

## **ABSTRACT**

### **PRINCIPAL COMPONENT ANALYSIS WITH MISSING DATA USING *NONLINEAR ITERATIVE PARTIAL LEAST SQUARES (NIPALS)* METHOD**

**By**

**Anindia Putri**

Principal component analysis (PCA) is a multivariate statistical technique which is used to reduce large dimension and correlated data into smaller dimension and uncorrelated data. The conventional PCA (i.e. by using the eigen vectors of the covariance or correlation matrix as the coefficient of the principal components) cannot be used to analyze data with missing values. One method that can be used to solve missing values problem in PCA is by using the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm. NIPALS algorithm decomposes data matrix  $\mathbf{X}$  into scores matrix  $\mathbf{T}$  and loadings matrix  $\mathbf{P}$ . In this research, investigation of PCA using NIPALS algorithm was conducted by using data with missing values 1%, 3%, 5%, 10%, 15%, 20%, 25% and 28%. Based on the bias, Mean Square Errors (MSE) and spectral norm, the result shows that NIPALS algorithm still can overcome the missing values as much as 28%. However, for the missing values as much as 25% or more, the resulted principal components were not accurate.

**Keyword : Principal component analysis, missing data, *nonlinear iterative partial least squares (NIPALS) algorithm***

## ABSTRAK

### ANALISIS KOMPONEN UTAMA DENGAN DATA HILANG MENGUNAKAN METODE *NONLINEAR ITERATIVE PARTIAL LEAST SQUARES* (NIPALS)

Oleh

**Anindia Putri**

Analisis komponen utama (AKU) adalah salah satu teknik statistika multivariat yang digunakan untuk mereduksi dimensi data yang berukuran besar dan saling berkorelasi menjadi dimensi data yang berukuran lebih kecil dan tidak saling berkorelasi. AKU konvensional (yaitu dengan menggunakan vektor eigen dari matriks varian kovarian atau korelasi sebagai koefisien dari komponen utama) tidak dapat digunakan untuk menganalisis suatu data yang mengandung nilai yang hilang. Salah satu cara untuk mengatasi data hilang dalam AKU yaitu dengan menggunakan algoritma *Nonlinear Iterative Partial Least Squares* (NIPALS). Algoritma NIPALS mendekomposisikan matriks  $X$  menjadi matriks *scores*  $T$  dan matriks *loadings*  $P$ . Dalam penelitian ini, AKU menggunakan data hilang dengan algoritma NIPALS yang memuat nilai hilang sebanyak 1%, 3%, 5%, 10%, 15%, 20%, 25%, dan 28%. Berdasarkan biasanya, *Mean Square Errors* (MSE) dan *spectral norm*, menyatakan bahwa algoritma NIPALS masih dapat mengatasi nilai hilang sebanyak 28%, namun untuk nilai hilang sebanyak lebih dari atau sama dengan 25% hasil komponen utama yang didapat sudah tidak akurat.

**Kata Kunci :** analisis komponen utama, data hilang, algoritma *nonlinear iterative partial least squares* (NIPALS)