The purpose of this analysis was to build a machine learning model to help Alphabet Soup determine which applicants for funding have the highest chance of utilizing the funds successfully.

* The target variable was the IS\_SUCCESSFUL column which indicates whether a particular applicant has successfully used funding in the past.
* The features were as follows:
  + Application type
  + Affiliation
  + Classification
  + Use case
  + Organization
  + Income amount
  + Requested funding amount
* The columns EIN and Name were removed because they are not relevant to the predictive model. After digging into the data more, I also removed the status column, because there were only 5 values for inactive, and additionally whether an organization is “active” in the system has little relevance to whether their spending is successful. Additionally I removed the “special considerations” column. While this information might be relevant, there was such a small number of values marked “yes” that it seems unlikely that it would have a big impact on the machine learning model.
* I chose to use four hidden layers, with differing activation functions and units, as shown in the image below. I arrived at this number mostly through trial and error in an attempt to get the accuracy above 75%. I stopped adding layers and units once it became apparent that it was no longer having a significant effect. A few times I attempted to reduce the amount of layers and units after doing other preprocessing, but each time it resulted in lower accuracy.

A screen shot of a computer program

Description automatically generated

* I attempted to get the accuracy above 75% many multiple times but ultimately was not able to get the accuracy above 74.8%. The steps I attempted were as follows:
  + I first tried adjusting the number of hidden layers, their units, and their activation functions as well as the number of epochs. I did a few rounds of this to try and fine-tune.
  + Next I went back to the Ask Amount column because I noticed it had very specific amounts, and looking back at the number of unique values for this column, there were over 8,000 different values. Therefore I decided to create bins. I spent a few rounds tweaking the number of bins and the values within them.
  + As mentioned above, I then removed the Status column and the Special Considerations column for reasons already mentioned.
  + I then spent a few more rounds tweaking the number of neurons, units, epochs, and activation functions, as well as experimenting with some of the other columns, but many of the things I tried reduced the accuracy instead of improving it so ultimately I went with the model that gave me the highest accuracy.