Module 21 Challenge

Deep Leaning Challenge: Charity Funding Predictor

Analysis overview

The objective of this analysis is to build and test an AI model to predict whether organizations that applied for funding from Alphabet Soup will succeed with their projects or not. This will help in choosing the organizations most likely to succeed. To achieve this, a neural network was used to sort funding requests into successful or unsuccessful categories, based on different features from the data.

	EIN	NAME	APPLICATION_TYPE	AFFILIATION	CLASSIFICATION	USE_CASE	ORGANIZATION	STATUS	INCOME_AMT	SPECIAL_CONSIDERATIONS	ASK_AMT	IS_SUCCESSFUL
0	10520599	BLUE KNIGHTS MOTORCYCLE CLUB	T10	Independent	C1000	ProductDev	Association	1	0	N	5000	1
1	10531628	AMERICAN CHESAPEAKE CLUB CHARITABLE TR	T3	Independent	C2000	Preservation	Co-operative	1	1-9999	N	108590	1
2	10547893	ST CLOUD PROFESSIONAL FIREFIGHTERS	T5	CompanySponsored	C3000	ProductDev	Association	1	0	N	5000	0
3	10553066	SOUTHSIDE ATHLETIC ASSOCIATION	T3	CompanySponsored	C2000	Preservation	Trust	1	10000-24999	N	6692	1
4	10556103	GENETIC RESEARCH INSTITUTE OF THE DESERT	T3	Independent	C1000	Heathcare	Trust	1	100000-499999	N	142590	1

Results

Data Preprocessing:

The target variable for the model is “IS_SUCCESSFUL”, which is a binary variable indicating whether the organization was successful in using the funding (1 for success, 0 for failure).

In the first model, the columns “EIN” and “NAME” were removed because they were considered non-useful identifiers for the prediction. However, in the second model, “NAME” was reintroduced as part of a binning process, along with the variables “CLASSIFICATION” and “APPLICATION_TYPE”. These features provide key information about the organization and the application, enhancing the model’s ability to predict the success of the funding.

	NAME	APPLICATION_TYPE	AFFILIATION	CLASSIFICATION	USE_CASE	ORGANIZATION	STATUS	INCOME_AMT	SPECIAL_CONSIDERATIONS	ASK_AMT	IS_SUCCESSFUL
0	BLUE KNIGHTS MOTORCYCLE CLUB	T10	Independent	C1000	ProductDev	Association	1	0	N	5000	1
1	AMERICAN CHESAPEAKE CLUB CHARITABLE TR	T3	Independent	C2000	Preservation	Co-operative	1	1-9999	N	108590	1
2	ST CLOUD PROFESSIONAL FIREFIGHTERS	T5	CompanySponsored	C3000	ProductDev	Association	1	0	N	5000	0
3	SOUTHSIDE ATHLETIC ASSOCIATION	T3	CompanySponsored	C2000	Preservation	Trust	1	10000-24999	N	6692	1
4	GENETIC RESEARCH INSTITUTE OF THE DESERT	T3	Independent	C1000	Heathcare	Trust	1	100000-499999	N	142590	1

Compiling, Training, and Evaluating the Model:

To achieve the target of at least 75% accuracy in the model, three hidden layers were used during training. These layers have progressively increased sizes with 7, 14, and 28 neurons each and employ the ReLU activation function. Additionally, an output layer with a single neuron and a sigmoid activation function is used for binary classification. This configuration is intended to capture complex patterns in the data and enhance the model’s ability to generalize.

Model: "sequential_1"

Layer (type)	Output Shape	Param #
dense_3 (Dense)	(None, 7)	721
dense_4 (Dense)	(None, 14)	112
dense_5 (Dense)	(None, 28)	420
dense_6 (Dense)	(None, 1)	29

Total params: 1,282 (5.01 KB)
Trainable params: 1,282 (5.01 KB)
Non-trainable params: 0 (0.00 B)

StandardScaler was applied to normalize the features and improve training convergence. The model was trained for 100 epochs, and a 15% validation split was used to prevent overfitting and ensure good generalization.

```
# Train the model
fit_model = nn.fit(X_train_scaled,y_train,validation_split=0.15, epochs=100)
```

```
Epoch 1/100
684/684 ━━━━━━━━━━━━━━━━━━━━ 4s 2ms/step - accuracy: 0.6686 - loss: 0.6082 - val_accuracy: 0.7681 - val_loss: 0.4814
Epoch 2/100
684/684 ━━━━━━━━━━━━━━━━━━━━ 2s 2ms/step - accuracy: 0.7555 - loss: 0.4938 - val_accuracy: 0.7769 - val_loss: 0.4734
Epoch 3/100
684/684 ━━━━━━━━━━━━━━━━━━━━ 3s 2ms/step - accuracy: 0.7686 - loss: 0.4789 - val_accuracy: 0.7676 - val_loss: 0.4737
Epoch 4/100
684/684 ━━━━━━━━━━━━━━━━━━━━ 3s 2ms/step - accuracy: 0.7636 - loss: 0.4825 - val_accuracy: 0.7709 - val_loss: 0.4721
Epoch 5/100
684/684 ━━━━━━━━━━━━━━━━━━━━ 2s 3ms/step - accuracy: 0.7638 - loss: 0.4802 - val_accuracy: 0.7738 - val_loss: 0.4716
Epoch 6/100
684/684 ━━━━━━━━━━━━━━━━━━━━ 3s 3ms/step - accuracy: 0.7654 - loss: 0.4761 - val_accuracy: 0.7714 - val_loss: 0.4727
Epoch 7/100
684/684 ━━━━━━━━━━━━━━━━━━━━ 2s 3ms/step - accuracy: 0.7671 - loss: 0.4778 - val_accuracy: 0.7753 - val_loss: 0.4722
Epoch 8/100
```

Summary

The deep learning model achieved an accuracy of 76.23%, surpassing the 75% target. This indicates that the model is effective at predicting whether organizations will successfully use the funding.

```
268/268 - 0s - 2ms/step - accuracy: 0.7623 - loss: 0.4877
Loss: 0.4877021908760071, Accuracy: 0.7623323798179626
```

While the current model performs well, there's potential for further improvement by experimenting with different models and techniques. Exploring options like ensemble methods, support vector machines (SVMs), or Random Forests could enhance overall performance and reliability. Additionally, more advanced neural network designs might help improve accuracy and handle the complex aspects of the classification task.