Abstract No: SO-25 **Software Intensive Systems**

Detection of cyber bullying on social media networks

Suchini Priyangika* and Shantha Jayalal

Department of Industrial Management, Faculty of Science,
University of Kelaniya, Sri Lanka
suchinipr92@gmail.com

Social Media is becoming an integral part of people's daily lives today. It is an effective way of sharing one's life experiences, special occasions, achievements and other events with their friends and family. Although it is a fruitful way to communicate with groups, some people find themselves being insulted or offended by others who are involved in certain post or conversations. These insulations can be based on racism, using profanity or any other vulgar or lewd language. This cyber bullying needs to be monitored and controlled by the social media site owners since it will highly effect on the number and safety of the active site membership. Currently, there is no automated process of identifying offensive comments by the social network site itself. It can be only diagnosed by humans after reading the comments, flagging or reporting them to the owner of the site or blocking the offender. Considering the massive big data set generated in social media daily, automatically detection of offensive statements is required to reduce insulation effectively. For this purpose, text classification approach can be applied where a given text will be categorized as insulting or not, through learning from a pre-learned model.

In order to develop the model, data was collected from the popular data repository site named www.kaggle.com. The dataset consists of comments posted on Facebook and Twitter. Firstly the dataset was divided into training data set and test data set. Then the collected data was preprocessed by removing the unwanted strings, correcting words and eliminating duplicate data fields. In the next step, features or keywords were extracted which are qualified to distinguish a statement as 'insulting' using N-grams model and counting methods. Feature selection is done using *Chi-Squared* test and finally apply classification algorithms for separating insulting comments and non-insulting comments from a dataset given. Machine learning algorithms such as *Support Vector Machines (SVM)*, *Naïve Bayes*, *Logistic Regression* and *Random Forest* are used for this. Out of the classification algorithms, *SVM* is to be performed better than other algorithms since this is a two-class classification problem and a comment is to be classified only into two separate classes which are 'insulting' and 'neutral'. With an exact separation of a given comment into 'insulting' and 'neutral' category, cyberbullying happening through offensive comments posted on social media sites can be detected.

Keywords: Social media, Big data, Feature extraction, Machine learning, Classification algorithm