Alphabet Soup Charity Report

# Overview

The purpose of the analysis is to build and optimize a neural network model that is able to predict whether applicants will be successful if funded by Alphabet Soup.

# 2.0 Results

## Data Preprocessing

What variable(s) are considered the target(s) for your model?

* Since the purpose is build a model that is able to predict the success of whether an applicant is funded, I used the variable, *IS\_SUCCESSFUL*, as the target
* What variable(s) are considered to be the features for your model?
* The variables considered for features in the model were: *APPLICATION\_TYPE, AFFILIATION, CLASSIFICATION, USE\_CASE, ORGANIZATION, STATUS, INCOME\_AMT, SPECIAL\_CONSIDERATIONS, ASK\_AMT*
* What variable(s) are neither targets nor features, and should be removed from the input data?
* The variables removed from dataset were: *EIN* and *NAME*

## 2.2 Compiling, Training, and Evaluating the Model

* How many neurons, layers, and activation functions did you select for your neural network model, and why?
* Decided on 4 layers with all activation functions as Relu, which is the best accuracy as seen in the table below
* The neurons were determined to be about two thirds of the features and reduced by half for each following layer
* Number of neurons:
* Layer 1 = 40
* Layer 2 = 20
* Layer 3 = 10
* Layer 4 = 5

|  |  |
| --- | --- |
| * **Layers and Activations Functions Table** | |
| * 2 Layers and Activation All Relu |  |
| * 3 Layers and Activations All Relu |  |
| * 3 Layers and Activations Mixture of Relu and Sigmoid |  |
| * 4 Layers and Activations All Relu |  |
| * 4 Layers and Activations Mixture of Relu and Sigmoid |  |

* Were you able to achieve the target model performance?
* I was not able to achieve the target model performance of 75 percent accuracy despite increase and decreasing the bin size for Application Type and Classification and drop columns from the charity dataset.
* What steps did you take to try and increase model performance?
* The first step was to determine the correct bin size for Application Type and Classification: <20, <10, or <5
* It was determined that the bin size <10 was the most optimal as shown in the table below

|  |  |
| --- | --- |
| * **Bin Size for Application Type and Classification Table** | |
| * Bin Size <20 |  |
| * Bin Size <10 |  |
| * Bin Size <5 |  |

* The second step is to determine if dropping a variable would cause a significant increase to accuracy
* It was determined that despite the relatively small differences in accuracy, the dropped variables that made the most significant impact were: Special Considerations and Ask Amount

|  |  |
| --- | --- |
| * **Dropped Variables Table** | |
| * Affiliation Variable Dropped |  |
| * Use Case Variable Dropped |  |
| * Organization Variable Dropped |  |
| * Status Variable Dropped |  |
| * Income Amount Variable Dropped |  |
| * Special Considerations Variable Dropped |  |
| * Ask Amount Variable Dropped |  |

# 3.0 Summary

Overall, the results illustrated that bin size, dropped variables, and layers and activation’s function made relatively small differences to the deep learning model. Although the results proved that a model with a bin size of less than 10, 4 layers with all Relu activation function, and an additional variable drop of Special Considerations was the most optimal with an accuracy score of 0.72944, it is important to note that this may have just been the result of variance in training the model. Since the differences are relatively small, another combination may be more optimal when the model is ran again. Thus, this raises the question if there is another model that could solve this classification problem better.

Before using a different, the Keras Tuner Python library could be used to determine programmatically the optimal set of hyperparameters for the deep learning model. If it results into an improved score of over 75 percent accuracy, another model could then be looked at. One such model could be using the logistic regression model instead of the neural network model. A logistic model is for binary classification problems to predict a binary outcome, which was the case for Alphabet Soup since they sought to predict whether applicants will be successful if funded by the charity.