**Overview**

The purpose of this analysis is to create an algorithm to predict whether or not applicants for funding a will be successful.

**Results**

Data Processing:

The target for this analysis is the “IS\_SUCCESSFUL” field of the data. The data is either a True (1) or False (0) if the applicant was successful.

The following fields were the features of the model: “APPLICATION\_TYPE”, “AFFILIATION”, “CLASSIFICATION”, “USE\_CASE”, “ORGANIZATION”, “STATUS”, “INCOME\_AMT”, “SPECIAL\_CONSIDERATIONS”, and “ASK\_AMT”.

The following variables were removed from the dataset: “EIN” and “NAME”.

Due to the quantity of variables in the “APPLICATION\_TYPE” and “CLASSIFICATION” fields, binning was performed to group several of the variables together that had low occurrences.

Compiling, Training, and Evaluating the Model

The model had two hidden layers and an output the layer.

The first hidden layer had 16 neurons and the “relu” activation function.

The second hidden layer had 16 neurons and the “relu” activation function.

The output layer had 1 neuron and the “sigmoid: activation function.

These were the settings to a similar activity during the class session, and thought this would be a good starting point, given that there would be some optimization.

First Analysis

Performing the first iteration of the model using the above parameters, the following was achieved:

|  |  |
| --- | --- |
| **Loss** | **Accuracy** |
| 0.55 | 0.72 |

The target was an accuracy of 0.75, so some optimization was attempted to achieve a better result.

Optimization

|  |  |  |
| --- | --- | --- |
| **Variable Changed** | **Loss** | **Accuracy** |
| First hidden layer, neurons changed from 16 to 32. | No change | No change |
| Epochs changed from 50 to 200 | No change | No change |
| Changed APPLICATION\_TYPE cutoff from 700 to 1000. Input dimension changed from 42 to 40. | No change | No change |
| Switched from relu to tanh. | No change | No change |
| Dropped STATUS and SPECIAL\_CONSIDERATIONS fields. | No change | No change |
| Random state changed from 78 to 20 | No change | No change |

The model did not change at all when each of these variables were changed individually. It was very surprising to see that the model did not fluctuate at all.

**Summary**

The original results from the neural network model showed some promise when the original accuracy was 72%. With this starting point, it was assumed that the accuracy could be improved with some modifications with the settings or a change in the preprocessing of the data. However, with changes to the neural network, and changes to the preprocessing of data, the results remained consistently 72%, with only a tenth of a percent difference at most.

Perhaps given the nature of the data, this is the best that can be expected from the result. It was surprising that dropping data did not alter the results at all, so perhaps the fields that were dropped, “STATUS” and “SPECIAL CONSIDERATIONS” did not have much influence on the target. Further preprocessing of the data could perhaps trigger the features with a more meaningful impact on the target. Also, perhaps it would be prudent to try a logistic regression model on this dataset as well.