Report

Overview of the analysis: Explain the purpose of this analysis.

Utilize the given data to predict successful funding.

- What variable(s) are the target(s) for your model?

  IS_SUCESSFUL

- What variable(s) are the features for your model?

  APPLICATION_TYPE; AFFILIATION; CLASSIFICATION; USE_CASE;ORGANIZATION; INCOME_AMOUNT; SPECIAL_CONSIDERATIONS

- What variable(s) should be removed from the input data because they are neither targets nor features? Dropped (EIN, ASK_AMT & STATUS)
  ASK_AMT -Highly imbalanced.
  STATUS- Not enough info the data
  EIN- Can be substituted by add NAME BACK

- How many neurons, layers, and activation functions did you select for your neural network model, and why?

  We notice by increase single layer neurons amount will boot the machine learning study run time. (Pick up accuracy faster)
  We notice add more layers cut off at 3 and above can increase accuracy.
  We must calibrate the epoch times- too many will overfit the training data; too low will cause not optimize the predict potential.

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 100)	28600
dense_1 (Dense)	(None, 100)	10100
dense_2 (Dense)	(None, 50)	5050
dense_3 (Dense)	(None, 120)	6120
dense_4 (Dense)	(None, 50)	6050
dense_5 (Dense)	(None, 10)	510
dense_6 (Dense)	(None, 10)	110
dense_7 (Dense)	(None, 10)	110
dense_8 (Dense)	(None, 1)	11

Total params: 56,661
Trainable params: 56,661
Non-trainable params: 0

- Were you able to achieve the target model performance?

  YES- Model 3 add back Name Test Model at 79.9% accuracy.

- What steps did you take in your attempts to increase model performance?

  1. Balanced Bin
  2. Add back future Name.
  3. Deepen the layer to 8.
  4. Reduce epochs to 100 to avoid overfit.
  5. Remove futures.

Summary

Name seems to be an important future since we do have reincurred application however it may not help Alphabet Soup since most likely it wont change name. We do notice Parent Booster USA has 1260 incidences, TOPS CWB INCS 765 incidences etc. Machine learning do be able to study the relation in a high dimensional space. It will help on the test data set due to the application in test data set will also contain applications from these names- if we assume the data is reasonably split.

Add complexity not necessary add predict power. In our case, 3 layers also can achieve similar result. Due to time constrain we stopped at 6 tiers. To simple the model also can further optimize our computing resource.