

**ENSEMBLE LEARNING APPROACH FOR YOUTUBE VIDEO
CLASSIFICATION BASED ON THEIR VIDEO CONTENT**

N.H.T.M. De Siva and R.A.H.M. Rupasingha*

*Department of Economics and Statistics, Faculty of Social Sciences and Languages,
Sabaragamuwa University of Sri Lanka, Sri Lanka.*

**hmrupasingha@gmail.com*

Many individuals use the platform YouTube to share videos with others worldwide. Different content and quality videos are available on the YouTube platform, and it would be beneficial to users if they could assess the content of the video before selecting a video to view. However, deciding and making predictions directly is difficult since different attributes need to be considered. This study is carried out to classify videos using six individual distinct machine learning algorithms and an ensemble learning algorithm as a solution to the above. The algorithms were selected based on the literature review. The gathered data set underwent essential pre-processing and attribute ranking. View count, like count, comment count, number of subscribers, tag count, and total views were identified as the main contributors to the study. For classification, Naive Bayes, Logistic Regression, Support Vector Machine (SVM), Decision Tree, Multilayer Perception (MLP), Random Forest individual algorithms, and Ensemble Learning algorithm that combined five individual algorithms were selected since those algorithms work effectively and efficiently with better results. Among the ensemble learning algorithm techniques, majority voting exhibits the best accuracy. For categorisation, the result was evaluated by 60% training data and 40% testing data. In each method, different parameters were changed for the evaluation and accuracy, recall, f-measure, precision, Root Mean Square Error, and Mean Absolute Error were taken into account. Random Forest demonstrated 96.89% accuracy, and Ensemble Learning showed 97.21% better accuracy than others. These findings indicate that this strategy is appropriate for YouTube content classification. It offers the information required to evaluate video content and make predictions about particular videos. Future research would focus on applying more deep learning algorithms to improve accuracy.

Keywords: Classification, Machine Learning, Video Content, YouTube