**Preprocessing**

The videos are transcribed using Google’s speech recognition to extract text from speech, then the text is passed through the following NLP pipeline to get cleaned:

1. Punctuation is removed.
2. Words are converted to lower case.
3. Tokenization is done.
4. Stop words are filtered out.
5. Words are lemmatized.

The cleaned text is then fed to a TF-IDF vectorizer to represent the sentences in a numerical format suitable for the ML and DL models.

**Model Training**

After preprocessing, the TF-IDF vectors are fed to various machine learning models like SVMs and XGB and SGD etc.. and deep learning models like a simple ANN, RNNs were not viable here due to the limited number of samples (only 120 samples) which is not sufficient to train an embedding layer.

**Important remarks and discoveries**

Other more complex approaches such as using Glove for vectorizing the text yield similar results to using TF-IDF so we chose to use TF-IDF because it is so much more efficient computation wise, the best model in this modality was the simple ANN yielding an accuracy of 79%, the machine learning algorithms’ were not so far behind achieving peek accuracies of around 75%.