SINCERITY IN QUORA

Abstract

Quora is an American question and answer website where questions are asked and edited by the Internet users, in the form of opinions, which include questions that are harmful to our community, and should be removed before they have a chance to cause harm(or spawn imitators).

The entire project would be consisting of predicting sincerity of questions on Quora. The dataset concerned here, consists of over 1.3 million questions, having the opportunity to train and test models, to detect insincerity and trolling based on real Quora questions. In conjunction with dataset, the project would be focused on solving this problem. We have used machine learning involving classification for the same. The entire project duration consists of training the available dataset with appropriate algorithms, to develop a model, and then testing the model with appropriate test dataset.

The aim of the analysis is to provide a safe and legit content website for Quora's 300 million monthly users. Count Vectorizer model was developed for each question, which was then fitted to Logistic Regression model, which classified the unseen questions with about 91 % accuracy.

*Keywords:* Quora, Insincerity , trolling, prediction, machine learning

RAGHAV AGARWAAL

RUCHA PATTNAIK

PREETIKA PATRA

SIMRAN ROUT

Under the supervision of :-Parijat Roy