Sentimental Analysis from Text Messages of Social Media

Farhana Akbar
Electrical & Computer Engineering
North South University
Dhaka, Bangladesh
farhana.akbar@northsouth.edu

Shamlima Deb Trena
Electrical & Computer Engineering
North South University
Dhaka, Bangladesh
shamlima.trena@northsouth.edu

Md. Khairul Alam
Eletrical & Computer Engineering
North South University
Dhaka, Bangladesh
munna.khairul@northsouth.edu

I. Introduction:

Social media has become super popular, with 5.35 billion people using it worldwide as of January 2024. Facebook, YouTube, and WhatsApp are all about connecting people, but with so many users, keeping things safe is tough. Cyberbullying and hate speech are big problems, affecting millions of users around the world. In Bangladesh, where 52.9 million people use Facebook, more than 60The issue is that tools to stop this stuff work well for languages like English, but not so much for languages like Bangla. There just aren't enough resources to check and understand what's being said online. Without good tools for languages like Bangla, it's hard to stop cyberbullying, hate speech, and false information. It lacks necessary computational linguistic resources like corpus, language models, and powerful machine learning methods for performing various NLP tasks.

II. Literature Review:

In this part, we review some related materials that were relevant to our research. For hate speech on Bangla comments, Khan, proposed an SVM based model where they used TF-IDF to process data and multiple classification models to yield the accuracy, proposed a model based on encoder-decoder. Comments were categorized into 7 distinct categories of speech. All contents were divided into hateful and non-hateful categories. To extract and encode features from Bangla comments 1D convolutional layers were used. And at last, to predict hate speech the attention mechanism LSTM and GRU(Gated Recurrent Unit) - based decoders were used. Romim, used many deep learning models which performed well. The approach they have conducted is a baseline experiment. Several deep learning models along with a lot of Bangla words(pre-trained) embedding such as FastText, Word2Vector, and BengFastText were used on this dataset. They have used 30k comments tagged by crowdsourc- ing where 10k is hate comments. Das Bandyopadhyay proposed an opinion polarity classification. The dataset was built from news text of Bangla sources and the authors used SVM to predict the accuracy Nabi, used TF-IDF feature extraction on Bangla text to generate results. We further review Countering online hate speech analyzing the dataset. It is a compiled a dataset with 6,898 counterspeech comments and 7,026 non-counterspeech comments. The psycholinguistic effects of counterspeech and non- counterspeech. The targeted communities for the dataset were Jews, African-Americans and LGBT communities.