

Project Proposal

1) Problem: What is your problem? What is your motivation behind choosing this problem for your project?

We are planning to build a classifier to predict toxic comments. The motivation behind it is the multitude of online forums, where-in people participate actively and make comments. Comments can be derogatory, abusing and they should be managed accordingly for well-being of all user. Blocking or reporting can be further action after classifying.

2) Methodology: What methods do you plan to design or employ to solve the problem?

We will be creating few visualization to understand data and check whether there are sufficient data for each classification and analyse influence of features on data. Then we will do pre-processing of data like dropping neutral words and then tokenizing the sentence. Then we plan to do cross validation on different models (**SVM, LogisticRegression**) after hyper-parameter tuning on training data and then check them on testing data.

3) Final product: What will be the outcome of this project? How will you measure the success of your course project? Will this project help you explore or learn something new?

It would help us learn about semantic analysis and how different tricks in pre-processing of data can contribute to a decrease in error rate. We would be exploring new ways to optimize accuracy. We would learn to solve a real world problem of multi-label classification, here one comment can be linked to different labels. Implementing different classifier would give us insight which type of model is suited for semantic analysis classification and their behaviour on our dataset.

NOTE: Professor, we have started working on machine learning through this course only. It is something that interests us, and we would like your valuable inputs and suggestions to improve this project and learn more. Thank you.