**GRU**

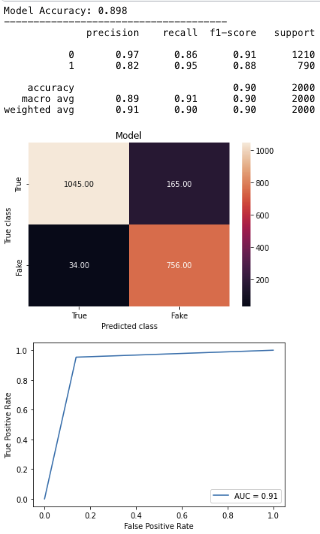
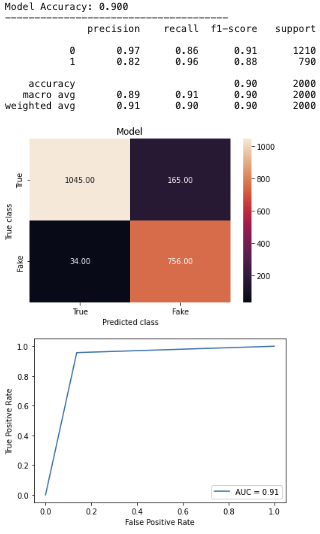
1. **Word Embedding: Glove**

Glove 300 dimensions is chosen for GRU over Word2Vec 300 dimensions. Both Glove and Word2Vec give very similar precision, recall and accuracy, but since Glove is slightly better. Therefore, Glove is adopted for training GRU models for the task.

gru\_gv / gru\_w2v

**(@Elaine you can search for “gru\_gv” in the code to see the model, call me if need clarify)**

Glove Word2Vec

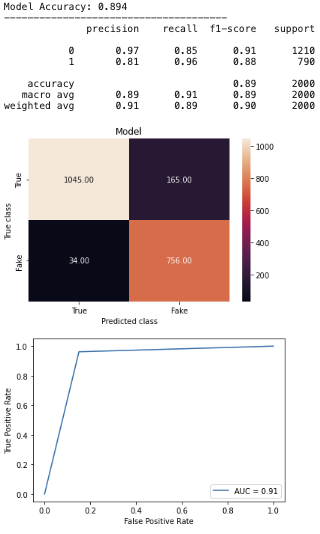
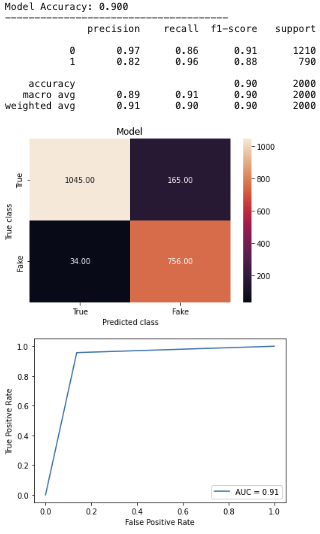


1. **Call backs: Learning Rate Decay Helps**

Both models have model checkpoint call back, but one also has learning rate decay while another one does not have. With learning rate decay, precision, recall and accuracy are better. Early stop is not adapted as there are only 20 epochs to run.

gru\_gv / gru\_gv3

With Learning Rate Decay Without Learning Rate Decay

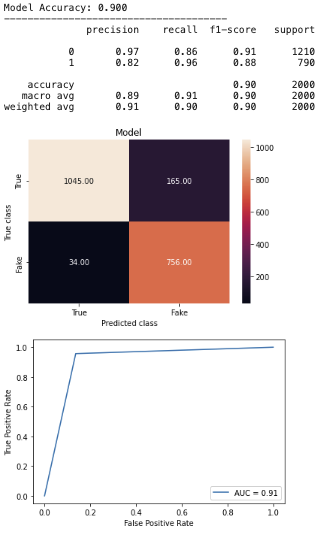
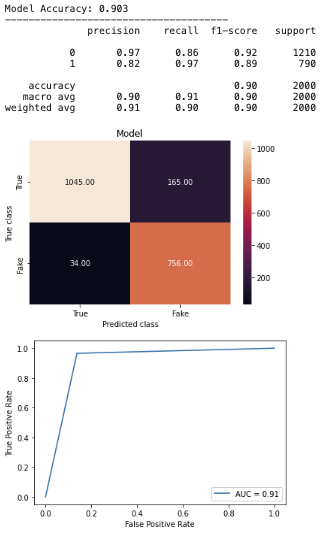


1. **Better without Bidirectional layer**

The accuracy and recall without Bidirectional layer are slightly better than that of with bidirectional layer.

