

## INTISARI

### **SPARSE RIDGE FUSION UNTUK MENGATASI MULTIKOLINEARITAS DALAM REGRESI LINEAR PADA DATA BERDIMENSI TINGGI**

Oleh

Putri Azizatun Hidayati

15/383369/PA/17029

Analisis regresi banyak digunakan diberbagai bidang, bahkan akhir-akhir ini analisis regresi sangat berguna untuk bidang kesehatan dan farmasi. Namun, sekarang ini banyak data yang jumlah observasinya lebih sedikit daripada jumlah prediktornya ( $p > n$ ) seperti pada data NIRS dan *microarray*. Data tersebut tentunya memiliki masalah multikolinearitas yang tinggi yang mana analisis regresi dengan *Ordinary Least Square* (OLS) tidak dapat menanganinya. Oleh karena itu, regresi terpenalti seperti *ridge*, *lasso*, *smooth lasso*, dan *elastic-net* mulai diperkenalkan. Regresi terpenalti yang disebut *Ridge Sparse Fusion* juga mulai diperkenalkan untuk mengatasi masalah tersebut.

Skripsi ini membandingkan performasi dalam bentuk *Mean Squared Error* antara *Sparse Ridge Fusion* dengan analisis *lasso* dan *smooth lasso* pada data yang berdimensi tinggi. Hasil yang diperoleh adalah *Mean Squared Error* dari metode *Sparse Ridge Fusion* lebih rendah daripada *Mean Squared Error* dari metode *lasso* dan *smooth lasso*. Hal tersebut menandakan bahwa metode *Sparse Ridge Fusion* dapat dikatakan lebih akurat daripada metode *lasso* dan *smooth lasso*.

Kata kunci: analisis regresi, *ridge*, *elastic-net*, *lasso*, *smooth lasso*, *sparse ridge fusion*, *high dimension*

## ABSTRACT

### ***SPARSE RIDGE FUSION FOR MULTICOLINEARITY IN LINEAR REGRESSION WITH HIGH DIMENSIONAL DATA***

by

Putri Azizatun Hidayati

15/383369/PA/17029

Regression analysis is widely used in various fields, even recently regression analysis is very useful for the health and pharmaceutical fields. However, at present there is a lot of data whose number of observations is less than the number of predictors ( $p > n$ ) as in the NIRS and microarray data. The data certainly has a high multicollinearity problem in which Ordinary Least Square (OLS) regression analysis cannot handle it. Therefore, penalized regression such as ridge, lasso, smooth lasso, and elastic-net was introduced. Penalized regression called Sparse Ridge Fusion was also introduced to overcome this problem.

This thesis compares performance in the form of Mean Squared Error between Sparse Ridge Fusion with lasso and smooth lasso analysis on high dimensional data. The results obtained are the Mean Squared Error of the Sparse Ridge Fusion method lower than the Mean Squared Error from the lasso and smooth lasso method. This indicates that the Sparse Ridge Fusion method can be said to be more accurate than the lasso and smooth lasso method.

Keywords: regression analysis, ridge, elastic-net, lasso, smooth lasso, sparse ridge fusion, high dimension