

INTISARI

Bayesian Model Averaging pada Regresi Logistik

Oleh

Muthmainah

13/352635/PA/15684

Salah satu metode statistika yang sering digunakan dalam menganalisis hubungan antara variabel independen dengan variabel respon biner adalah regresi logistik. Model regresi logistik terbaik diperoleh berdasarkan uji signifikansi variabel serta kriteria pemilihan model tertentu. Inferensi terkait prediktor hanya terpusat pada model tunggal yang terpilih untuk prediksi. Hal ini tentu mengabaikan ketidakpastian terkait variabel yang tidak masuk ke dalam model maupun kemungkinan model lain yang dirasa layak untuk dipertimbangkan.

Bayesian Model Averaging memberikan solusi untuk mengatasi ketidakpastian terkait pemilihan model serta variabel, dengan merata-ratakan distribusi posterior dari sekelompok model, yang diboboti oleh probabilitas posterior modelnya. Metode ini melibatkan seluruh model yang mungkin terbentuk dari kombinasi faktor-faktornya. Untuk mempermudah komputasi, model yang dimasukkan ke dalam analisis dipilih berdasarkan algoritma *Occam's Window*. Kontribusi setiap faktor dalam model yang terpilih dapat dilihat melalui probabilitas posterior dari koefisien variabel, sehingga diperoleh interpretasi yang lebih jelas terkait pengaruh faktor terhadap variabel respon yang diteliti.

Studi kasus yang digunakan dalam skripsi ini bertujuan untuk mengetahui faktor yang mempengaruhi warna air sungai di Kabupaten Bantul. Model terbaik yang diperoleh dari regresi logistik akan dibandingkan dengan model hasil analisis dari metode *Bayesian Model Averaging*. Berdasarkan kemampuan prediksi dan persentase akurasi, diperoleh bahwa metode *Bayesian Model Averaging* memberikan model prediksi yang lebih baik daripada regresi logistik.

Kata kunci : *Bayesian Model Averaging*, Regresi logistik, *Occam's window*

ABSTRACT

Bayesian Model Averaging for Logistic Regression

by

Muthmainah

13/352635/PA/15684

One of the most commonly used statistical methods in analyzing the relationship between independent variables and binary response variable is logistic regression. The best logistic regression model is obtained based on the significance variable test and certain model selection criteria. Inference about the predictors is then made from a single model chosen for prediction. This subsequently ignores uncertainty associated with both variables not included into the model and possibility of other worthy considered models.

Bayesian Model Averaging provide solution to overcome uncertainty regarding model and variable selection, by averaging the posterior distribution of a set of models, weighted by their posterior model probabilities. This method involve all possible models which formed from combination of factors. In order to facilitate computation, the models included into the analysis are selected using Occam's Window algorithm. Contribution of each factor in the selected models can be seen through the posterior probability of coefficient of variable, which has a clearer interpretation about factors influencing the studied response variable.

The case study used in this thesis aims to determine factors that affect the color of river water in Bantul regency. The best model obtained from standard logistic regression will be compared with model from Bayesian Model Averaging method. Based on predictive performance and accuracy percentage, it was found that Bayesian Model Averaging method gives better prediction model than standard logistic regression.

Key words : Bayesian Model Averaging, Logistic regression, Occam's window