**A**

**Report**

**on the**

**Analysis of Alphabet Soup Charity Funding Predictor**

**using**

**Deep Neural Network Model**

**Contents**

[**Introduction** 2](#_Toc146294330)

[**Purpose of the Analysis** 2](#_Toc146294331)

[**1.** **Data Preprocessing** 2](#_Toc146294332)

[*Initial Model* 2](#_Toc146294333)

[*Optimised Model* 2](#_Toc146294334)

[**2. Model Architecture** 3](#_Toc146294335)

[*Initial Model* 3](#_Toc146294336)

[*Optimised Model* 3](#_Toc146294337)

[**3. Model Performance** 4](#_Toc146294338)

[**4. Model Optimization Techniques** 4](#_Toc146294339)

[**5. Model Limitations** 4](#_Toc146294340)

[**6. Other Observations** 5](#_Toc146294341)

[**Alternative Models** 5](#_Toc146294342)

[**Summary** 5](#_Toc146294343)

**Analysis of Alphabet Soup Charity Funding Predictor using Neural Network Model**

# **Introduction**

# **Purpose of the Analysis**

The overarching goal of this analysis is to aid Alphabet Soup Charity in making informed decisions regarding funding applicants. Leveraging the power of neural networks, this analysis aims to predict which organizations are likely to use the funds effectively, thereby maximizing the impact of the charity's donations.

# **Data Preprocessing**

*Initial Model*

In the initial model, the target variable was ‘IS\_SUCCESSFUL’ column, which contained binary classification (0 or 1) indicating whether the organization used the funds effectively. The features were ‘APPLICATION\_TYPE’, ‘AFFILIATION’, ‘CLASSIFICATION’, ‘USE\_CASE’, ‘ORGANIZATION’, ‘INCOME\_AMT’, ‘ASK\_AMT’, ‘STATUS’, and ‘SPECIAL\_CONSIDERATIONS’.

In this model, only ‘EIN’, and ‘NAME’ columns were identified as variables that lacked predictive power. Hence, they were removed. Binning was done for only the categorical columns that had 10 unique values (‘APPLICATION\_TYPE’, and ‘CLASSIFICATION’).

*Optimised Model*

While the target column (‘IS\_SUCCESSFUL’) was the same as the original model, the feature variables changed. Additional variables like ‘STATUS’, and ‘SPECIAL\_CONSIDERATIONS’ were identified to lack variability; as such, were dropped. Thus, leaving the feature variables to only 7 including ‘APPLICATION\_TYPE’, ‘AFFILIATION’, ‘CLASSIFICATION’, ‘USE\_CASE’, ‘ORGANIZATION’, ‘INCOME\_AMT’, and ‘ASK\_AMT’. Also, outliers were removed from the numerical column ‘ASK\_AMT’.

Additionally, the ‘ASK\_AMT’, ‘APPLICATION\_TYPE’, and ‘CLASSIFICATION’ columns were binned.

# **2. Model Architecture**

*Initial Model*

This model comprised of an input layer, a hidden layer, and an output layer. The input feature consists of 43 neurons and the model was designed using ReLu activation function and sigmoid on the outer layer. Figure 1 shows the initial model architecture. The following outlines the activation function used in each layer.

* Input layer : 80 neurons, using the ReLU activation function.
* Hidden layer : 30 neurons, using the ReLU activation function.
* Output layer: 1 neuron, using the Sigmoid activation function (for binary classification).

A screenshot of a computer program

Description automatically generated

Figure 1: Initial model Architecture

*Optimised Model*

In the optimised model, additional layers were incorporated, and activation functions were varied. As well, experimentation was done with the learning rates, and the epochs. Figure 2 shows the model architecture for the optimised model including the activation function used in each layer.

A screenshot of a computer

Description automatically generated

Figure 2: optimised model Architecture

# **3. Model Performance**

Initial Model: Achieved an accuracy of approximately 72.92% with a loss of 56%.

Optimized Model: Achieved an accuracy nearing 75.09% with a loss of 56%.

# **4. Model Optimization Techniques**

Techniques like introducing more hidden layers, using different activation functions, experimenting with the number or epochs, and learning rate. were employed.

The dataset underwent further preprocessing, like binning rare occurrences of numeric variable and dropping less significant columns.

# **5. Model Limitations**

* The model is sensitive to the quality and quantity of data. Inaccuracies or biases in the data can influence predictions.
* Neural networks are challenging to optimise and tune.

# **6. Other Observations**

While accuracy is essential, ensuring that the model's recommendations align with the charity's objectives is equally crucial.

# **Alternative Models**

One could consider employing a *Random Forest* classifier for this problem. Reasons for this choice include:

* Interpretability: Random forests provide feature importance metrics, offering insights into which variables are most influential in predicting an organization's success.
* Handling Categorical Variables: Random Forest can inherently handle categorical variables without the need for one-hot encoding.
* Less Prone to Overfitting: Due to the ensemble nature of Random Forests, they are less likely to overfit compared to deep neural networks.
* Speed: Training can be faster, especially when the dataset is very large.

# **Summary**

The neural network model provided a reasonably accurate tool for Alphabet Soup Charity to assess potential funding candidates. It harnessed the power of deep learning to discern patterns and make predictions about the effective use of funds by organizations. However, traditional machine learning models like Random Forest can sometimes be more appropriate depending on the nature of the dataset and the specific requirements of the analysis.