ASSIGNMENT 3 [CSE 574]

1. **Rajiv Nagesh – UBIT: rajivnag**
2. **Ishansh Sahni – UBIT: ishanshm**
3. **Shreya Joshi – UBIT: shreyajo**

Part 1: Performing Manual Annotation

1.2.1 Based on what we discussed in class about this dataset and the task of stance annotation (here, for attitudes towards vaccines), on what percent of the tweets that you annotate do you think the two annotators will agree? There is no right answer here, obviously, but provide a justification of your response.

Well, I think that the percent of agreement among the annotators of our team could be around 60-65%. I am setting this high number as there is a similar bias in us. This bias includes regional, political, and societal emphasis. Three of us come from the same part of India and our country was high on the vaccination drive which has tuned is in that thinking where if any individual is taking the vaccine is kind of a person who is inclined to be a pro-vaccine if not literally a pro-vaccine individual. Furthermore, we have been nurtured from social and political media that we always should be vaccinated and looking at the tweets in the dataset provided to us, this bias that prevails in us would tend people being vaccinated a pro-vaccine.

1.2.2 Based on what we discussed in class about this dataset, what percent of the tweets that you annotate do you think will be labeled *pro-vaccine*? Justify your answer.

This answer would be similar for us as we answered in 1.2.1. The regional and political bias would play a picture in answering this. But to answer the question, probably 30% of the tweets provided in the dataset could be labeled as pro-vaccine as there are other factors as neutral, anti-vaccine and not providing any information related to vaccine as well. So, 30% of entire tweets being labeled as pro-vaccine is quite high. Consequently, terms like COVID, Vaccination, etc. on twitter were fiddled by those who were mostly on the learned or well-read individuals and, the latter would also opt to be vaccinated and be a pro-vaccine which is the right thing to do. Given the justification for this answer you could also infer that I am pro-vaccine.

Part 3: Reflecting

2.1 What was your overall percent agreement, and how did this compare to your estimate from 1.2.1? Why were these similar/different?

The computed percentage agreement after the manual annotations and calculating the Krippendorf's Aplha came to 49.33%. The earlier estimated percentage agreement was around 60-65%. We are short of 15% which is not bad as compared to the results. We overestimated the agreement keeping bias in mind but a machine learning model with its computation didn't work in the similar manner. Though it is a lower percentage agreement, we are keeping in mind of the manual annotation and going ahead with the value because 3 annotators are separate individuals with various thoughts on aspects and societal science. Just found some statistics on percentage agreement which I thought would be insightful to site here-

If it's a sports competition, you might accept a 60% rater agreement to decide a winner. However, if you're looking at data from cancer specialists deciding on a course of treatment, you'll want a much higher agreement — above 90%. In general, ***above 75%*** *is considered acceptable for most fields.*

2.2 Was the annotation harder or easier than you expected? Why, and what was the hardest part?

The manual annotation was a times taking task as we had to think in the shoes of the person who tweeted and not just in our own perspective. The annotation also included headlines of articles and news headlines which made it tougher to be conclusive on whether the person who tweeted that was pro-vaccine or just stating a general news headline for the sake of it. Such aspects differ annotator to annotator as they annotate and perceive the information thereby giving each tweet one of the accepted annotations. There was a time when our team faced a dilemma on a particular tweet which was a headline but had a mix of both positive and negative, inferring an annotation for such open-ended tweets were tough. Furthermore, some of the tweets were not in English and deciphering the translated version of it was tough as that did not give the actual connotation of the tweet. Discrepancy of such key information also matters in a machine learning model. Conclusively, we can say that the annotation was easy individually but to arrive on the final annotation of a particular tweet after discussion with your group mates was a relatively tougher task than expected contradicting to our earlier answer in 1.2.1 which stated that we would agree more because of the political and regional background we come from.

2.3 Does this change your perspective on how people talk about health-related content on social media? Why or why not?

I don’t think this change my perspective on how people talk on social media as that is up to them. Some of them take a stance and some are just oblivious about the whole scenario. There were tweets which were just vague and derogatory which also complemented the health-related issue of the vaccine. People’s take on an issue on social media cannot be used as a metric to take a stance or perspective about a world-wide cause. For instance, I cannot decide COVID is a hoax just because some person from Texas tweeted that COVID is fake and an international scam. Would I rely on the person from Twitter or an organization like WHO, who work day in and out for fighting against COVID even to this date? So, no, how people talk on social media would sure have a say on my perspective by a little bit but sure won’t change my whole idea about a cause.

2.4 What did you learn from this annotation task that will change your perceptions of how NLP models are trained moving forward?

We learned that NLP models thrive on agreement, we (group) should be more aware about a case and study it to its depth before performing any manual annotation. Because there have been cases where inter-group dilemma causes a fiasco. NLP models would work the way the annotators annotate a text of information. We should keep in mind the connotation of the text and be neutral towards it without a prior bias which is relatively tough but should be kept in mind moving forward.

2.5 How do you think your team's agreement (say, in terms of Krippendorf's Alpha) will compare to the other teams? Why do you think so?

This again is a hypothesis of ours, but we believe that our percentage agreement coming from an Indian background would be more or less similar to those who come from the same part i.e., India. As stated in 1.2.1, the regional bias would prevail in most of the peers having an Indian group. For example, not considering ML bias and other bias if two individuals from different continents of the world were picked to annotate a single tweet, there are chances they would probably they would annotate them differently. This is the bias they have in mind about COVID and the prevalent information they are inculcating with whatever various forms of intel. Regional bias plays a big role in this assignment for the annotations thereby resulting the same in the percentage agreement. The difference between our group and say a group from the US would not be much but there will be a significant difference in the annotations and the reasoning behind each annotation.