A Random Forest Classification for Predicting Pedestrians'

Red Light Violation Behavior

Deepak Raj Shah ^{a*}, Rojee Pradhananga ^a

^a Department of Civil Engineering, Institute of Engineering, Pulchowk Campus, Lalitpur, Nepal.

A decision tree is a non-parametric supervised learning algorithm used for both classification and regression problems. Random forest is a widely utilized machine learning algorithm that combines the output of an ensemble of random decision trees to reach a substantial result. Traffic signals allow the sequential separation of vehicular and pedestrian flow. But, red-light violation, a common phenomenon among lower-middle income countries, can cause pedestrians to be exposed to the unwanted risk of a vehicular collision. Existing literature claims Random Forest algorithms are more accurate than conventional prediction algorithms. The current study proposes a robust Random Forest model that accurately predicts the red light violation behavior of pedestrians in Kathmandu Valley. In addition, the model estimates the relative importance of explanatory variables and how they relate to red light violation decision. Data for 1790 Pedestrians were sampled from the video footage. 70% of the dataset was allocated for model training and 30% was allocated for validation. The results reveal perceived volume, remaining red time until green phase initiation, and length of the crosswalk as the most important out of the 9 input parameters, and gender and age as the least important. The Precision, Recall, and F1 score are 0.8648, 0.9437, and 0.9025 respectively, which proves that the developed model is very capable of predicting in unseen datasets as well.