An Ensemble Approach to Detect Hate Speech in Tamil Tweets

F.H.A. Shibly[1], Uzzal Sharma[2] & H.M.M. Naleer[3]

[1]*Dept. Arabic Language, South Eastern University of Sri Lanka, Oluvil, Sri Lanka*

[2]*Department of Computer Applications, Assam Don Bosco University, Guwahati, India*

[3]*Department of Mathematical Sciences, South Eastern University of Sri Lanka, Sammanthurai, Sri Lanka*

[1]*shiblyfh@seu.ac.lk*, [2]*uzzal.sharma@dbuniversity.ac.in*, [3]*drnaleer@seu.ac.lk*

ABSTRACT

People have converged on a worldwide level because of advancements in communication technologies. They are critical in ensuring freedom of speech by allowing individuals to express their thoughts, behaviours, and opinions openly. Although this presents an excellent chance for racism, trolling, and exposure to a flood of offensive online content. As a result, the exponential growth of hate speech on social media significantly impacts society. In this research, we applied machine learning and deep learning algorithms to detect hate speech and compared the performances of those algorithms to develop an ensemble model. Researchers collected and combined two Tamil languages hate speech tweets datasets created by Bharathi Raja Chakravarthi et al. Tweets in this dataset are classified into two categories: not offensive and offensive. This dataset contains 10,129 tweets. Also, researchers selected six machine and deep learning algorithms for this study. Support Vector Machine (SVM), Logistic Regression (LR), Naïve Bayes (NB), Bidirectional LSTM, Multi-layer Perceptron (MLP) and Multilingual BERT were applied. Regarding detecting hate speech, SVM (82%) and LR (82%) have the best Accuracy. Furthermore, researchers developed two ensemble algorithms to construct the most efficient model. The first ensemble model was created by combining SVM, LR and NB and the second ensemble was developed using SVM and LR. Four algorithms, including the two ensemble models, obtained the same Accuracy. Therefore, the researchers compared the F1 score and found that the ensemble model 02 outperformed other classifiers. The findings of this research study are essential because these findings can be utilized as a model study for Tamil language hate speech to evaluate future research works using different machine learning algorithms for detecting hate speech more accurately and efficiently.

Keywords*: Machine Learning, Deep Learning, Algorithms, Hate Speech, Detection and Ensemble Model.*

The Manuscript has been withdrawn from the proceedings according to the authors' request

ISBN: 978-624-5736-55-3