Overview

In this project I have given database of The nonprofit foundation Alphabet Soup. The database contains 34,299 rows of data and 11 columns of data. Columns includes

- EIN and NAME—Identification columns
- APPLICATION_TYPE—Alphabet Soup application type
- AFFILIATION—Affiliated sector of industry
- CLASSIFICATION—Government organization classification
- USE_CASE—Use case for funding
- ORGANIZATION—Organization type
- STATUS—Active status
- INCOME_AMT—Income classification
- SPECIAL_CONSIDERATIONS—Special considerations for application
- ASK_AMT—Funding amount requested
- IS_SUCCESSFUL—Was the money used effectively

I have used IS_SUCCESSFUL column as a Target Variables

Results

I have created 6 models by adjusting inputs. To create various models, I Dropped some columns, created more bins for rare occurrences in columns, added more neurons to a hidden layer, Added more hidden layers etc

Model 1

Columns Dropped - EIN and NAME
Bins – 9 bins for APPLICATION_TYPE, 6 bins for CLASSIFICATION
Scaled data – X_train, X_test

Number of Hidden Layers and nodes – Layer 1 with 80 nodes, Layer 2 with 30 nodes
Activation functions – Input Layer 'Relu', Output Layer 'Sigmoid'
Epochs – 10
Accuracy – 52.61%

```
268/268 - 0s - loss: 0.7632 - accuracy: 0.5261 - 434ms/epoch - 2ms/step
Loss: 0.7632343173027039, Accuracy: 0.5260641574859619
```

Model 2

Columns Dropped - EIN and NAME
Bins – 9 bins for APPLICATION_TYPE, 6 bins for CLASSIFICATION
Scaled data – ASK_AMT column
(Added Scaled ASK_AMT to data frame then dropped Ask_AMT)
Number of Hidden Layers and nodes – Layer 1 with 50 nodes, Layer 2 with 30 nodes
Activation functions – Input Layer 'Relu', Output Layer 'Sigmoid'
Epochs – 100
Accuracy – 72.66%

```
268/268 - 1s - loss: 0.5658 - accuracy: 0.7266 - 507ms/epoch - 2ms/step
Loss: 0.5657655000686646, Accuracy: 0.7266472578048706
```

Model 3

Columns Dropped - EIN and NAME
Bins – 9 bins for APPLICATION_TYPE, 6 bins for CLASSIFICATION
Scaled data – X(all the features column)
Number of Hidden Layers and nodes – Layer 1 with 64 nodes, Layer 2 with 32 nodes
Activation functions – Input Layer 'Tanh', Output Layer 'Sigmoid'
Epochs – 200

Accuracy – 72.64%

```
268/268 - 0s - 2ms/step - accuracy: 0.7264 - loss: 0.5682
Loss: 0.5682398676872253, Accuracy: 0.7264139652252197
```

Model 4

Columns Dropped - EIN and NAME
Bins – 9 bins for APPLICATION_TYPE, 6 bins for CLASSIFICATION
Scaled data – X(all the features column)
Number of Hidden Layers and nodes – Layer 1 with 128 nodes, Layer 2 with 64 nodes, Layer 3 with 32 nodes
Activation functions – Input Layer 'Relu', Output Layer 'Sigmoid'
Epochs – 100
Accuracy – 72.89%

```
268/268 - 0s - 2ms/step - accuracy: 0.7289 - loss: 0.5805
Loss: 0.5804538130760193, Accuracy: 0.728863000869751
```

Model 5

Columns Dropped – EIN, NAME, STATUS, SPECIAL_CONSIDERATIONS
Bins – 8 bins for APPLICATION_TYPE, 6 bins for CLASSIFICATION
Scaled data – X(all the features column)
Number of Hidden Layers and nodes – Layer 1 with 128 nodes, Layer 2 with 64 nodes, Layer 3 with 32 nodes
Activation functions – Input Layer 'Relu', Output Layer 'Sigmoid'
Epochs – 50
Accuracy – 73.12%

```
268/268 - 0s - loss: 0.5597 - accuracy: 0.7312 - 477ms/epoch - 2ms/step
Loss: 0.559714674949646, Accuracy: 0.731195330619812
```

Model 6

Columns Dropped – EIN, STATUS, SPECIAL_CONSIDERATIONS
Bins – 96 bins for NAME, 8 bins for APPLICATION_TYPE, 6 bins for CLASSIFICATION
Scaled data – ASK_AMT column
(Added Scaled ASK_AMT to data frame then dropped Ask_AMT)
Number of Hidden Layers and nodes – Layer 1 with 128 nodes, Layer 2 with 64 nodes, Layer 3 with 32 nodes
Activation functions – Input Layer 'Relu', Output Layer 'Sigmoid'
Epochs – 100
Accuracy – 76.51%

```
268/268 - 1s - loss: 0.5594 - accuracy: 0.7651 - 501ms/epoch - 2ms/step
Loss: 0.559353232383728, Accuracy: 0.7651311755180359
```

Summary

This report evaluates six neural network models designed for predictive analysis, each varying in architecture, data preprocessing, and training parameters. The models consistently dropped columns like EIN and NAME except model 6, while some incorporated additional feature engineering such as scaling the ASK_AMT column. Model 1 achieved an accuracy of 52.61% with a simple structure, while subsequent models improved significantly, with Model 6 achieving the highest accuracy of 76.51% by utilizing three hidden layers and incorporating more relevant features.

The analysis highlights the importance of model complexity, feature selection, and activation functions in achieving better performance. As the models evolved, increasing the number of nodes and hidden layers led to higher accuracies, particularly in Models 4, 5, and 6. Additionally, longer training epochs and effective preprocessing contributed to improved results. Future efforts should focus on hyperparameter tuning and regularization techniques to enhance model robustness and prevent overfitting. In Future we can also try to use different supervised learning models like xgboost.