

A COMPARATIVE STUDY OF MULTIPLE MACHINE LEARNING APPROACHES FOR LANDSLIDE PREDICTION

Madawala C.N¹ and Kumara B.T.G.S. ^{1*}

¹Department of Computing & Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Sri Lanka

*cnmadawala@std.appsc.sab.ac.lk

A Landslide is defined as a collapse of a mass of earth. Irregular development activities on mountains and inadequate attention to construction aspects have led to the increase of landslide and consequently sustaining damages to lives and properties. According to the National Research Building Organization (NBRO) reports, within the study area, nearly 3275 sq.km of the area expanded over the Ratnapura District; and 2178 sq.km area is to be highly prone to land sliding. If a proper investigation were performed in time, most of the landslides could be predicted relatively and accurately. The main objective behind this study is to landslide-hazard mapping to discover the real scope and austerity of landslide processes, such that knowledge will produce the extreme benefit to government officials, and the general public in avoiding the landslide hazards and mitigating the losses. Initially, the Machine learning algorithms such as Support Vector Machine (SVM), Naïve Bayes, Decision Tree (DT) and Random Forest algorithms were used to develop the landslide prediction model. Also, execute the Ensemble Learning techniques based on Bagging, Boosting and Stacking to develop the landslide prediction model. Then, both modelling results have compared and finally, investigate the most appropriate prediction model. This study has a strong capability to predict landslides by considering triggering factor; rainfall and causative factor; slope angle, land cover, elevation, intensity, Bedrock geology, and Soil materials. While applying the Ensemble approach; Voting has obtained a 98% accuracy rate, Bagging (74%), Boosting (AdaBoostM1) 94% and Stacking (StackingC) 76% of accuracy rates obtained respectively. Moreover, the experimental results after applying Machine Learning Algorithms; SVM gained 96% accuracy rate, Naive Bayes 94% accuracy rate, DT 89% accuracy rate, and Random Forest gained 84% accuracy rate. As the empirical results of this study researcher concluded that Ensemble Learning techniques have achieved the highest accuracy over other approaches therefore, Novel Ensemble Approach has the best degree to fit for building a landslide prediction model. Moreover, an improvement of the hazard monitoring, accuracy of early warning and disaster mitigation are performed.

Keywords: *Ensemble approach, Support Vector Machine (SVM), Naïve bayes, Decision Tree (DT), Random forest*