**Alphabet Soup Analysis**

**Overview**

The nonprofit foundation Alphabet Soup wants a tool that can help it select the applicants for funding with the best chance of success in their ventures. Using machine learning and neural networks along with their provided dataset, we need to use the features available to create a binary classification to achieve this endeavor.

**Results**

Data Preprocessing

* The target for this model will be the “IS\_SUCCESSFUL” variable, which simply determines if the funding of that venture was or was not successful.
* All other variables with exception of “EIN” were considered features for the model. These include, application type, affiliation, classification, use case, organization, status, income amount, special considerations, and ask amount. Initially the “Name” variable was not used as a feature, but was later added in for optimization.
* The “EIN” variable was not considered a feature of this model because it was simply a supplemental identification for the “Name” variable.

Compiling, Training, and Evaluating the Model

* Initially, two hidden layers were used outside of the output layer. Layer one used 80 neurons, and layer two used 30, with both activations using “relu”. These were chosen as a starting point taking the number of columns into account, which at the time was 44.
* With the above model, we were only able to achieve 72.94% accuracy in the model, this was below the target threshold of 75%, so optimization was required.
* To optimize the model, the first thing we did was add the “Name” variable, which was excluded outright before. This ended up taking our number of columns from 44 to 454. After that, the model may have been overfitting, so the number of neurons was scaled back to 20 on layer 1 and 15 on layer two, still using relu on both. With these adjustments we were able to cross the threshold and got up to 79.89% accuracy.

**Summary**

Summarize the overall results of the deep learning model. Include a recommendation for how a different model could solve this classification problem, and then explain your recommendation.

The main issue with the initial model was the exclusion of the each company’s name, once that was added in an binned, we were able to see much more accurate results as it provided the necessary relevance to the data. It is possible that a random forest model could have been appropriate for this job as it could have told us which variables ended up having a greater weight from the start. Below are the final output results:

A screenshot of a computer program

Description automatically generated

A screen shot of a computer code

Description automatically generated