

Approach

I am planning to use two models which are bidirectional LSTM and Bert to compare the accuracy.

I started with Bidirectional LSTM since it has a simple structure compared to BERT but since BERT performs better with a large dataset and is trained on a large library of data, I want to compare their performance on this particular dataset.

Process

Since the data provided is not balanced a third data set is used for training which is more balanced. There are other ways to generated data from the given dataset by using different resampling techniques or creating synthetic data but since there was already a well-made dataset, I used that.

After removing all the unnecessary columns, only the labels and text were left which was eventually used. The text was preprocessed by removing all the unnecessary words such as pronouns, stop words etc. and converting the words to their root words using NLTK library performing these tasks beforehand for both can contribute to the accuracy of the model. One hot encoding is done on the words in the test to convert them to vectors and then padding is done to convert them to a fixed dimension. Doing one hot encoding before the embedding may or may not give better outcomes. Here both embedding and one hot encoding method is used for better outcome. A sequential model in Keras for a binary classification task using an embedding layer followed by a bidirectional LSTM layer and a dense output layer. For now, only specific parameters are explored, hyperparameter tuner can be used to explore other options such as changing the activation function, choosing the embedding size etc. Model is fit and the epoch number and batch size are also a hit and trail.

Architecture

Therefore, there are three layers in your neural network:

Embedding Layer

Bidirectional LSTM Layer

Dense Output Layer

The embedding layer captures the semantic meaning of words, the bidirectional LSTM processes sequences bidirectionally, and the dense output layer produces the final classification output. More dense layer can be added to the architecture but that could lead to overfitting, to counter that regularization techniques can be incorporated.