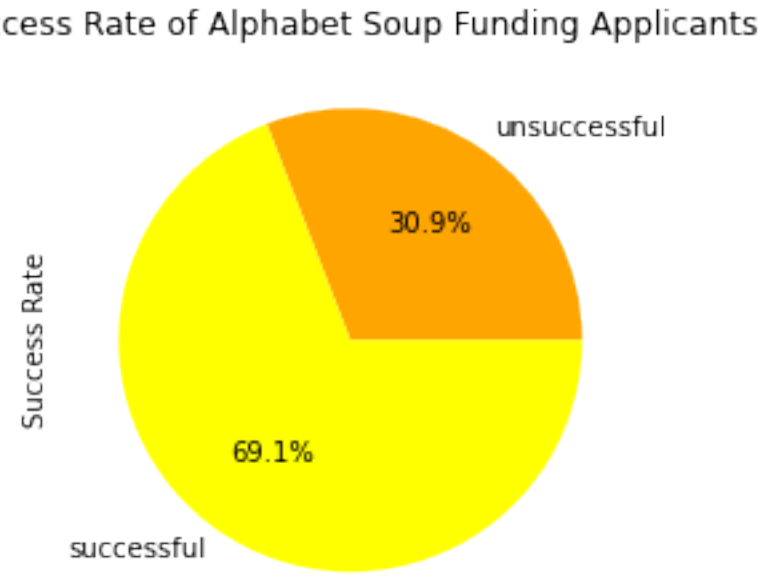
Alphabet Soup Charity Funding: Predicting Success

Overview

Based on the data provided on the success or failure of previous Alphabet Soup charity funding applicants, 69.1% of funded applicants are successful in the use of their funding and 30.9% are unsuccessful.



While a success rate of almost 70% isn't terrible, it would obviously be preferable to be able to better filter applicants to ensure that funding goes to those that will succeed. In order to predict the probably outcome of any candidate for funding, machine learning can be applied. In this instance, the use of deep learning in the form of a neural network for binary classification has been requested.

Results

Data Preprocessing

- The target is the IS_SUCCESSFUL column, which shows whether a funding applicant was successful in their use of funds.
- The features are the following columns: APPLICATION_TYPE, AFFILIATION, CLASSIFICATION, USE_CASE, ORGANIZATION, INCOME_AMT, and ASK_AMT.
- EIN and NAME columns are neither targets nor features and do not provide useful data for the neural network model so they are removed from the data. SPECIAL_CONSIDERATIONS and STATUS columns are also removed because they do not contribute meaningfully to the data due almost all of the rows containing the same value.

Compiling, Training, and Evaluating the Model

- The best neural network model had three layers (including the output layer), with 8, 10, and 1 units, respectively. The first and second layer used tanh activation function and the final layer used sigmoid.
- Although this final model was slightly better than the previous ones, it did not achieve the target of 75% accuracy.
- After trying an initial model, three attempts at optimization were made after slight changes in data preprocessing, namely the dropping of SPECIAL_CONSIDERATIONS and STATUS. The first attempt involved adding more neurons to first and second layer. The second attempt involved adding another hidden layer. The third attempt removed the added hidden layer and switched the first and second layer from relu to tanh activation functions.

Summary

Despite several attempts at optimization, even the best version of the neural network model was not able to meet the goal of 75% accuracy, peaking at about 73% accuracy. In addition, the loss was concerningly high. While it would be possible to continue to attempt to optimize this model to meet that goal of 75% accuracy, it could be argued that a different type of machine learning model, such as a relatively simple logistic regression model, might be better suited to predict the success of funding applicants.