# Alphabet Soup Deep Learning Model for Funding Success Prediction

## Overview of the Analysis

The purpose of this analysis is to create a binary classification model using deep learning techniques to predict if an organization funded by Alphabet Soup will be successful in their venture. The model utilizes a dataset of over 34,000 organizations that have received funding from Alphabet Soup, containing metadata about each organization.

## Results

### Data Preprocessing

- \*\*Target variable(s) for the model:\*\* The target variable for the model is `IS\_SUCCESSFUL`.

- \*\*Feature variable(s) for the model:\*\* The feature variables for the model include `APPLICATION\_TYPE`, `AFFILIATION`, `CLASSIFICATION`, `USE\_CASE`, `ORGANIZATION`, `STATUS`, `INCOME\_AMT`, `SPECIAL\_CONSIDERATIONS`, and `ASK\_AMT`.

- \*\*Variable(s) removed from the input data:\*\* The `EIN` and `NAME` columns were removed from the input data as they are identification columns and not useful as features or targets.

- \*\*Feature variable\*\* `NAME` has been brought back in the last model.

### Compiling, Training, and Evaluating the Model

- \*\*Neurons, layers, and activation functions selected for the neural network model and rationale:\*\* The model consists of three hidden layers with 80, 30, and 1 neurons, respectively, and ReLU activation functions. The output layer uses a sigmoid activation function for binary classification. The structure was chosen to provide a balance between complexity and the potential for overfitting, while maintaining the ability to learn complex patterns in the data.

<img width="581" alt="Screenshot 2023-04-27 at 10 37 39 pm" src="https://user-images.githubusercontent.com/117792685/234865718-4ae56158-73e5-4324-b4db-c7ed0e045ada.png">

This model did not achive desired accuracy of 75%.

<img width="485" alt="Screenshot 2023-04-27 at 10 37 33 pm" src="https://user-images.githubusercontent.com/117792685/234865887-77b6367f-9b9b-4628-b0ee-ee7bf1b45db6.png">

In this project, we ran about 7 models. The first few models removed the EIN and NAME columns and with applying difirrent neurons and layers and binning just achived accuracy of 73%.

Model "3-AlphabetSoupCharity\_Optimisation" has the best accuracy of 78.85%.

<img width="482" alt="Screenshot 2023-04-27 at 10 37 25 pm" src="https://user-images.githubusercontent.com/117792685/234866318-e80f5f4a-bc89-43b5-bcae-2a1b9517d37c.png">

Our \*\*Feature variable(s) for the model:\*\* The feature variables for the model include `APPLICATION\_TYPE`, `AFFILIATION`, `CLASSIFICATION`, `USE\_CASE`, `ORGANIZATION`, `STATUS`, `INCOME\_AMT`, `SPECIAL\_CONSIDERATIONS`, `NAME` and `ASK\_AMT`. and \*\*Target variable(s) for the model:\*\* The target variable for the model is `IS\_SUCCESSFUL`. The model consists of three hidden layers with 14, 7, and 1 neurons, respectively, The output layer uses a sigmoid activation function for binary classification and used Relu for other layers.

<img width="595" alt="Screenshot 2023-04-27 at 10 37 20 pm" src="https://user-images.githubusercontent.com/117792685/234866452-5839f16d-3ea7-428b-899f-989777193357.png">

- \*\*Achievement of the target model performance:\*\* The model did achieve the target performance of 78.85% accuracy. However, multiple attempts were made to optimize the model, including adjusting input data, modifying the structure of the neural network, and modifying the training regimen.

- \*\*Steps taken in attempts to increase model performance:\*\* To increase model performance, the following steps were taken:

- Dropping additional irrelevant columns from the input data.

- Creating more bins for rare occurrences in columns and adjusting the number of values for each bin.

- Adding more neurons to a hidden layer.

- Adding or removing hidden layers.

- Using different activation functions for the hidden layers.

- Increasing or decreasing the number of epochs in the training regimen.

### Summary

The deep learning model achieved the desired performance of 78.84% accuracy in predicting the success of Alphabet Soup-funded organizations. Several attempts were made to optimize the model through data preprocessing and neural network structure adjustments, ultimately leading to this improved performance.

As an alternative recommendation, a Random Forest Classifier or a Gradient Boosting Classifier could be used to solve this classification problem. These ensemble learning methods have the potential to provide better accuracy and generalization without being as prone to overfitting as deep learning models. Further exploration of these alternative models may lead to even better performance in predicting the success of funded organizations.