**Meeting 26/11/20 Notes**

* Literature review – need to add tables and figures e.g. table for classifiers with classifier, references, dataset used on, accuracy/other metrics if comparable. Same for features – could have separate tables depending on how related. Need to include research questions and deliverables (basic, intermediate, advanced); in conclusion - hint at how to move forward; summarize current limitations and gaps (and how aims related to literature limitations) – limitations section works as a reminder for what needs expanding later on. Should extend literature review to resemble mini version of final paper; can be > 6 pages, take time on it
* Classifier – balance into bins for each score – 5 toxic scores per gender so 10 sets (do for entire dataset – women may lack confidence and give more neutral scores), remove ‘other’ from dataset, handle 0 as separate neutral case, combine [-2,-1], [0], [1,2] into classes – solve as 3 class classification (should be 50:50 if unbiased), get distribution of true negative and true positive results (does it get extremes right? is it normally distributed?), look at 10 bins for anything useful e.g. female in neutral bin all right? – see if can predict any bins accurately for male/female; can experiment with different models; use all female data – match number in other bins to smallest (female toxic)
* Can build 1-class classifier to compare results – drop the toxicity label
* Look at detoxify link sent and think of how to apply/incorporate – if state-of-the-art then compare against it