

Overview of the Analysis

I used a dataset of historical funding decisions made by Alphabet Soup to build a model that can identify applicants for funding with the best chance of success in their ventures.

Data Preprocessing

The targets of the model was the variable named 'IS_SUCCESSFUL', which means was the money used effectively.

The features of the model included application type, affiliation, classification, use case, organization type, status, income, special considerations, and funding amount requested. The EIN variable was removed from the input data as it was neither a target nor a feature.

I used three different ML models on the data. The first two models displayed accuracy rates of approximately 73% (which was less than the desired 75% rate) despite using different numbers of layers and nodes. However, keeping the names variable as a feature increased the accuracy rate of the model to 78%.

Results

Here you will find the accuracy scores, loss scores, and parameters of the three machine learning models I used, along with what I did to increase model performance.

Machine Learning Model 1: I used two hidden layers with activation function 'relu' and included 43 as the input (number of columns) along with 80 as the units (approximately double the input), with the following results:

- Accuracy Score: 72.97%
- LossScore:55.68%

Machine Learning Model 2: I added another hidden layer, which decreased the performance of the model but by a minimal amount, with the following results:

- Accuracy Score: 72.79%
- LossScore:56.32%

Machine Learning Model 3: Knowing the social sector due to my own professional experience and that organization names carry weight, I decided to keep the name column and bin all names with less than 10 occurrences as 'other'. This required an increased number of inputs (266 - number of columns) and units (530 - approximately double the input) and I decided to maintain 3 hidden layers, resulting in the following results:

- Accuracy Score: 78.32%
- LossScore:54.76%

Summary

The third machine learning model performs best as it has a higher accuracy score - it will be correct nearly 80% of the time. This model takes into account the organization that is applying for funding,

which is crucial to determine the previous track records of organizations and if they have the best chances of success in their ventures.