Alphabet Soup Charity Prediction Model Report

Overview of the Analysis

The purpose of this analysis was to create a binary classifier that is capable of predicting whether applicants will be successful if funded by Alphabet Soup. The data from Alphabet Soup's business team consists of more than 34,000 organizations that have received funding over the years. Within this dataset are a number of columns that capture metadata about each organization such as the following:

Data Preprocessing

- **Target Variable for the Model:** The target for our model was the "IS_SUCCESSFUL" column. This column indicates whether the money was used effectively.
- **Feature Variables for the Model:** The features for our model included every column except for "IS_SUCCESSFUL", "EIN", and "NAME". These included "APPLICATION_TYPE", "AFFILIATION", "CLASSIFICATION", "USE_CASE", "ORGANIZATION", "STATUS", "INCOME_AMT", "SPECIAL_CONSIDERATIONS", and "ASK_AMT".
- **Non-Target or Non-Feature Variables:** The "EIN" and "NAME" columns were removed from the input data as they are identification information and do not contribute to the model's predictive ability.

Compiling, Training, and Evaluating the Model

- **Neurons, Layers, and Activation Functions:** The deep learning model was a binary classifier that was composed of two hidden layers with 80 and 30 neurons respectively. The input feature data determined the number of input features (number of neurons for the input layer). The activation function used for the hidden layers was ReLU or Rectified Linear Activation. Because our output is binary, the activation function used in the output layer was Sigmoid.
- **Achieving Target Model Performance:** The model did not initially reach the desired performance. The target model's accuracy was not achieved on the first run.
- **Steps to Increase Model Performance:** To attempt to increase model performance, the number of neurons in the hidden layers was increased, additional hidden layers were added, and different activation functions were tested. Also, the data was bucketed differently for some of the features to reduce the feature space.

Summary

The deep learning model was able to predict the success of the grant applications with an accuracy around 73%. Although this is a good start, there is definitely room for improvement.

For future work, I would recommend trying different models to improve accuracy. A Random Forest Classifier or Gradient Boosting could potentially provide a better result. These models work well with tabular data and might be able to pick up on patterns that the deep learning model missed. Additionally, further feature engineering and selection could potentially improve the model's performance. For example, using PCA (Principal Component Analysis) to reduce the dimensionality of the data might be beneficial. It's also worth considering gathering more data if possible, as having more examples for the model to learn from could potentially improve its performance.