

Using Data-mining Techniques for the prediction of the severity of road crashes in Cartagena, Colombia

Holman Ospina-Mateus; Leonardo Augusto Quintana Jiménez; Francisco José López-Valdés; Natalie Morales-Londoño; Katherinne Salas-Navarro

Abstract

Objective: Analyze the road crashes in Cartagena (Colombia) and the factors associated with the collision and severity. The aim is to establish a set of rules for defining countermeasures to improve road safety. **Methods:** Data mining and machine learning techniques were used in 7894 traffic accidents from 2016 to 2017. The severity was determined between low (84%) and high (16%). Five classification algorithms to predict the accident severity were applied with WEKA Software (Waikato Environment for Knowledge Analysis). Including Decision Tree (DT-J48), Rule Induction (PART), Support Vector Machines (SVMs), Naïve Bayes (NB), and Multilayer Perceptron (MLP). The effectiveness of each algorithm was implemented using cross-validation with 10-fold. Decision rules were defined from the results of the different methods. **Results:** The methods applied are consistent and similar in the overall results of precision, accuracy, recall, and area under the ROC curve. **Conclusions:** 12 decision rules were defined based on the methods applied. The rules defined show motorcyclists, cyclists, including pedestrians, as the most vulnerable road users. Men and women motorcyclists between 20–39 years are prone in accidents with high severity. When a motorcycle or cyclist is not involved in the accident, the probable severity is low.

Keywords

Road crashes; Prediction; Data mining; Severity