

Overview of the analysis:

This analysis aims to evaluate the performance of a deep learning neural network model on funding applications for a charity organization. The goal is to predict whether a funding application will be successful or not based on various features and metadata.

Data Preprocessing:

The target variable for the model is the "IS_SUCCESSFUL" variable, indicating whether funding was used successfully.

The feature variables for the model include:

- APPLICATION_TYPE: Alphabet Soup application type
- AFFILIATION: Affiliated sector of industry
- CLASSIFICATION: Government organization classification
- USE_CASE: Use case for funding
- ORGANIZATION: Organization type
- INCOME_AMT: Income classification
- ASK_AMT: Funding amount requested.

The variables that should be removed from the input data are:

- EIN: Secondary identification variable for an organization
- SPECIAL_CONSIDERATIONS: Special considerations for the application
- STATUS: Active status

Compiling, Training, and Evaluating the Model:

The neural network model consists of three layers:

- First hidden layer: 7 Neurons with a ReLU activation function
- Second hidden layer: 14 Neurons with a ReLU activation function
- Output layer: 1 Neuron with a sigmoid activation function

These choices were made to create a model with sufficient capacity to capture the data patterns while keeping the computational demand manageable.

The target model performance was achieved, with an accuracy of 77%.

Steps taken to increase model performance:

Examination and removal of outliers in the ASK_AMT variable.

Retaining the NAME variable but binning values with a count less than 100 as "Other" to reduce the number of unique values.

Binning fewer values for the APPLICATION_TYPE and CLASSIFICATION variables as "Other" to simplify the data.

Reducing the number of neurons in the model to speed up computing time.

Summary:

The deep learning neural network model achieved the target performance of over 75% accuracy in classifying funding applications as successful or unsuccessful. Preprocessing steps, such as outlier removal and binning, contributed to the model's performance improvement. However, the inclusion of the NAME variable may lead to overfitting and bias towards known organizations.

To address this classification problem, a recommendation is to use a Random Forest model. Random Forest models are suitable for binary classifications and allow for assessing variable importance. This would enable the elimination of unnecessary variables and potentially enhance the model's performance further.