Overview

The goal of this analysis was to predict the success of funding applications submitted to Alphabet Soup by creating and evaluating a deep learning model. The neural network is used to analyze historical data and identify patterns that contribute to successful funding outcomes. This model will enable the foundation to make data-driven decisions in prioritizing applications that are more likely to succeed which will optimize resource allocation and impact.

Results

* Data Preprocessing
  + The target variable, ‘IS\_SUCCESSFUL’, is used because it indicates whether the funding application was successful or not.
  + Some of the feature variables include ‘APPLICATION\_TYPE’, ‘AFFILIATION’, ‘CLASSIFICATION’, ‘USE\_CASE’, ‘ORGANIZATION’, ‘STATUS’, ‘INCOME\_AMT’, ‘SPECIAL\_CONSIDERATIONS’ and ‘ASK\_AMT’. These columns are used because they include either categorical values or numerical values that correlate with success.
  + The removed variables at the very beginning, ‘EIN’ and ‘NAME’ are removed because they contain unique identifiers that are unnecessary to the analysis. A little further in the code, ‘AFFILIATION\_Family/Parent’, “CLASSIFICATION\_C2000’, ‘APPLICATION\_TYPE\_T3’, ‘STATUS’, ‘SPECIAL\_CONSIDERATIONS\_N’, ‘INCOME\_AMT\_5M-10M’ are removed to improve the model’s performance by removing unnecessary static.
* Compiling, Training, and Evaluating the Model
  + Neurons, Layers and Activation Funtions
    - This model consists of:
      * Input Layer: Matches the number of features in the dataset
      * Hidden Layer 1: 100 neurons with the tanh activation function
      * Hidden Layer 2: 30 neurons with the tanh activation function
      * Output Layer: 1 neuron with sigmoid activation function to handle binary classification.
    - This setup was chosen after many iterative tests trying to find the balance between model complexity and accuracy
  + Target Performance Achieved
    - This model achieved an accuracy of approximately 72% which is slightly below the target of 75%
      * I was able to achieve a 74% accuracy rating at one point but when the code was reran with no changes made, it dropped to 72%. I don’t know what fluke happened to get the higher accuracy though.
    - Steps taken to improve performance:
      * The number of neurons in each hidden layers was adjusted many times.
      * I started off using the activation function ReLU but switched to tanh and was able to maintain a higher accuracy rating using tanh.
      * Removed low-correlation column features from the dataset. Originally, I removed all that had negative correlation but that gave the worst accuracy score of all so many were readded.
      * Batch size, learning rates and epochs were changed and adjusted many times for each iterative change.

Summary

The deep learning model developed for this analysis achieved a maximum accuracy of 74%, though it’s performance fluctuated slightly during retraining. Through many iterative tests and changes, the final accuracy settled at 72%. While the model came close to the target of 75%, further optimization may be needed to achieve slightly higher consistent performance.

This model effectively demonstrated the potential of using a neural network to predict funding application success. It highlights the importance of preprocessing, iterative tuning and feature selection in improving performance.

**Model Recommendation**

While the neural network performed well, Random Forest Classifier might be more suitable for this classification problem. Random Forest models excel at handling categorical and numerical data with minimal preprocessing. Random Forest is also less sensitive to hyperparameter tuning then neural networks. Additionally, they provide feature importance metrics, which can offer insights into which features have the most significant impact on predictions.

Edit: I tested everything again shortly before turning this in with the same code I got the 73% accuracy score and no longer got 73%. I tried to edit all incorrect accuracy results but in case I missed any, 73% accuracy is supposed to be 72%.