

Deep Learning Challenge: Charity Funding Predictor

Overview:

The non-profit foundation Alphabet Soup wants to create an algorithm to predict whether applicants for funding will be successful. With knowledge of machine learning and neural networks, we must use the features in the provided dataset to create a binary classifier that can predict whether applicants will be successful if funded by Alphabet Soup.

Results:

To begin the data processing, we removed any irrelevant information. After dropping EIN and NAME the remaining columns were to be considered features for the model. Name was added back into the second test for binning purposes. The data was then split for training and testing sets. The target variable for the model was labelled "IS_SUCCESSFUL" and has the value of 1 for yes and 0 for no. APPLICATION data was analysed and "CLASSIFICATION" value was used for binning. We used several data points as a cutoff to bin "rare" variables together with the new value of "Other" for each unique value. Categorical variables were encoded by get dummies () after checking to see if the binning was successful.

Compiling, Training, and Evaluating the Model:

There were three layers total for each model after applying Neural Networks. The number of hidden nodes were dictated by the number of features.

```
In [26]: # Define the model - deep neural net, i.e., the number of input features and
# hidden nodes for each layer.
num_input_features = len(x_train[0])
layer_1_nodes = 80
layer_2_nodes = 30

nn = tf.keras.models.Sequential()

# First hidden layer
nn.add(
    tf.keras.layers.Dense(units=layer_1_nodes,
                           input_dim=num_input_features,
                           activation='relu')
)

# Second hidden layer
nn.add(
    tf.keras.layers.Dense(units=layer_2_nodes,
                           input_dim=layer_1_nodes,
                           activation='relu')
)

# Output layer
nn.add(
    tf.keras.layers.Dense(units=1,
                           input_dim=layer_2_nodes,
                           activation='sigmoid')
)

# Check the structure of the model
nn.summary()
```

5,981 parameters were created by a three-layer training model. The first attempt was just over 73% accuracy which was under a desired 75% but not too far off

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 80)	3520
dense_1 (Dense)	(None, 30)	2430
dense_2 (Dense)	(None, 1)	31
Total params: 5,981		
Trainable params: 5,981		
Non-trainable params: 0		

```
In [22]: # Evaluate the model using the test data
model_loss, model_accuracy = nn.evaluate(x_test_scaled,y_test,verbose=2)
print(f"Loss: {model_loss}, Accuracy: {model_accuracy}")
```

```
268/268 - 1s - loss: 0.5542 - accuracy: 0.7347 - 1s/epoch - 5ms/step
Loss: 0.5542473196983337, Accuracy: 0.7346938848495483
```

Optimization:

The second attempt with the "NAME" column in the dataset, achieved an accuracy of almost 79%. This is 4% over the target 75% with 42,951 parameters

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 100)	39600
dense_1 (Dense)	(None, 30)	3030
dense_2 (Dense)	(None, 10)	310
dense_3 (Dense)	(None, 1)	11
Total params: 42,951		
Trainable params: 42,951		
Non-trainable params: 0		

Multiple layers should be used for deep learning models since it learns how to predict and classify information based on computer filtering inputs through layers

```
In [21]: # Evaluate the model using the test data
model_loss, model_accuracy = nn.evaluate(x_test_scaled,y_test,verbose=2)
print(f"Loss: {model_loss}, Accuracy: {model_accuracy}")

268/268 - 1s - loss: 0.4531 - accuracy: 0.7917 - 913ms/epoch - 3ms/step
Loss: 0.45312708616256714, Accuracy: 0.7917200922966003
```