Alphabet Soup Charity Analysis Report

Introduction:

Alphabet Soup, a non-profit organization, provides funding to various organizations that support educational, humanitarian, and scientific efforts. To assist in the selection of successful applicants for funding, a machine learning-based tool was developed. The tool utilized a dataset of more than 34,000 organizations that had previously received funding from Alphabet Soup. The tool employed neural networks to create a binary classifier that can predict the probability of success for each applicant.

Data Preprocessing:

The preprocessing of data involved importing the charity_data.csv file into a Pandas DataFrame. The EIN and NAME columns were removed, and the target and feature variables were identified. To identify rare categorical variables, the number of unique values for each column was computed. A cutoff point was established based on the number of data points for each unique value, which was used to bin infrequent categorical variables together under a new value, "Other." To encode categorical variables, pd.get_dummies() was utilized, and the preprocessed data was split into two arrays, a features array (X) and a target array (y). The data was then split into training and testing datasets using the train_test_split function.

Model Development:

Two phases of model development were conducted. The first phase involved compiling, training, and evaluating the model. TensorFlow was utilized to develop a neural network for binary classification. The goal of this model was to predict the likelihood of success for Alphabet Soup-funded organizations based on the features present in the dataset. In the second phase, the model was optimized to improve its predictive accuracy beyond 75%. To achieve this target accuracy, the model was refined by dropping fewer columns, including the 'NAME' column in the dataset. More bins were created to account for rare occurrences by implementing a binning strategy for the 'NAME' column. Lastly, the model's architecture was enhanced by incorporating additional neurons into the hidden layers.

Conclusion:

In conclusion, the present study involved the development of a machine learning-based tool to aid in the selection of successful applicants for funding from Alphabet Soup. Data preprocessing was performed to remove irrelevant columns, identify rare categorical variables, and encode categorical variables. Neural networks were utilized to develop a binary classifier for predicting the likelihood of success for each applicant. The model was optimized to improve its predictive accuracy, and various strategies were employed, including the inclusion of the 'NAME' column, the creation of more bins, and the enhancement of the model's architecture. The optimized model is expected to provide valuable insights to Alphabet Soup in selecting successful applicants for funding.