Deep learning Challenge: Charity Funding Predictor

A homework assignment from the University of Birmingham Data Analytics Bootcamp (April 2022)

James Fairgrieve

# Overview

The non-profit foundation Alphabet Soup wants to create an algorithm to predict whether applicants for funding will be successful. The organisation has provided a CSV file containing more than 34,000 organisations that have received funding from Alphabet Soup over the years. Within this dataset are a number of columns that capture metadata about each organisation, such as the following:

* **EIN** & **NAME** – Identification columns
* **APPLICATION\_TYPE** – Alphabet Soup application type
* **AFFILIATION** – Affiliated sector of industry
* **CLASSIFICATION** – Government organisation classification
* **USE\_CASE** – Use case for funding
* **ORGANIZATION** – Organisation 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

The foundation wants the final model to achieve a final model accuracy of over 75%.

# Results

## Data Pre-processing

The obvious target in the model is the **IS\_SUCCESSFUL** column, as this is the indicator for whether an applicant is successful or not. The variables that could be considered to be potential features of the final model are:

* **NAME**
* **APPLICATION\_TYPE**
* **AFFILIATION**
* **CLASSIFICATION**
* **USE\_CASE**
* **ORGANIZATION**
* **INCOME\_AMT**
* **ASK\_AMT**
* **IS\_SUCCESSFUL**

At this stage, this leaves **EIN**, **STATUS** & **SPECIAL\_CONSIDERATIONS** as the remaining variables that can be removed.

For the model, there were two hidden layers used with 5 & 10 neurons. The final output layer had one unit as this model looked at the binary yes or no for whether the funding was successful.

Graphical user interface, text, application, email

Description automatically generated

Including the **NAME** variable to create more parameters in the model resulted in a higher accuracy than in previous attempts. Including this column resulted in a model accuracy of 77.3%, surpassing the target set by Alphabet Soup.

Table

Description automatically generated

# Summary

In this case, simply increasing the number of parameters in the model has increased the accuracy, but there are other ways that a more accurate model can be achieved:

* Adding more hidden layers
* Adding more neurons to the hidden layers