Overview:

The purpose of this analysis is to find out the mechanism of whether the money was used effectively by analyzing the relationship between all variables provided by Alphabet Soup's business team. The objective is to use the deep learning model to achieve and optimize the accuracy until it is higher than 75%.

Results:

In the data preprocessing, variable of "is_successful" is the target for the model, and other variables are features of the model in achieving the result. The variables of "name" and "EIN" were removed from the analysis because they are simply identification of each row and provide no meaning to this analysis.

In the model prediction and evaluation, the initial analysis is conducted with two hidden layers. The first layer contains 80 neurons, the second layer contains 30 neurons, and both layers were set with activation function of "relu." The layers and neurons were chosen based on the data complexity. Based on the information provided, these data are moderately complex, so two hidden layers with 80 on first and 30 on second should be sufficient to analyze. The activation function of "relu" is simply the default as we first assume the data is linearly correlated.

Unfortunately, based on the analysis, we only achieved 73% accuracy, which is slightly lower than 75%. There are three steps I attempted to optimize the model and improve it to more than 75%. The first step is to add another hidden layer, the second step is to double the neurons in hidden layers, and the third step is to change the activation function. However, these three steps fail to improve accuracy.

Summary:

After these three major steps that could potentially improve model's accuracy failed, there is huge possibility that the problem lays in data preprocessing. Perhaps, collecting more useful data or removing irrelevant data can help with classification problems. After a glance at the data preprocessing, there are a lot of categories on application type, government organization classification, and income classification. Reducing them to a reasonable number of categories could mitigate unnecessary influences on the model analysis.