Alphabet soup report

Overview of the Analysis

The purpose of this analysis was to develop a machine learning model capable of predicting the success of projects funded by a nonprofit foundation, Alphabet Soup. By utilizing historical data from over 34,000 projects that received funding, the goal was to create a binary classifier that could accurately identify projects with the best chance of success. This analysis involved preprocessing the data, designing a neural network model, compiling, training, evaluating its performance, and exploring various strategies to optimize the model's accuracy.

Data Preprocessing

Target Variable: The target for our model was `IS\_SUCCESSFUL`, indicating whether the funding was used effectively.

Feature Variables: Features included variables such as `APPLICATION\_TYPE`, `AFFILIATION`, `CLASSIFICATION`, `USE\_CASE`, `ORGANIZATION`, `STATUS`, `INCOME\_AMT`, `SPECIAL\_CONSIDERATIONS`, and `ASK\_AMT`, which provided various categorical and numerical insights into each application.

Variables Removed: The variables `EIN` and `NAME` were removed from the input data as they were identifiers that provided no predictive power for the analysis.

Compiling, Training, and Evaluating the Model

Neurons, Layers, and Activation Functions: The initial model consisted of two hidden layers with 80 and 30 neurons, respectively, using ReLU activation functions for hidden layers and a sigmoid activation for the output layer. This configuration was chosen to start with a relatively simple model that could be iteratively improved. Dropout layers were also included to reduce overfitting.

Achievement of Target Model Performance: The model achieved an accuracy of approximately 72.9%, falling short of the desired 75% accuracy target.

-Performance Improvement Attempts:

- Adjusted the number of neurons and layers.

- Implemented early stopping and learning rate schedules to optimize training.

- Experimented with different activation functions and dropout rates.

- Conducted feature engineering to refine the input data and tried different approaches for handling class imbalance.

- Explored different model architectures and hyperparameter tuning.

Summary

The deep learning model developed in this project demonstrated the potential of using neural networks for predicting the success of funded projects, achieving an accuracy close to 73%. Despite various optimization attempts, the model did not reach the target performance of 75% accuracy. This outcome suggests that while the neural network was able to capture complex patterns in the data, there might be limitations due to the inherent characteristics of the data or the chosen model architecture.

Recommendation for a Different Model

Gradient Boosting Machines (GBM): Given the tabular nature of the dataset, a Gradient Boosting Machine, such as XGBoost or LightGBM, could be more effective. GBMs have shown excellent performance on structured data and might capture the nuances of the dataset more effectively than a neural network due to their ability to handle varied feature types and interactions.

Reason for Recommendation: GBMs are highly efficient, less prone to overfitting, and require less data preprocessing compared to deep learning models. They also provide feature importance scores, which can offer insights into which variables most significantly impact project success, guiding further data collection and feature engineering efforts.

In conclusion, while the neural network provided valuable insights, exploring alternative models like GBMs could offer a path to achieving and potentially surpassing the desired accuracy threshold. Future work should also consider incorporating more diverse data sources, advanced feature engineering, and domain-specific knowledge to further enhance model performance.