Progress Report for Tweet Classification Competition

About the project: Our team is working on classifying the tweets in sarcasm and non-sarcasm categories. We intend to use deep learning algorithms in natural language processing to classify the tweets and get highest level of accuracy, thus, not only beating the baseline but will also aim to secure the top rank in the competition.

Progress made so far: We've analyzed the problem carefully. We looked into multiple deep learning algorithms which can be used to solve our problem. We also looked into multiple transformers based deep learning algorithms too, to solve our problem.

We did lot of data cleaning in the input file, so as to feed the right input to the classification algorithm. We used Python's Pandas, Reg Ex on Anaconda's Jypyter notebook for data cleaning.

Based on the algorithms, which we tried so far, we've been able to solve the tweet classification problem and beat the baseline. We are currently ranked #16 in the leaderboard. We'll however try to improve on our rank.

We've used Google's T5 based classification transformer to receive best result so far on the classification problem.

On the Leaderboard in Livelab, our submission can be found with the id: pdwivedi08

Remaining Tasks: We'll continue to try to improve our ranking in the competition. Additionally, we've following tasks left:

1. Detailed documentation of the project
2. Presentation on the project execution
3. Code cleaning and comments update

Challenges: We don't have any major challenge at this time in the project completion. We did face some challenge earlier to decide the right algorithm to use and our research and self-study did help in that.