Social and Information Networks(CSE3021)

J Component-Review 1

Faculty: Dr. Parvathi R,

Slot: A1

Team Members

Ninad Sabnis(18BCE1211)

Mehul Raj(18BCE1146)

Abstract

A lot of people today receive abusive of threatening messages online from people who exploit the relative anonymity of social media to spread hatred and hurt between communities.In a civilized online conversation, most of us have encountered a ‘troll’ who just for the fun of it likes to disrupt a mature discussion with childish comments which can be disrespectful or hurtful to the individuals who are talking leading one of them to leave the forum out of frustration.

This project deals with identifying such ‘trolls’ or unwanted accounts spreading hateful messages online so that further action can then be taken on these individuals as deemed fit which may even include reporting their details to the law enforcement officials. To accomplish this task we plan to use a few machine learning(ML) and natural language processing(NLP) algorithms like K-means clustering and logistic regression classifier.

Dataset

- The dataset that we have chosen consists of two csv files named test and train.
- The train dataset contains 15276 rows while the test dataset contains 5057 rows.
- The train dataset contains three columns:
  - The tweet id
  - The text in the tweet
  - The classification values it is given.

- The test dataset on the other hand contains only the first two columns with third column being predicted by our ML and NLP algorithms.

Pre-Processing

- Stopwords
- Word_tokenize
- CountVectorizer
- Tokenizer

Classes

- Toxic
- Severe toxic
- Obscene
- Threat
- Insult
- Identity hate

Models

- MultinomialNB
- SGDClassifier
- LogisticRegression

Further Work:

2D CNN

Comparison of Models

- Accuracy_score
- RMSE(Root Mean Square Error)

These will be used to find the deviation between expected values and obtained values and compare between models.