gru\_gv1 / gru\_gv

Without bidirectional layer With bidirectional layer

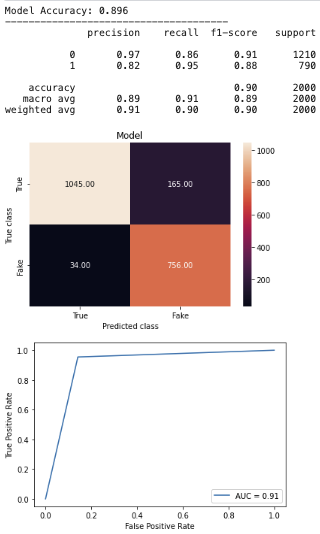
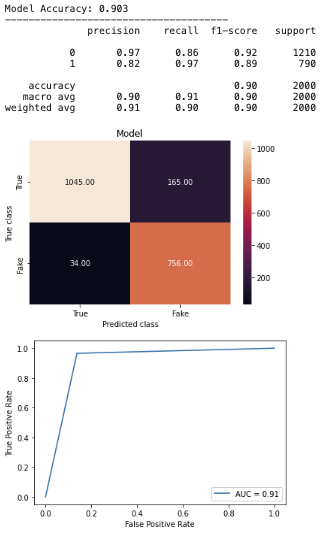
****

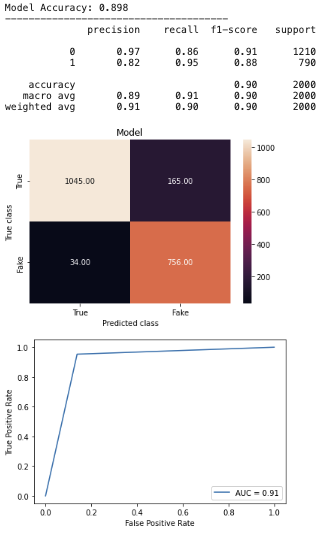
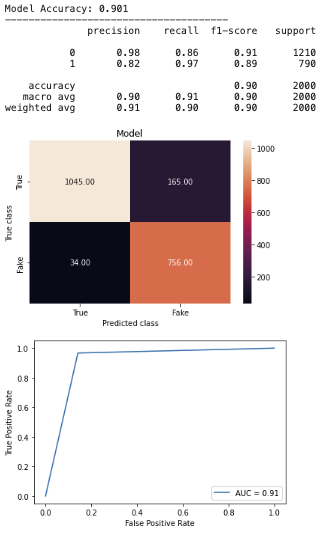
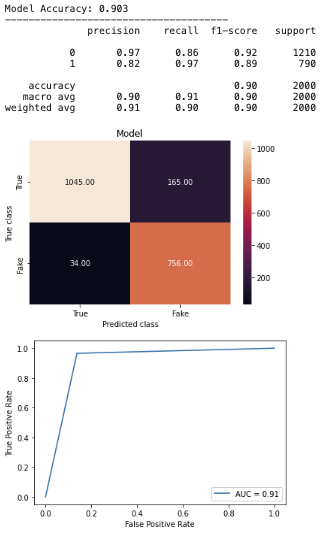
1. **Better with fewer GRU layers**

The recall of fake news is slightly better in the model with 1 layer of GRU than that of with 2 layers.

gru\_gv1 / gru\_gv4

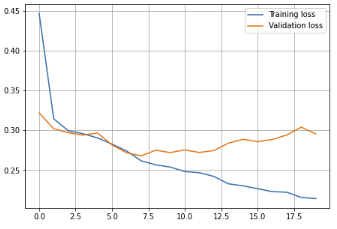
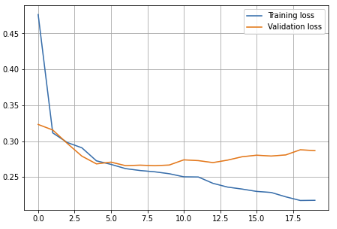
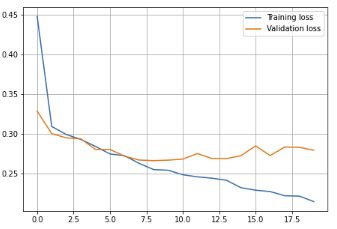
With Single GRU layer With Double GRU layer

****

1. **Better with a Dropout layer at rate of 0.1**

The 0.1 dropout rate model is the best in terms of precision and recall, and has the flattest and smoothest validation loss curve.

gru\_gv6 / gru\_gv1/gru\_gv5

Dropout rate = 0.1 No Dropout Dropout rate = 0.2

**Best Model:**

gru\_gv6

The best GRU model is a model that has one non-bidirectional GRU layer, a dropout layer of 0.1 dropout rate, followed by a dense layer and output layer. Dense units and batch size are also fine-tuned.

