Detecting emotions in conversations with Eliza

Tessel Haagen – 2825310

## Section 1: Methods

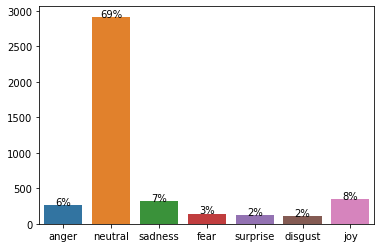
In this experiment, we compare two machine learning models for detecting emotions in conversations with Eliza. Eliza is a chatbot designed as a psychotherapist by replying to human utterances based on basic word cues in the input. For this experiment, we used the Ekman emotions: anger, sadness, fear, surprise, disgust, and joy. Furthermore, we have added a seventh emotion neutral to this set. Students had a fake conversation with Eliza and annotated their utterances with the set of emotions. In total 44 different students participated, which resulted in 2129 test instances, excluding the utterances of Eliza. The distribution of the test data is visible in Figure 1, which shows highly biased data towards the seventh emotion neutral. Due to this bias, machine learning models may exhibit reduced recall. When the training data is more balanced than the test data, the models tend to underpredict 'neutral' instances, resulting in increased false negatives. Moreover, this bias will impact precision as other emotional categories are more frequently predicted, leading to higher false positives. With decreased recall and precision, the machine learning models are likely to demonstrate lower F1-scores.

Figure : Distribuation of the test data

The first model is a fine-tuned BERT sentiment analysis pipeline on GO emotions. The GO emotions are more nuanced as it consist of 27 emotion classes. In order to address this, we incorporated additional steps for post-processing. After applying the fine-tuned BERT model to our test instances, we map the scores of the GO-emotion labels to the Ekman emotion labels, resulting in a list of scores. To get the final scores for the Ekman emotion labels, we averaged these scores. We use a threshold to decide if a score can count for the total Ekman emotion score. This threshold is set to 0.04 since randomized scores for the GO emotion will lead approximately to a score of 0.04. Lastly the

## Section 2: Ekman classification

## Section 3: Comparison of GO-classifier and SVM

## Section 4: Improvements of the SVM

## Appendix A: Distribution per speaker

?