Deep Learning Homework - Charity Funding Predictor

Preprocessing the Data

Using my knowledge of Pandas and scikit-learn's StandardScaler(), I went ahead and preprocessed the dataset. This step prepared me for Step 2, where I would compile, train, and evaluate the neural network model. First, I removed any irrelevant information around the dataset, those being the EIN and NAME columns. This would only include the features that deemed appropriate for the model, even though for the final test I included back the NAME column. The data was then split into training and testing sets. The target variable for this specific model is "IS_SUCCESSFUL" and is verified by the value, 1 being yes and 0 being no. The APPLICATION data was what was analyzed, and the CLASSIFICATION value was used for binning. Then, the all categorical variables were encoded by the "pd.get_dummies()" function.

Compile, Train, and Evaluate the Model

The initial test only had two layers in total, but eventually was increase to three to optimize the model. Neurons were adjusted for each different test to effectively test the model effectively, with the first node layer being the highest value, and last being lowest. All three initial tests scored well below 75%, leading to our final test that was run to get the final test score above 75%. The three variables adjusted were number of nodes (increased), modified the thresholds for the APPLICATION_TYPE and CLASSIFICATION columns, changing the number of epochs, and adding the third hidden layer.

Summary and Final Optimization

For the fourth and final test, the NAME column was added back into the dataset, which achieved 78.4% accuracy, well above the original target, and had a total of 137,782 params.

I believe the Deep learning training model should always include multiple layers, and because it is machine learning based it teaches a computer to filter whatever inputs there are through the layers included to learn how to classify the data and information.