Alphabet Soup Charity Analysis

Overview: The purpose of this analysis was to determine whether applicants will be successful if funded. By using a provided dataset, we will create an algorithm to predict the success rate of such applications.

Results:

* Data Preprocessing:
  + The variables considered for targeting were whether the funded applicants were successful. We could see this in the column named “IS\_SUCCESSFUL”, by using this output we were then able to focus on our features.
  + The features used for the initial model were application types, application counts, classification types, and classification counts. After the initial model we optimized the model by accounting for the name of the applicants as well.
  + The variables removed from the initial model were the employee identification number (EIN), and the name. For the optimized model however, we kept the name category, and dropped EIN, Special Considerations, and Status for optimization.
* Compiling, Training, and Evaluating the Model:
  + In our initial model we used 4 neural layers, complete with input, 2 hidden layers, and an output layer. For activation of our initial model, we used 2 ‘relu’ and 1 ‘sigmoid’. Finally our model was optimized using ‘adam’.

For the optimized model we had 5 neural layers, complete with input, 3 hidden layers, and an output layer. For activation we used 1 ‘relu’, and 3 ‘sigmoids’. The model was optimized using ‘adam’.

‘Sigmoid’ was used after ‘relu’ as it is able to transform the data received from ‘relu’ into a value between 0 and 1, thus showing the efficiency of the models on a smaller scale.

* + Our initial model had an accuracy of 73% so further steps could be taken to achieve target performance.
  + In order to increase the model’s accuracy we went back to our initial data to see what columns could be dropped that had no significant bearing on whether an applicant was successful. To see what columns could be dropped we checked the value counts of the columns that had only 2 outputs. In this case these additional columns we were able to drop were Special Considerations, and Status. This is also when we added the neural layers for greater filtering, and changed our activations.
* Summary:
  + Our initial model did not meet the Charity’s requirements of 75% therefore that model must be dropped from consideration, however, after keeping the name column we are able to tweak our final output to a 79% accuracy rate making our optimized network the best identifier. Further models used were a random forest classifier, but this did not beat the neural network as it only had a 77% accuracy rate for applicants.