**Alphabet Soup Charity Funding Predictor**

Joseph March

2021-12-13

**Overview**

**Purpose**

The purpose of the analysis was to create a binary classifier capable of predicting whether applicants for funding will be successful if funded by Alphabet Soup. A CSV containing more than 34,000 organizations that have received funding from Alphabet Soup was provided for use in training the model. Testing the model for accuracy and making improvements was also part of the purpose of the analysis.

**Results**

**Data Preprocessing**

The **IS\_SUCCESSFUL** variable, which determines if the funded money was used effectively, was considered the target for the model. The following variables were considered features for the model:

* APPLICATION\_TYPE – Alphabet Soup application type
* AFFILIATION – Affiliated sector of industry
* CLASSIFICATION – Government organization classification
* USE\_CASE – Use case for funding
* ORGANIZATION – Organization type
* INCOME\_AMT – Income classification
* SPECIAL\_CONSIDERATIONS – Special consideration for application
* ASK\_AMT – Funding amount requested

The variables deemed to be neither targets nor features, and thus removed from the input data, were:

* EIN – An identification column
* NAME – An identification column
* STATUS – Active status

**Compiling, Training, and Evaluating the Model**

The initial attempt to build a neural network model for this data resulted in a 0.728 accuracy. After completing the initial model, attempts to improve the model were performed, three of which were saved for comparison, as model 1, model 2, and model 3. The best model (referred to as model 2) resulted in a 0.731 accuracy. It used the ReLU activation function (currently the most used activation function for binary classification problems), had 3 hidden layers of 256, 128, and 64 neurons (based on testing of different numbers of layers and neurons), had 41 input features, and 1 output unit with the output layer relying on Sigmoid activation.

Other elements tested were changing the number of bins for rare occurrences in columns, changing the number of values for bins, dropping unneeded columns, and changing the number of epochs. These were in addition to the previously mentioned changes to number of neurons, number of layers, and activation functions.

**Initial Model:**



**Model 1 (Optimized):**



**Model 2 (Optimized):**



**Model 3 (Optimized):**



**Summary**

**Results**

While model 2 was the best, with an accuracy score of 0.731, none of the models made it to the desired 0.75 score. It is possible, with more robust data, the model could be tuned further to reach the desired score.

**Recommendation**

With the model as it currently exists, Logistic Regression could be an alternative option to using a neural network model. This was tested, for comparison purposes, and returned a 0.72 accuracy score. While this is a slightly lower accuracy, the reduction in processing power may make up for the reduction in accuracy.