Measurement of Toxicity Levels in Sentences using NLP and CNN with Word Embedding System

Mahmudul Hasan Shakil
Dept. of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
mahmudul.hasan.shakil@g.bracu.ac.bd

Zawad Alam
Dept. of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
zawad.alam@g.bracu.ac.bd

Sazal Kanti Kundu
Dept. of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
sazal.kanti.kundu@g.bracu.ac.bd

Arnob Kumar Dey
Dept. of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
arnob.kumar.dey@g.bracu.ac.bd

Shihab Sharar
Dept. of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
shihab.sharar@g.bracu.ac.bd

Annajiat Alim Rasel
Dept. of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
annajiat@bracu.ac.bd

***Abstract*— We are living in an era of technology. The rapid growth of information technology and the disruptive transformation of social media has happened in recent years. Websites like Facebook, Twitter, Instagram, where people can express their thoughts or feelings by posting text, photos, or videos, have become incredibly popular. But it has also a bad reputation for spreading toxicity. There are many existing works in this field but those are not fully successful yet to provide accuracy at a satisfactory level. A group technique of convolution neural network (CNN) and natural language processing (NLP) can be utilized which portions poisonous and non-poisonous words. Then it can be classified into different categories like toxic, severe_toxic, insult, etc. This architecture can be organized around data processing, using data cleaning processes and NLP techniques such as tokenization, stemming, and word embedding can be used to convert a word into a vector technique. And finally, CNN classifier may label each sentence into different toxicity levels.**

Keywords*— *Toxic comment classification, NLP, word embedding, fast text, tokenization, CNN