimensi memperoleh hasil yang berarti sebagai gambaran untuk menghilangkan gangguan-gangguan terhadap sinyal, dan berdasarkan perhitungan akurasi yang dilakukan terhadap sinyal EKG yang berbasis PCA diperoleh bahwa Metode KNN memiliki keakuratan yang lebih baik dibandingkan Naïve Bayes.

Penelitian lanjutan dapat dilakukan dengan mengimplementasikan fitur-fitur ekstraksi lainnya seperti KPCA(*Kernel Principal Component Analysis*), *Isomap* dan *Laplacian Eigenmaps*

REFERENSI

- [1] L. F. Giraldo, E. Delgado, J. C. Riano, and G. Castellanos, "Feature selection using hybrid evaluation approaches based on genetic algorithms," in Proc. IEEE Comput. Soc. Press Proc., Electron., Robot. Automotive Mech. Conf. (CERMA 2006), vol. 2, pp. 242–247.
- [2] Laurens van der Maaten, Eric Postma, and Jaap van den Herik, *Dimensionality Reduction: A Comparative Review*, Elsevier, 2008
- [3] C.J.C. Burges. *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*, chapter Geometric Methods for Feature Selection and Dimensional Reduction: A Guided Tour. Kluwer Academic Publishers, 2005.
- [4] J. Wang, Z. Zhang, and H. Zha. Adaptive manifold learning. In *Advances in Neural Information Processing Systems*, volume 17, pages 1473–1480, Cambridge, MA, USA, 2005. The MIT Press
- [5] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. ed. New York: Spring-Verlag, 2002
- [6] D. Blei, A. Ng, M. Jordan, and J. Lafferty. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:2003, 2003