

CS 613 NLP: Assignment 3

This assignment is based on an ongoing competition, [Sentimix: A code-mixed sentiment analysis competition at SemEval 2020](#)

Deadline: 11:59 PM on November 17, 2019. No late submission is allowed.

Tasks

1. Download the dataset from [here](#).
2. Train data has 15131 tweets. Test data has 1869 tweets. Format of the data is:
meta uid sentiment
token lang_id
token lang_id
.....
.....
 - *uid* is a unique id for each tweet.
 - *lang_id* is 'HIN' if the token is in Hindi, 'ENG' if the token is in English, and 'O' if the *token* is in neither of the languages.
 - The *sentiment* is either positive, negative or neutral.
3. Build a classifier that can classify the ENG-HIN code-mixed tweets based on their sentiments. You are free to use any technique. Neural network architectures are preferable.
4. Some good references for code-mixed sentiment analysis: [1](#), [2](#), [3](#)

Submission

Submit a pdf file in the Goole classroom with the following details:

1. Link to GitHub repository
2. Description of the model used in 130-150 words. You can add an image describing the architecture.
3. Report the performance of your system using the following metrics:
 - Accuracy
 - Precision across all three classes.
 - Recall across all three classes.
 - F1-score across all three classes

Evaluation

1. The novelty in the model proposed. [5 marks]
2. Performance of proposed architecture. [5 marks]