**Quora Insincere Questions Classification**

***Problem Statement:*** *“To develop a system that can identify insincere questions present on Quora by developing models to detect toxic and misleading content”*

Quora has 190 million unique visitors and over 100 million users visit Quora every month. It caters to different groups of people around the world belonging to different races, castes and age groups. The questions available on Quora are of different types: questions seeking factual answers, hypothetical questions, those with experiences of people, questions with opinions or surveys on certain topics and insincere questions (rants, trolls, and biased opinions).

It is extremely important to detect these insincere questions and handle them appropriately to avoid hurting the sentiments of any of the diverse groups of users on Quora. Formally, an insincere question is one that is intended to make a statement rather than look for helpful answers and has any of the following characteristics: a non-neutral tone, disparaging or inflammatory content, scenarios or circumstances that are not grounded in reality or one that uses sexual content for shock value and not to seek genuine answers.

Detection of these insincere questions is a very difficult task. The first reason for this being such a herculean problem is that the insincerity of a question is subjective and may change based on a user’s perspective. Secondly, most insincere questions might look quite similar to the sincere ones. It is therefore difficult to distinguish between the two. Also, understanding the intent of a given piece of text is a benchmark task in NLP because it is difficult. Due to these reasons, the classification of insincere questions on Quora is a popular Kaggle challenge.

The dataset provided by Kaggle has 1.3 million rows in the training dataset and approximately 400*k* rows in the testing set. Each row has three fields: a unique question identifier, question text, and a target value. The target value is 0 in case of sincere questions and 1 for insincere questions.

After reviewing 25 different solutions that were attempted in the Kaggle challenge, we have listed out several approaches that can be followed to solve this problem. The approaches that we plan to follow in our solution would comprise the following modules:

**Data Preprocessing:**  spell check, removal of stop words, case conversion, stemming and lemmatization punctuation removal

**Word Embeddings:** word2vec, glove, elmo, bert, wiki, GoogleNews, paragram, a combination of more than one of these

**Dimension Reduction:** Max pooling, spatial dropout, PCA, projection meta embedding, random sampling

**Core Architecture:** Vanilla RNN, use of single bi-LSTM, convolution layers, bi-GRU, FC Net

**Model Optimisation:** Hyperparameter tuning, Cross-validation, batch normalization, dropout

We would be choosing our initial approach based on theoretical proof. After its successful completion, we would be trying out different modules to take empirical results into consideration. Finally, our evaluation would be performed on the test dataset using F₁ score which is a well-balanced measure. It considers both the precision ***p*** and the recall ***r*** to compute the score and is thus a very good measure.