

### SI 650 Homework3

Muhan Yuan

UMID: 3159 1807

Display Name: MuhanYuan

It takes three steps to conduct the personal attack detection:

First step is text preprocessing. Since in most cases, punctuation doesn't contain any insulting message, I deleted all punctuation from the raw comments. But after taking closer look at the data, I noticed that exclamation mark sometime indicates strong emotions, which could have some relation with personal attack, so I choose to put it back. Also, I referred a "swear words" corpus and manually adding weight to these words in comments by repeating these words.

Next thing is vectorizing the text. I used Vectorized method in sklearn package and used Tfidftransformer to make the transformation to improve the model's performance. I tried different values of the parameter "ngram" and in most cases, "(1,2)" has the best performance, which tells us using both one word and two words next to each other as feature is most informative but also doesn't lead to overfitting.

I chose LinearSVC to perform the classification, because it has advantage over other methods on the operation efficiency and performance. In this project, there are significantly more comments in class 0 (no attack), so I need to balance the class weight (Setting the class\_weight= 'balanced'). In most cases, when parameter "C" equals 10, the result is most accurate, which means we should have a moderate penalty on the error terms.

The final Score is 0.76196.