**Recurrent Neural Network**

Mengting Ding

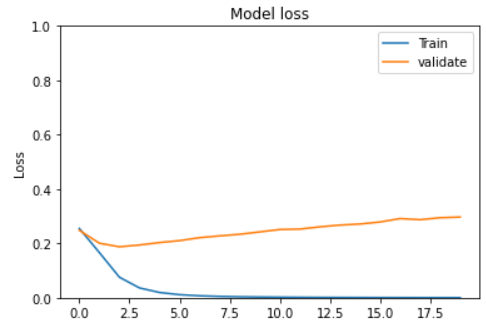
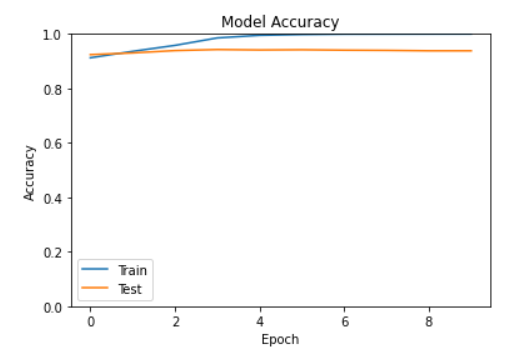
This twitter dataset is downloaded from kaggle, the main goal is to identify if there is any hate speech in the test set. To prepare the data, train set and test set are read through pandas read\_csv with utf8 encode.

Test size is further split into 80% of training and 20% of validation set. The y target is the label column, also split into 80% of training and 20% of validation. Words in the tweets are preprocessed by tokenizer and padded to the max length.

**First model:** No LSTM and dropout layer

Maxlength = 200, with only one embedding layer, one flatten layer and one dense layer with sigmoid as activation function.

Accuracy: 1.00, val\_acc: 0.94, loss:0.0021, val\_loss:0.2329

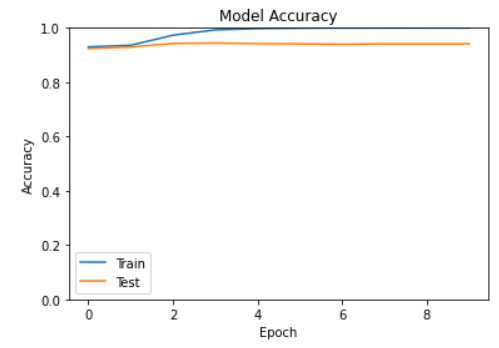
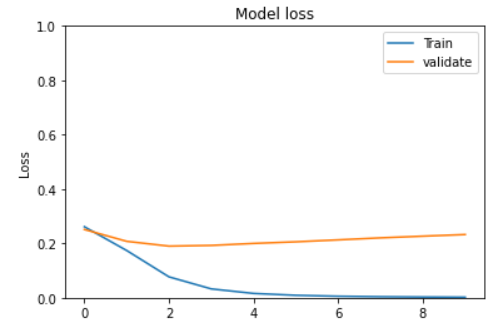


Predicted percentage of hate speech tweets in test set: 1%.

**Second model:** No LSTM and dropout layer

Maxlength = 100

Accuracy: 1.00, val\_accuracy:0.94, loss: 0.002, val\_loss:0.233

The percentage of hate speech deteched is 0.5%.

Run time is 8 seconds.

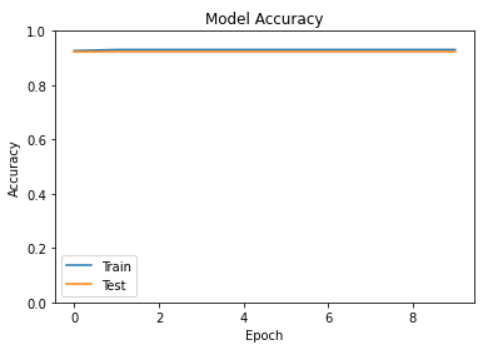
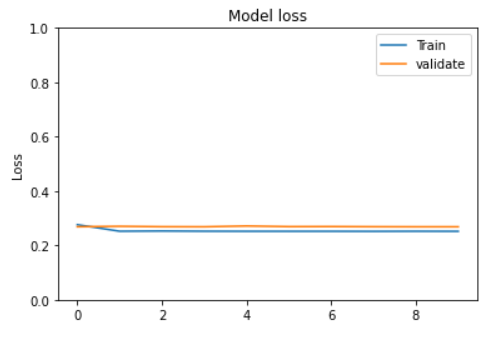
**Third model:** withLSTM

Maxlength = 100

Without a Dropout layer, the model runs faster than the one with dropout layer.

