Comparative Survay New

mohamed1504002

December 2018

1 Abstract

Cyberbullying is an emerging problem that faces a lot of people on using the social media. there were many proposed systems to automatically detect and block the cyberbullying but every model had a lot of problems that we are going to try to solve it in our proposed model

2 Proposed Model

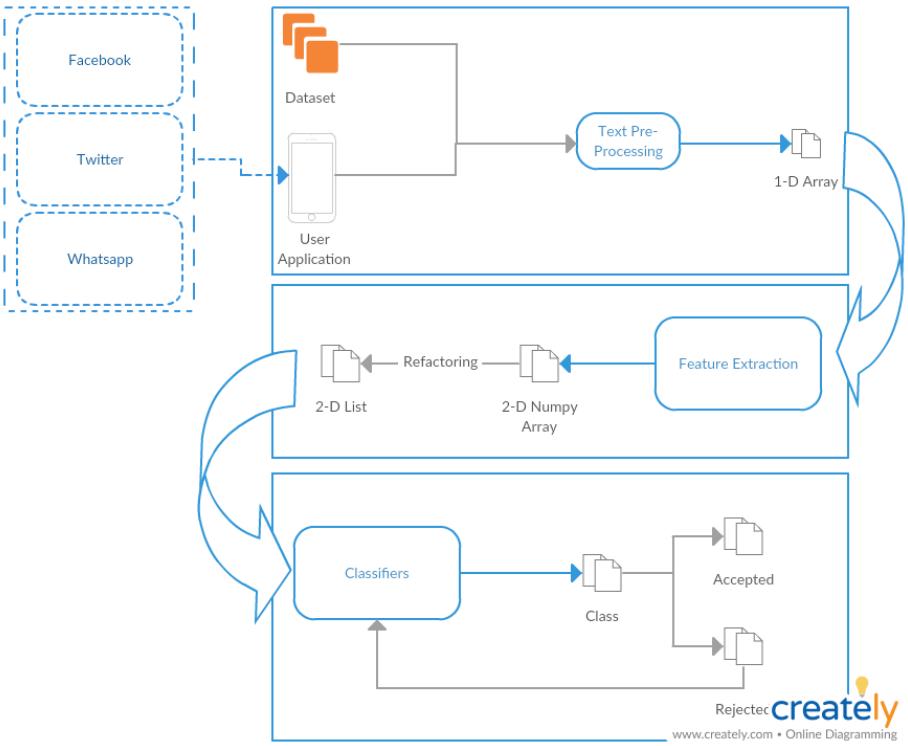

Figure 1: System Overview

Our Proposed system consists of three main stages which are Pre-Pr Feature Extraction and Classification these three stages arocessing of words ,e ttext he maentersi in order nclassified as bullying text or not besides or proposed model has a self learning feature which enables the trained model to train on the error it has made in the classification based on the user rating and the approval of the system administrator

2.1 Pre-Processing

In this stage of the system we are making the text ready for the feature extraction stage by tokenizing the text,applying stemming and lemmitization , also we used Bing API for word correction, and we removed the encoding parts in the text.

2.2 Feature Extraction

In this stage we extract the Features from the text that we are going to push it in the model for the training and predicition for the feature extraction we are going to use three method which are TF-IDF , Linguistic Inquiry and finally the sentiment analysis. we are going to use the sentiment analysis to eliminate the false positive problem which faced many previous systems, the sentiment analysis depend on the context of the chat it get the polarity of the text overall not every sentence. and for the TF-IDF it gets the wights of the words in a certain text

2.3 Classification

This stage is the actual stage of the classification to either bullying or not and for this stage we are going to use three different models that we are going to train with different data. we are going to use logistic regression and Random Forrest and Naive Bayes. These three classifiers are going to be trained with different and for every classification process we are going to make the features enters in the three classifiers and make every classifier make its own prediction and then we are going to make a voting between the result and the voting isn't going to make all the classifiers has equal votes we are going to make the classifier with better accuracy has more voting points than the classifier that has less accuracy.

2.4 Self Learning

For this stage it is only stage to make the model self maintained which means that we receive rating from every user about every classification process and according to this rating the administrator asses this rating and either approve it so the classifier is going to train on the same data with the correct class of reject this rating and in this case the rating is dropped and never reach the classifier for training.