

Use of LIME for Human interpretability in Sinhala document classification

P. K. S. Kumari

*Faculty of Graduate Studies and Research
Sri Lanka Institute of Information Technology,
Sri Lanka
suriyaa.k@sliit.lk*

P. S. Haddela

*Department of Information Technology
Sri Lanka Institute of Information Technology,
Sri Lanka
prasanna.s@sliit.lk*

Abstract

With advancement of technology in Sri Lanka, use of Sinhala text usage has grown rapidly over the time where automatic categorization is helpful for efficient content management. As a result, experts tend to use machine learning application to categorize this large volume of data in an efficient and accurate manner. Most of these learning models are operating in a black-box where there is no way to understand how the model has decided which category an instance is assigned. Understanding the reason behind why learning model makes these predictions is very important to trust such models and to provide reasonable justifications in real world application. Intention of this research is to present the work carried on related to document classification model prediction interpretation where a set of text classifiers has been studied with use of SinNG5, freely available Sinhala Document corpus.

Keywords: *Machine learning, Sinhala text, document classification, human interpretability*