

## INTISARI

### ANALISIS KOMPONEN UTAMA *ROBUST SPARSE* DENGAN PENDEKATAN *PROJECTION-PURSUIT* PADA DATA BERDIMENSI TINGGI

Oleh

Dwi Resti Indah Puspitawati

14/364229/PA/15957

Pada era digital ini, fenomena *big data* mulai bermunculan. Hal yang paling menonjol dari sebuah *big data* yaitu dimensi datanya yang tinggi, dan karenanya perlu dilakukan reduksi dimensi. *Principal Component Analysis* merupakan salah satu metode reduksi dimensi yang didasarkan pada matriks kovarians yang tidak *robust* terhadap *outlier*. Oleh karena itu diperlukan *Robust PCA* pada data yang mengandung *outlier*. Salah satu metode *Robust PCA* yaitu dengan menggunakan pendekatan *Projection-pursuit*, yaitu mencari arah *robust PCA* yang mampu memaksimalkan *projection index*, dan dalam hal ini digunakan  $Q_n^2$  sebagai *projection index*. Telah diketahui bahwa *Classical PCA* dan *Robust PCA* merupakan kombinasi linear dari seluruh variabel. Hal ini akan menghambat proses interpretasi komponen utama, karena peneliti tidak dapat mengetahui variabel mana yang berperan penting dalam pembentukan komponen utama. Untuk itu, diperlukan *Sparse PCA* guna mempermudah proses interpretasi.

Pada skripsi ini akan dibahas metode pembentukan komponen utama *Robust Sparse PCA* dengan menggunakan pendekatan *Projection-pursuit*, yang merupakan kombinasi dari *Robust PCA* dan *Sparse PCA*. Pada skripsi ini akan dilihat performa dari metode *Classical PCA*, *Robust PCA*, *Sparse PCA*, dan *Robust Sparse PCA* dari sisi kemudahan dalam proses interpretasi komponen utama, persentase variabilitas data yang dapat dijelaskan oleh  $k$  komponen utama, dan kemampuan dalam mendeteksi serta membedakan jenis *outlier*.

Kata kunci : reduksi dimensi, *outlier*,  $Q_n^2$ , *projection-pursuit*, *Robust PCA*, *Robust Sparse PCA*

## **ABSTRACT**

### **ROBUST SPARSE PRINCIPAL COMPONENT ANALYSIS WTH PROJECTION-PURSUIT APPROACH ON HIGH DIMENSIONAL DATA**

By

Dwi Resti Indah Puspitawati

14/364229/PA/15957

In this digital era, big data phenomenon began to emerge. The most prominent thing of a big data is its high data dimension, and hence need a dimension reduction. Principal Component Analysis is one of the dimension reduction methods based on covariance matrices that are not robust to outliers. Therefore Robust PCA is required on data containing outliers. One of the Robust PCA methods is to use the Projection-pursuit approach, which is looking for a robust PCA direction that maximizes the projection index, and in this case  $Q_n^2$  is used as the projection index. It is known that Classical PCA and Robust PCA are linear combinations of all variables. This will inhibit the process of interpretation of the PCs, because researchers can't know which variable plays an important role in the formation of the PCs. For that, Sparse PCA required to simplify the process of interpretation.

In this thesis will be discussed the Robust Sparse PCA using the Projection-pursuit approach, which is a combination of Robust PCA and Sparse PCA. In this paper we will see the performance of Classical PCA, Robust PCA, Sparse PCA, and Robust Sparse PCA methods in terms of convenience in the process of interpretation of the PCs, the percentage of data variability that can be explained by the k PCs, and the ability to detect and differentiate the types of outliers.

**Keywords:** dimensional reduction, outlier,  $Q_n^2$ , projection-pursuit, Robust PCA, Robust Sparse PCA