

Charity Funding Predictor

Nigel Itchon

Data Science Bootcamp, University of Pennsylvania

November 17, 2022

Overview

Deep learning is a subset of machine learning methods seeking to imitate the way humans gain knowledge. Based on a neural network, deep learning uses multiple layers to extract higher-level features from raw input. Using Alphabet Soup's dataset of 34,000 organizations that have received funding, the purpose of the analysis is to develop a tool that can help charity funding select applicants with the best chance of success in their ventures. The analysis follows the machine learning lifecycle beginning with preprocessing the data, walking through compiling, training, and evaluating the model. After the first pass, the model is optimized using the TensorFlow library to achieve a target predictive accuracy higher than 75%.

Results

The target variable for the funding predictor model is the Boolean "IS_SUCCESSFUL" in the provided dataset. Indicated by 0 or 1, the target variable identifies whether a specific venture is funded. The feature variables include application type, affiliation, classification, use case, organization, income amount and ask amount. The variables removed include EIN, name, status, and special considerations. These variables have no effect on the neural network, specifically the identifying features of EIN and name. The status feature is primarily populated with active status. There are few data points with special considerations that keeping the feature would not serve to accurately predict the targets. The data was filtered to only include ask

amounts less than \$10,000. The rest of the preprocessing binned unique values in feature columns as “Other,” if their counts were less than 500.

Using the Keras tuner to search for the optimal hyperparameters, the optimized and final neural network contains two hidden layers, 21 and 16 nodes respectively. Each hidden layer uses the hyperbolic tangent function as its activation function. The output layer uses the sigmoid function. The model achieved a model loss and model accuracy of 52.9% and 75.0% respectively. Achieving the requisite accuracy is attributed to the preprocessing steps of binning as well as filtering the data for ask amounts less than \$10,000. Prior to those steps, the accuracy score was 72.9%. The keras tuner also served to deliver the most optimal hyperparameters.

Summary

Using two hidden layers and hyperbolic tangent function, the deep learning model reaches a 75.0% accuracy score for ask amounts less than \$10,000. Amounts greater than \$10,000 may need a different classification approach, such as random forests or hierarchical clustering. The data should be broken up into relative groups of ask amounts because larger funding requests may have different considerations than smaller funding requests. Another recommendation is the further subdivide use cases or include more descriptive features to help the classifications. The scope of the use case feature appears to be limited, but would be helpful in determining targets.