For the TF-IDF model I use a fully connected network consisting of 5 linear layers with dropout and ReLU activations. I find the FCN performs well on the validation set, again though it suffers on the test set. This is likely due to the TF-IDF features being restricted to the vocabulary of the training set only. It does not make sense to utilise other types of models for TF-IDF such as CNNs or RNNs as the order of the features does not convey information due to the bag of words approach.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Technique | Dataset | Weighted F1 | MCC | AUC | Accuracy |
| FCN | Validation | 0.934 | 0.844 | 0.948 | 0.932 |
|  | Test | 0.434 | 0.279 | 0.622 | 0.460 |

For the Transformer, I train a classification head using the CLS token from the transformer. I experiment with freezing vs not freezing the transformer and also using Longformer.

[TABLE]

I find that…

[TABLE]

Dataset imbalance is not an issue here…

I find the most effective to be … with hyperparameters … and loss function …