Maxlength = 100, with a single embedding layer, one LSTM layer, one flatten layer and one dense layer

Accuracy:0.931; val\_accuracy: 0.924, loss: 0.251, val\_loss:0.268

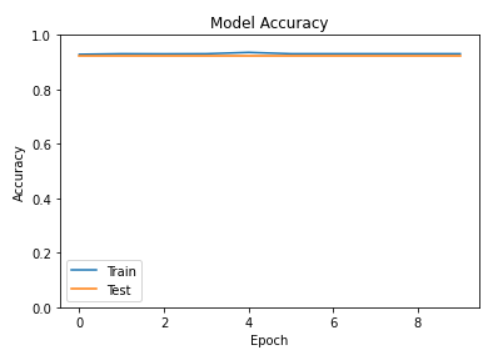
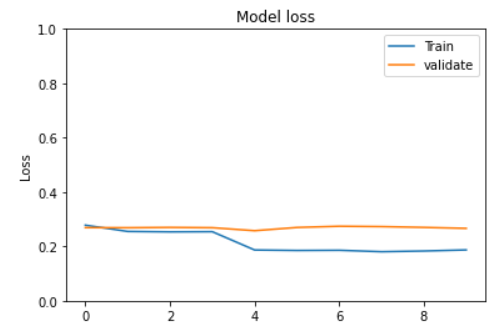
 

Percentage of hate speech is 0.

Run time 17 mins.

**Fourth model**: LSTM and 0.2 dropout rate

Maxlength = 50

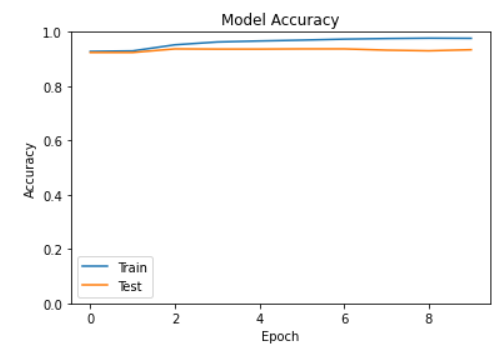
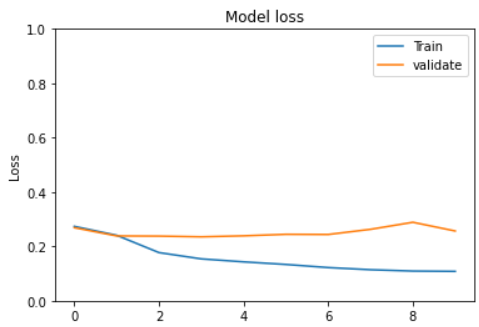
Accuracy: 0.93, val\_accuracy:0.92, loss: 0.187, val\_loss:0.2663

Run time: 8mins 26s

No hate speech predicted.

**Fifth model**: With LSTM

This model contains one embedding layer, one LSTM layer, one flatten layer and one Dense layer with activation sigmoid. The max length is equal to 50.

Accuray: 0.976, val\_accuracy: 0.934, loss: 0.108, val\_loss: 0.256.

Running time is 10 mins 30s.

Prediceted percentage of hate speech is 2.3%.

**Conclusion:**

According to the graphs and accuracy rate and loss rate, the fifth model seems performs the best among them all. As it is not like the third and fourth model that there is no significant improvement during training, nor like the first and second model which have a trainning accuracy of 1 and is clearly overfitting the model.

The models may not perform well due to the imbalanced dataset as there are only 2242 entries labelled as hate speech compared to 29720 labelled as non-hate speech in the training set. The percentage of hate speech in training set is 7.5%, hence it is reasonable for the model to predict a very low percentage of hate speech in the testing set.