

INTISARI

SELEKSI VARIABEL DENGAN METODE *SCREENING* KOLMOGOROV-SMIRNOV UNTUK DATA SURVIVAL BERDIMENSI TINGGI

Oleh

Syarto Musthofa

17/418729/PPA/05513

Ada banyak metode *screening* variabel yang bisa menangani data berdimensi tinggi. Beberapa dari metode tersebut bisa mengurangi dimensi data secara efektif dan menjamin semua variabel aktif tetap muncul dengan probabilitas tinggi. Namun, kebanyakan prosedur *screening* yang ada saat ini dikembangkan hanya untuk data lengkap berdimensi tinggi dan tidak layak diterapkan pada data survival dengan informasi tersensor. Metode *Screening* Kolmogorov-Smirnov dapat dimodifikasi untuk mengatasi masalah ini dengan mengganti fungsi distribusi kumulatif dengan fungsi survival yang diestimasi dengan estimator Kaplan-Meier. Metode ini dapat bekerja dengan berbagai tipe kovariat baik itu kontinu, diskrit, maupun kategorikal. Performa dari metode ini diukur berdasarkan studi simulasi. Suatu contoh data riil mengenai ekspresi gen juga digunakan sebagai aplikasi dari metode ini.

Kata kunci: Metode *screening*, data berdimensi tinggi, data survival

ABSTRACT

VARIABLE SELECTION BY KOLMOGOROV-SMIRNOV SCREENING METHOD FOR HIGH-DIMENSIONAL SURVIVAL DATA

By

Syarto Musthofa

17/418729/PPA/05513

There are numerous variable screening methods available for high-dimensional data. Some of the methods can effectively reduce the dimensionality while ensuring that all the active variables can be retained with high probability. However, most existing screening procedures are developed for high-dimensional complete data and cannot be applicable to censored survival data. The Kolmogorov-Smirnov Screening Method could be modified to overcome this problem by replacing the cumulative distribution function with survival function which estimated by Kaplan-Meier estimator. This method can work with many types of covariates including continuous, discrete, and categorical variables. The performance of this method presented via simulation study. A real data example of gene expression is used to illustrate the application of the method.

Keywords: Screening method, high-dimensional data, survival data