

Alphabet Soup: Charity Funding Predictor Report

Overview of the Analysis

The purpose of this analysis is to develop a binary classifier capable of predicting whether applicants for funding by Alphabet Soup will be successful. By leveraging machine learning and neural networks, the analysis aims to identify patterns in the provided dataset to make accurate predictions about funding outcomes. Achieving a model accuracy of at least 75% is the target for success.

Results

Data Preprocessing

- **Target Variable:**
 - The target variable for the model is IS_SUCCESSFUL, where:
 - 1 indicates that the applicant was successful.
 - 0 indicates that the applicant was unsuccessful.
- **Feature Variables:**
 - The features for the model include:
 - EIN
 - NAME
 - APPLICATION_TYPE
 - AFFILIATION
 - CLASSIFICATION
 - USE_CASE
 - ORGANIZATION
 - STATUS
 - INCOME_AMT
 - SPECIAL_CONSIDERATIONS
 - ASK_AMT
- **Removed Variables:**

- The following variables were removed as they were deemed irrelevant or non-predictive:
 - EIN: An employer identification number with no inherent predictive value.
 - NAME: Initially removed, but later reintroduced for feature engineering purposes to group or bin rare values effectively.
-

Compiling, Training, and Evaluating the Model

- **Model Architecture:**

- The model was structured with three layers:
 1. **First Hidden Layer:** 5 neurons, ReLU activation function.
 2. **Second Hidden Layer:** 10 neurons, ReLU activation function.
 3. **Output Layer:** 1 neuron, Sigmoid activation function (for binary classification).

- **Model Parameters:**

- Total Parameters: 451
- Trainable Parameters: 451
- Non-Trainable Parameters: 0

- **Activation Functions:**

- ReLU was used in the hidden layers to introduce non-linearity and help the model learn complex patterns.
- Sigmoid was used in the output layer to output probabilities for binary classification.

- **Model Performance:**

- **First Attempt:**
 - Accuracy: 73%
 - Loss: Higher compared to the final model.
 - Did not meet the 75% target accuracy.
- **Second Attempt:**
 - Accuracy: 78.86%

- Loss: 0.4614
 - Surpassed the target accuracy of 75%, marking a significant improvement.
 - **Steps Taken to Improve Performance:**
 1. **Feature Engineering:** Reintroduced the NAME column for binning purposes, grouping rare or unique values into an “Other” category.
 2. **Optimization of Layer Structure:** Adjusted the number of neurons in the hidden layers to better capture the relationships between features.
 3. **Data Preprocessing:** Encoded categorical variables using one-hot encoding and scaled numerical variables to optimize training efficiency.
-

Summary

The deep learning model successfully achieved its goal of predicting funding outcomes, with the second model reaching an accuracy of 78.86%, exceeding the 75% target. The improvements made during feature engineering and model optimization were key contributors to this success.

Recommendation for Improvement

Multiple layers should be used for deep learning models since it learns how to predict and classify information based on filtering inputs through layers.