Alphabet Soup Deep Learning Model Report

Overview of the Analysis

The purpose of this analysis was to create, train, and evaluate a deep learning model using a neural network to predict whether an organization will be successful in securing funding from Alphabet Soup. By using historical data and applying classification techniques, we aim to help the non-profit organization make informed decisions about funding opportunities.

Results

Data Preprocessing

- Target Variable:
  - IS_SUCCESSFUL: This binary variable (1 = funded, 0 = not funded) is the classification target.

- Feature Variables:
  - All remaining relevant numeric and categorical variables (e.g., application type, classification, organization income category) were used after being encoded.

- Removed Variables:
  - EIN and NAME: These identifiers were dropped as they do not contribute to model performance.

Compiling, Training, and Evaluating the Model

- Neural Network Design:
  - Input Layer: Based on the number of input features (43 after encoding).
  - Hidden Layers:
    - Layer 1: 80 neurons, ReLU activation
    - Layer 2: 30 neurons, ReLU activation
  - Output Layer: 1 neuron, Sigmoid activation (for binary classification)

- Model Performance:
  - Accuracy achieved: ~72.5%
  - Target accuracy of 75% was not reached, but it was close.

Alphabet Soup Deep Learning Model Report

- Steps Taken to Improve Performance:
  - Adjusted the number of neurons and layers
  - Tried different activation functions (ReLU, tanh)
  - Applied additional dropout layers to prevent overfitting
  - Experimented with batch size and epochs
  - Normalized the input data

Summary and Recommendation

The deep learning model achieved moderate success, with an accuracy of ~72.5%. While this does not meet the 75% target, it does show that a neural network can reasonably classify funding outcomes. To improve performance, we recommend exploring alternative machine learning models such as Random Forest or Gradient Boosting (e.g., XGBoost). These models often outperform neural networks on structured/tabular data due to their ability to handle categorical features and nonlinear relationships more efficiently.