Charity Funding Predictor Analysis

Amina hajisaid

2022

# Charity Funding Predictor Analysis

# Overview

The non-profit foundation Alphabet Soup wants to create an algorithm to predict whether or not applicants for funding will be successful. With your knowledge of machine learning and neural networks, you’ll use the features in the provided dataset to create a binary classifier that is capable of predicting 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 the following:

* **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 consideration for application
* **ASK\_AMT**—Funding amount requested
* **IS\_SUCCESSFUL**—Was the money used effectively

# Results:

## Data Processing

Using libraries like Pandas, Scikit-Learn and TensorFlow the dataset (charity.csv) has been pre-processed, compiled, trained, evaluated and then optimised to achieve the highest possible accuracy.

The variable considered the target for the model is the column named **IS\_SUCCESSFUL** as it lists if the application was successful or unsuccessful. The variables that could be considered possible features of the model are **APPLICATION\_TYPE, AFFILIATION, ‘CLASSIFICATION’, ‘USE\_CASE’, ‘ORGANIZATION’, ‘STATUS’, ‘INCOME\_AMT’, ‘SPECIAL\_CONSIDERATIONS’, ‘ASK\_AMT’**

The EIN and NAME columns were then dropped from the model input data as they were non-beneficial ID columns; neither targets nor features.

## Compiling, Training, and Evaluating the Model

In the initial attempt, I selected two Hidden layers with 90, 40 neurons. Relu was used as the activating function for the first layer and sigmoid for the second and output layers.

Graphical user interface, text, application, email

Description automatically generated

However, this only achieved a model accuracy of 72.2% and failed to achieve the target model performance.

A picture containing diagram

Description automatically generated

For the second attempt, I aimed to the increase the number of neurons in the first layer to 100 and slightly decrease the second layer to 30. The activation functions were kept the same as the first attempt.

Graphical user interface, text, application, email

Description automatically generated

On the contrary, these changes resulted in only a slight increase in the model accuracy. The model performance was 72.5% at the end of the second attempt and it did not achieve the target.

A picture containing application

Description automatically generated

In the final and third attempt, I decided to add a third layer. The first two layers were kept the same as the second attempt and the third layer was added with 10 neurons. The activation functions were kept the same for the first two layers and the output layer. The activation for the third layer was set as sigmoid.

Text

Description automatically generated

Similar to the second attempt, the third layer in the final attempt did not result in the target model performance being achieved. The accuracy in the third attempt did not show a significant increase with only 72.7% being achieved.

A picture containing Word

Description automatically generated

## Summary

In summary, adding more hidden layers and increasing the number of neurons in the three attempts only resulted in a slight increase with regards to the accuracy of the model. However, the target model performance was not achieved. Perhaps, other variables need to be taken into consideration such as the input data to create more parameters; for example, the name column could be included. Another example is changing the number of bins.