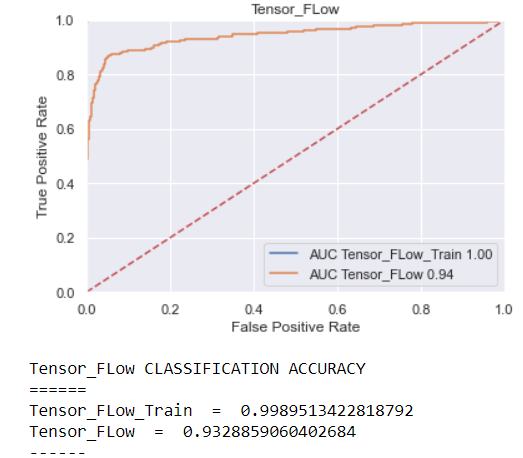
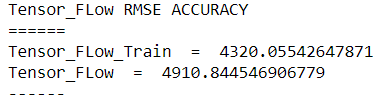
MSDS-422-SEC56\_Assignment 4 Daniel Zaremba

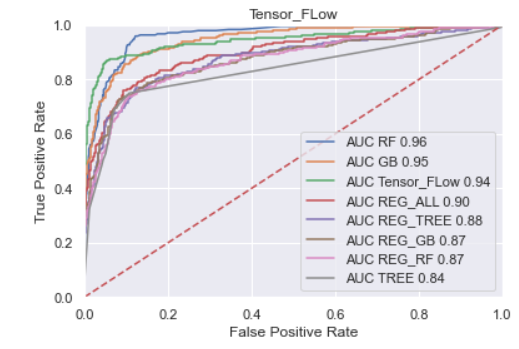
**TensorFlow Observations**

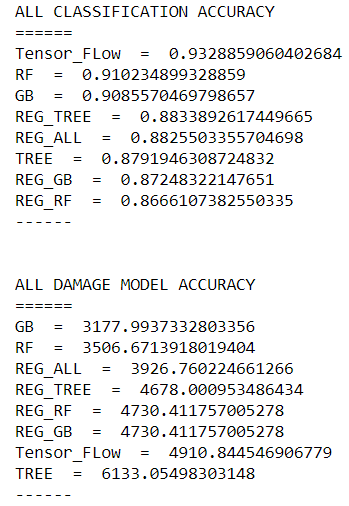
**TensorFlow model using all variables:**



**TensorFlow model using Gradient Boosting model variables:**







**Observations**

Having played around with the TensorFlow parameters, I initially had the units at two times the gradient boosting model variable count divided by three and the epoch at 100 this resulted in a test accuracy of .875, changing the unit to be just two times the variable count, increased the epoch to 500 and fed the TensorFlow model all variables resulted in a lift in test accuracy to .933. I would recommend using the TensorFlow model as it has the highest classification accuracy. I would also recommend using the Gradient Boosting model for predicting Loss Amounts. I tried different activation functions like softmax and sigmoid but these both performed worse than relu.

The Default predictive variables for the most part make sense, M\_DEBTINC makes sense as individuals that don’t provide their debt to income ratio may be hiding high debt and be more risky, on the other hand I’m not sure about the M\_VALUE field, I would reach out to a subject matter expert to see if a missing home value is predictive of default. With respect to the Loss Amount variables again we see M\_DEBTINC and also TRUNC\_IMP\_DEBTINC which makes sense, also showing as predictive is TRUN\_IMP\_CLNO which is a double edge sword, the fewer the tradelines the riskier but any individual may have a lot of tradelines and run up a lot of debt which is also risky.