

INTISARI

Regresi Buckley-James dengan Boosting untuk Data Tersensor Kanan

Oleh

Muhammad Rizki Ramadhan

13/347975/PA/15409

Ekspresi gen yang diperoleh dari hasil *DNA microarray* menghasilkan jumlah yang sangat banyak. Untuk melihat ekspresi gen mana yang mempengaruhi penyakit suatu pasien menjadi kesulitan tersendiri. Metode *Boosting* telah banyak digunakan pada klasifikasi dan juga regresi untuk memilih variabel independen dalam pembentukan suatu model. Kesulitan bertambah jika pada data tersebut mengandung data yang tersensor. Salah satu metode yang sering digunakan untuk menangani data yang tersensor adalah regresi Cox atau yang biasa dikenal dengan *Cox's proportional hazard model* yang diperkenalkan oleh Cox (1972). Regresi Cox tersebut membutuhkan asumsi hazard proporsional dimana semua individu dianggap memiliki satu nilai *baseline hazard* yang sama lalu kemudian nilai dari *baseline hazard* tersebut berubah sesuai dengan karakteristik pada masing - masing individu. Jika asumsi tersebut sulit untuk terpenuhi, regresi Buckley-James dapat menjadi salah satu alternatif untuk mengestimasi nilai variabel dependen yang mengandung data tersensor.

Regresi Buckley-James dengan metode *Boosting* dapat digunakan untuk mengatasi permasalahan data tersensor dengan jumlah variabel independen sangat banyak dimana pada studi kasus ini bertujuan untuk memprediksi tingkat survival dari seorang pasien penderita kanker limfoma yang telah menjalani kemoterapi dengan melihat ekspresi gen hasil dari *DNA microarray*.

Dari hasil analisis diperoleh Regresi Buckley-James dengan metode L_2 *boosting* menghasilkan 26 ekspresi gen terpilih, sedangkan untuk metode *twin boosting* menghasilkan 10 ekspresi gen terpilih.

Kata kunci : Survival, data tersensor, Buckley-James estimator, *boosting*, seleksi variabel



ABSTRACT

Buckley-James Regression with Boosting for Right-Censored Data

By

Muhammad Rizki Ramadhan

13/347975/PA/15409

Gene expression obtained from the DNA microarray yields a very large number. Which gene expression influences a patient's disease becomes a difficulty. Boosting has been widely used in classification and regression to select independent variables in the modelling. The difficulty increases if the data contains censored data. One method that is often used to handle censored data is Cox regression or commonly known as Cox's proportional hazard model introduced by Cox (1972). The Cox regression requires the assumption of proportional hazard where all individuals are deemed to have the same baseline hazard value then changes according to the characteristics of each individual. If these assumptions are difficult to fulfill, the Buckley-James regression can be an alternative to estimate the value of the dependent variable containing censored data.

Buckley-James regression with the Boosting method can be used to overcome the problem of censored data with a large number of independent variables which in this case study aims to predict the survival rate of a patient with lymphoma cancer who has undergone chemotherapy by looking at the expression of the resulting gene. DNA microarray.

From the analysis, it was obtained that the Buckley-James regression using the L_2 boosting resulted in 26 selected gene expressions, while the twin boosting method resulted in 10 selected gene expressions.

Keyword : Survival, censored data, Buckley-James estimator, Boosting, variable selection