ABSTRCT

Classification of Diphteria Incidence in Bangkalan District with Logistic Regresion, Classification Trees, and Multivariate Adaptive Regression Spline (MARS)

Diphtheria is an endemic disease that widelyspread in many developing countries. South East Asia Region (SEAROi always rank first in most diphtheria cases in the world. Back in 2012, diphtheria was spread in 19 provinces. East Java had the most number compared to other provinces, with 955 cases. In 2013, Bangkalan district came in second after Surabaya city, reaching 76 cases includng 4 mortalities. Thus, diphtheria problem deserves attention in order to predict future incidence of diphtheria in Bangkalan district.There are many risk factors that influence the incidence of diphtheria. In this study, the classification of diphtheria incidence were done using logistic regression, classification trees and MARS to determine the most significant characteristics and factors. After the classification it is found that from all 5 predictor variables, 3 most significant factors were contact, behavior and mobilization.

Logistic regression model showed that the percentage of those without diphtheria (sensitivity) was 91.7% vs those with diphtheria (specificity), 47.2%. Classification trees resulted in sensitivity of 79.2 % and specificity of 69.4%. Meanwhile, MARS model showed 90.3% sensitivity and 55.6 % specificity. From this study, MARS has the most accurate classification with 78.7%.

Keywords : Diphtheria, logistic regression, classification trees, MARS