FFNN

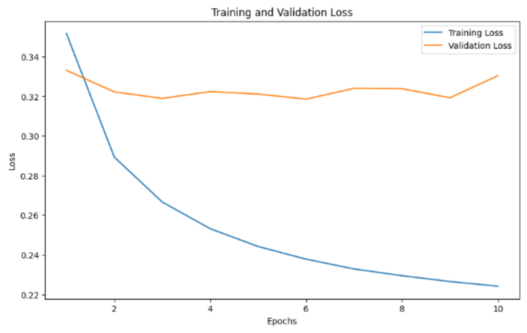
Use and organization of Datasets

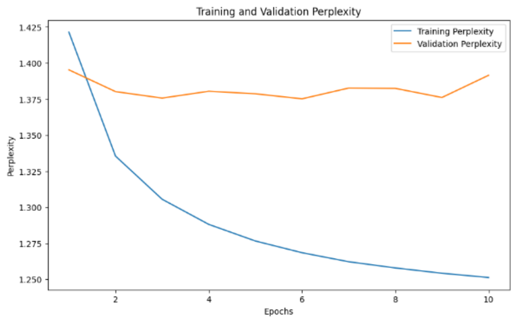
The model only uses the mix dataset for training, validation, and testing. Each bio is split into sequences according to the window size parameter. The sequences serve as the training examples. The label for each sequence is a binary label of either 0 or 1, depending on if the sequence came from a real or fake bio. The sequences are trained in batches and shuffled.

Operation and Hyperparameters of Model

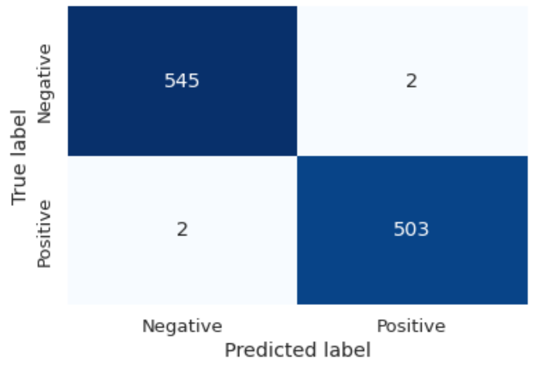
The model starts with an embedding layer with dropout, then 2 hidden linear layers, that ultimately feeds into a sigmoid activation function. We used the following hyperparameters: window_size=10, dropout_rate=0.3, d_model=100, d_hidden=100, batch_size=32, epochs=10, loss=BCELoss, optimizer=Adam, lr=0.0001

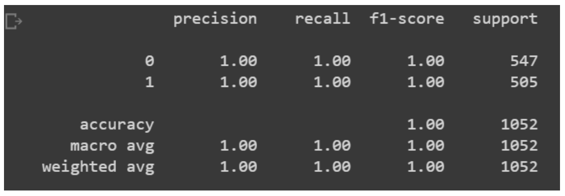
Learning curves, Perplexity for Training, Validation and Testing





Fakes Detection Result in Confusion Matrix





```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       547
           1       1.00      1.00      1.00       505

    accuracy                           1.00      1052
   macro avg       1.00      1.00      1.00      1052
weighted avg       1.00      1.00      1.00      1052
```

Limitations and Possible Solutions

The model is likely overfitting since it performs at a 99%+ accuracy and the validation loss does not increase. Previous attempted solutions (did not improve validation loss): increasing dropout rates, using L2 regularization and using a learning rate scheduler, use probabilities over a sequence (too computationally expensive). Other solutions: increase training epochs, make use of the other datasets to increase the number of training examples, finetune learning rate, batch size, optimizer, dropout

LSTM

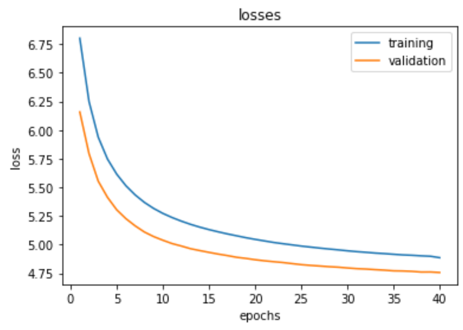
Use and organization of Datasets

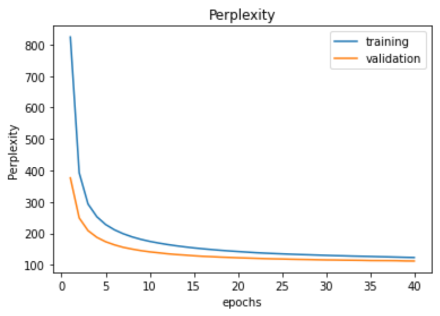
The model only uses the mix dataset for training, validation, and testing. Each bio is split into sequences according to the window size parameter. The sequences serve as the training examples. The label for each sequence is the word following the sequence. The sequences are trained in batches. The final prediction is determined by the probability of the token following <end_bio> (i.e. [REAL] or [FAKE] is more likely)

Operation and Hyperparameters of Model

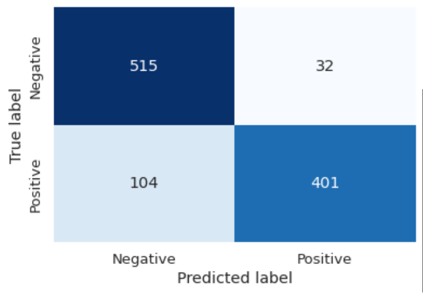
The model uses the pytorch lstm cell with drop out and hidden layers dependent on the hyperparameters. The output is a linear layer that is used to store log probabilities. Softmax is applied externally during inference. We used the following hyperparameters: window_size=5, dropout_rate=0.3, d_model=100, d_hidden=100, batch_size=32, epochs=40, loss=BCELoss, optimizer=Adam, lr=0.0001, n_layers=2

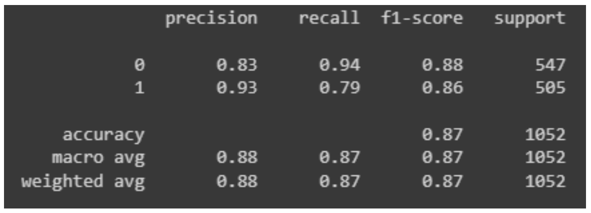
Learning curves, Perplexity for Training, Validation and Testing





Fakes Detection Result in Confusion Matrix





```
              precision    recall  f1-score   support

           0       0.83      0.94      0.88       547
           1       0.93      0.79      0.86       505

    accuracy                           0.87      1052
   macro avg       0.88      0.87      0.87      1052
weighted avg       0.88      0.87      0.87      1052
```

Limitations and Possible Solutions

The model can suffer from overfitting and vanishing/exploding gradients. Overfitting can be addressed with regularization techniques like dropout or weight decay. Vanishing/exploding gradients can be mitigated with gradient clipping or a different activation function. Small datasets may also limit the model's ability to learn effectively, so increasing the dataset size or using data augmentation can help.

Results

Why we selected the model?

FFNN has a much better accuracy, and we believe that it will be able to generalize to new data better despite concerns with overfitting.

Limitations and possible solutions

Again overfitting can be addressed similarly as stated above. We can also add rule based algorithms in order to improve generalization and shortcomings of language model based classification, such as looking for fake and unreasonable dates.