Module 21 Challenge: Report

OVERVIEW: This analysis is being used to predict the success or failure of crowdfunding campaigns for Alphabet Soup.

RESULTS:

* Data Preprocessing
  + What variable(s) are the target(s) for your model? The primary targets are the ‘Application Type’ and ‘Classification’ fields of the data.
  + What variable(s) are the features for your model? The feature we hope to predict is the ‘Is Successful’ variable.
  + What variable(s) should be removed from the input data because they are neither targets nor features? I removed the variables ‘Affiliation’, ‘Use Case’, and ‘Special Considerations’ because I didn’t feel they were material to the final outcome.
* Compiling, Training, and Evaluating the Model
  + How many neurons, layers, and activation functions did you select for your neural network model, and why? I chose to use two hidden layers with units of 10 and 20. For my initial attempt I used the Relu activation for both hidden layers and Sigmoid for the output layer. I thought this seemed like a good starting place and provided change points that I could use for future models.
  + Were you able to achieve the target model performance? No, with my initial model I only had 65.6559% accuracy.
  + What steps did you take in your attempts to increase model performance? I made two additional attempts to find a suitable model. First I added back in the removed variables in case they were more determinate of the outcome than I had assumed. My first optimization attempt was to add a hidden layer for a total of 3 and change the units and activations for the layers. My hidden layers had the following units/activations: 6/Relu, 6/Tanh, 6/Tanh. I hoped that reducing the complexity by reducing the units while adding an additional layer would make a more accurate and efficient model. My final attempt was a keras tuner that used as many as 30 units in each layer and could add as many as 5 hidden layers. My hope was that with this amount of variability in the models a highly accurate combination of units and layers could be found by the tuner.

SUMMARY: Overall the models performed better with the variables I had removed added back in. The first optimization improved the accuracy to 72.6297% and the tuner was able to get a slightly better 72.89% on its best model. I wonder if finding the most influential variables and then fine tuning the model around them would help increase the accuracy above the 75% threshold or even higher.