

INTISARI

PENYELESAIAN REGRESI SEMIPARAMETRIK DENGAN MENGGUNAKAN REGRESI RANDOM FOREST

Oleh

ALFINA NUR FIRMANI

12/331386/PA/14644

Analisis regresi merupakan analisis yang sering digunakan untuk mengetahui hubungan yang terjadi antara variabel dependen dan independennya. Namun metode ini masih cenderung terbatas pada asumsi dan keadaan data tertentu, terutama dalam penyelesaian regresi semiparametrik. Oleh karena itu muncul metode regresi yang sangat fleksibel dan mudah, yang merupakan pengembangan dari pohon keputusan. Metode ini dinamakan regresi *Random Forest*. Regresi *Random Forest* merupakan gabungan dari banyak CART yang ditumbuhkan sehingga akurasi yang dihasilkan akan lebih akurat dari pohon tunggal. Studi kasus yang digunakan pada penelitian adalah mengestimasi persentase kemiskinan berdasarkan faktor-faktor yang mempengaruhinya. Dari hasil regresi ini didapatkan hasil bahwa faktor angka melek huruf merupakan faktor utama penyebab kemiskinan. Dari hasil tersebut diusulkan saran penanggulangan kemiskinan dengan memfokuskan pada faktor utama, sehingga penekanan jumlah penduduk miskin menjadi lebih efisien.

Kata kunci : Regresi Semiparametrik, Pohon Keputusan, CART, *Random Forest*, *Variable Importance*

ABSTRACT

SEMIPARAMETRIC REGRESSION SOLVING USING RANDOM FOREST REGRESSION

by

ALFINA NUR FIRMANI

12/331386/PA/14644

Regression analysis is an analysis which often used to know the relation between dependent and independent variable. However this method is still limited on the assumption and condition of a certain data, especially in the semiparametric regression solving. Therefore a very simple and easy regression method was made, which appeared as the improvement of decision tree. This method named Random Forest regression. Random Forest regression is a combination from several CARTs which is built so that the accuracy obtained is more accurate than decision tree. The study case which is used in this experiment is predicting poverty percentage based on several factors that influence it. The result from this regression shows that lackness of reading skill is the major factor causing poverty. From this result, an advice suggested to decrease poverty is focusing at the major factor, so that decreasing the number of poor people become more efficient.

Keyword : Semiparametric regression, Decision tree, CART, Random Forest, Variable Importance