Predictive models for hotspots occurrence using decision tree algorithms and logistic regression.

Abstract

Predictive models for hotspots (active fires) occurrence are essential to develop as one of activities in forest fires prevention in order to minimize damages because of forest fires. This work applied the decision tree algorithms i.e., ID3 and C4.5, as well as logistic regression on spatial data of forest fires for Rokan Hilir District in Riau Province in Indonesia. The data consist of ten explanatory layers (physical, weather and socio-economic data) and a target layer. Target objects in the target layer are hotspots 2008 and non-hotspot points which were randomly generated near hotspots. As many 561 target objects were prepared through several data preprocessing tasks. The results show that the C4.5 algorithm has better performance than the ID3 algorithm in terms of accuracy and the number of generated rules. The C4.5 decision tree has the accuracy of 65.24% with number of generated rules is 35 and the first test attribute of the tree is peatland type. Furthermore, the logistic regression model outperforms the decision tree algorithms with the accuracy of 68.63%.