**Week 4**

Using WordVec+LSTM model I have first results with labeled by team almost 7000 tweets. The notebook of model is here:

<https://colab.research.google.com/drive/1fwCWig1ZOgoT59ov3fl6gJzoPEydblTr#scrollTo=jjkyf9D2YuhO>

[**Samir Sheriff**](https://app.slack.com/team/UM92VBKGU) [12:07 PM](https://omdenavoice4impact.slack.com/archives/CMF4NMRBN/p1568011057089600)

I have created a preliminary pipeline to use transfer learning with the ULMFit model based on a pre-trained AWD\_LSTM backbone for sentiment analysis, using the labeled data from [#task4\_labeling](https://omdenavoice4impact.slack.com/archives/CMJMB69D5). I haven't played around with all hyperparameters yet, so the results aren't good (I'm guessing mostly due to the skewed nature of the dataset). For now, I'm sharing the notebook here if anyone is knows more about this model and wants to lend a hand. <https://colab.research.google.com/drive/1OcK2AWQBwfs-FmtiNnboGsvWUfFNQQIj>

<https://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/>

Basically I detected some communities using Girvann Newmann Algorithm, calculated the average pagerank of that community and the number of signal words that is associated with all the influential people in the community and tried to make some sense out of the data. However, the program might get exhausted due to limitation of colab GPU if we use large datasets. Right now, The bottom line is- if we can extend this particular for the datasets we are labeling or some other datasets where we have a knowledge that certainly we have some gang members in it, then we can genuinely test the credibility of this analysis. The motivation of this work was the basic concept that was discussed in the paper I shared the other day. Some inputs and suggestions would be really appreciated since there have been quite a low amount of activity since last few days. Here is the link to the colab. Lots of codes and ideas have been shared in the colab and I think this can be a very good starting point for those who are wondering where to start and how to cooperate. Just go through this colab and you will find the contents. Thank you and looking forward for some suggestions and collaboration. Thanks in advance. <https://colab.research.google.com/drive/1boFlLhO8APrJ3KeeUPyW6kwjVOlS9RX_>