

***PRINCIPAL COMPONENT LOGISTIC REGRESSION (PCLR) UNTUK
MULTIKOLINEARITAS DALAM REGRESI LOGISTIK BINER***

Oleh
NAYLA AZMI AFIFI
11/316862/PA/13984

INTISARI

Regresi logistik digunakan untuk memprediksi variabel respon yang biner dengan satu set variabel penjelas (prediktor). Estimasi parameter dapat menjadi tidak akurat, serta interpretasi *odd ratio* bisa salah jika terdapat multikolinearitas di antara prediktor. Untuk mengatasi masalah estimasi parameter dengan multikolinearitas, diperlukan pengurangan dimensi prediktor dengan menggunakan komponen utama optimal dari prediktor tersebut. Komponen utama ditentukan dengan menggunakan matriks kovariansi dari data yang telah dilakukan *centering*. Banyaknya komponen utama optimum dipilih berdasarkan beberapa kriteria yaitu *Kaiser's Stopping Rule*, *Scree Test*, dan *Percent of Cumulative Variance*. Pada akhirnya, performa model regresi logistik komponen utama dibandingkan dengan model regresi logistik yang variabel multikolinearnya telah dihapus. Dengan melihat *Bayesian Information Criterion (BIC)*, model *PCLR* dianggap lebih baik daripada model yang variabel multikolinearnya telah dihapus karena memiliki nilai *BIC* yang lebih kecil. Berbeda dengan model yang variabel multikolinearnya telah dihapus, *PCLR* mampu mempertahankan prediktor sehingga tidak terjadi hilangnya informasi yang diperlukan.

Kata kunci: regresi logistik, multikolinearitas, komponen utama

***PRINCIPAL COMPONENT LOGISTIC REGRESSION (PCLR) FOR
MULTICOLLINEARITY IN BINARY LOGISTIC REGRESSION***

**By
NAYLA AZMI AFIFI
11/316862/PA/13984**

ABSTRACT

Logistic regression is used to predict binary response variabel by a set of independent variables (predictors). The parameter estimation can be not too accurate, also the interpretation of odd ratio may be erroneous when there is multicollinearity among the predictors. In order to improve the estimation of the parameters under multicollinearity, we need to reduce the predictor dimension using optimum principal components of the covariates. Principal components are determined using covariance matrix of the centered data. The number of optimum principal components are selected based on several criterias, i.e. Kaiser's Stopping Rule, Scree Test, and Percent of Cumulative Variance. Finally, the principal component logistic regression model performance is compared with the logistic regression model with removed multicollinear predictors. By looking at the Bayesian Information Criterion (BIC), PCLR considered better than logistic regression model with removed multicollinear predictors because it has a smaller BIC. In contrast to logistic regression model with removed multicollinear predictors, PCLR can maintain the predictors so there is no loss of information needed.

Keyword: logistic regression, multicollinearity, principal component