Ed Spiezio-Runyon

UTSA 21: Deep Learning Project
Charity Funding Success Predictor

Overview

Background:

The nonprofit foundation Alphabet Soup wants a tool that can help it select the applicants for funding with the best chance of success in their ventures. With your knowledge of machine learning and neural networks, you'll use the features in the provided dataset to create a binary classifier that can predict whether applicants will be successful if funded by Alphabet Soup.

From Alphabet Soup's business team, you have received a CSV containing more than 34,000 organizations that have received funding from Alphabet Soup over the years. Within this dataset are a number of columns that capture metadata about each organization, such as:

- **EIN** and **NAME** — Identification columns
- **APPLICATION_TYPE** — Alphabet Soup application type
- **AFFILIATION** — Affiliated sector of industry
- **CLASSIFICATION** — Government organization classification
- **USE_CASE** — Use case for funding
- **ORGANIZATION** — Organization type
- **STATUS** — Active status
- **INCOME_AMT** — Income classification
- **SPECIAL_CONSIDERATIONS** — Special considerations for application
- **ASK_AMT** — Funding amount requested
- **IS_SUCCESSFUL** — Was the money used effectively

Project Goal:

Design and optimize a neural network model that will achieve greater than 75% accuracy in predicting the success of charitable applicants based on their application, in order to best allocate Alphabet Soup's donation resources.

Ed Spiezio-Runyon

UTSA 21: Deep Learning Project
Charity Funding Success Predictor

Results

Data Pre-Processing:

Since they are neither targets nor features, the variables EIN and NAME have been removed from the input data. This has an added benefit of anonymizing the data and removing cross-references in the data for applicants who have previously received multiple donations. All remaining columns were considered in the analysis.

As deep learning techniques help teach a computer to filter through inputs to predict and classify information, deep learning models should consist of multiple neural layers. I chose to use three neural layers in the model, with 10, 20, and 30 neurons each, respectively, to give the computer a sufficient number of connections to filter.

Target variable:	IS_SUCCESSFUL.
Feature variables:	APPLICATION TYPE
	AFFILIATION
	CLASSIFICATION
	USE_CASE
	ORGANIZATION
	STATUS
	INCOME_AMT
	SPECIAL_CONSIDERATIONS
	ASK_AMT

Compiling, Training, and Evaluating the Model

Attempt 1:

Study variables:	APPLICATION_TYPE, CLASSIFICATION
Neural layers:	3 (with 10, 20, and 30 neurons)
Result:	72.4%
Success/Failure:	Failure

Ed Spiezio-Runyon

UTSA 21: Deep Learning Project
Charity Funding Success Predictor

Attempt 2:

Modification:	Changing study variables
Rationale:	Examine other variables for predictive connection
Study variables:	ORGANIZATION, INCOME_AMT
Neural layers:	3 (with 10, 20, and 30 neurons)
Result:	72.8%
Success/Failure:	Failure

Attempt 3:

Modification:	Changing study variables
Rationale:	Examine other variables for predictive connection
Study variables:	INCOME_AMT, ASK_AMT
Neural layers:	3 (with 10, 20, and 30 neurons)
Result:	72.4%
Success/Failure:	Failure

Attempt 4:

Modification:	Changing study variables
Rationale:	Examine other variables for predictive connection
Study variables:	APPLICATION_TYPE, INCOME_AMT
Neural layers:	3 (with 10, 20, and 30 neurons)
Result:	73.1%
Success/Failure:	Failure

Attempt 5:

Modification:	Return to original study variables, modify neural layers
Rationale:	Examine greater number of neural layers
Study variables:	APPLICATION_TYPE, CLASSIFICATION
Neural layers:	4 (10, 20, 40, and 80 neurons)
Result:	72.8%
Success/Failure:	Failure

Ed Spiezio-Runyon

UTSA 21: Deep Learning Project
Charity Funding Success Predictor

Attempt 6:

Modification:	Return to original study variables, re-add NAME
Rationale:	Examine original data with NAME to determine if model could make accurate prediction with cross-references to applicants who have previously received multiple awards.
Warning note:	This does remove the anonymity and returns the cross-connections with previous award winners who have received multiple donations. An improved accuracy in this case may not be properly predictive of future success for applicants.
Study variables:	APPLICATION_TYPE, CLASSIFICATION
Neural layers:	3 (with 10, 20, and 30 neurons)
Result:	71.4%
Success/Failure:	Failure

Summary

Although all attempts were within 4% and one was within 2% of the target accuracy, the deep learning model was unable to develop the requested level of accuracy for the project. It is possible that the dataset is not large enough to provide sufficient opportunity for the computer to discern a pattern. It is also possible that there is no discernable pattern to predict at or above the accuracy level of 75%, due to variables existing in the applicants' management of the projects and beyond Alphabet Soup's view.

Since the desired predictive outcome is binary – predicted success or failure – it is possible that a supervised machine learning model such as the Random Forest Classifier, combining decision trees to generate a classification outcome, may be more suitable for the project.