**Overview**

The purpose of this analysis was to create a neural network model for the non-for-profit company Alphabet Soup that would help them to select which applicants to provide funding for. Using historical data on companies that had previously been funded by Alphabet Soup a model was trained that was able to predict with 72.8 % accuracy whether a company would be successful if they were funded by Alphabet Soup.

**Results**:

* **Data Preprocessing**
  + **What variable(s) are the target(s) for your model?**

The target for the model was the IS\_Successful column which stated whether a company was successful with their venture after having received funding from Alphabet Soup.

* + **What variable(s) are the features for your model?**
* The features of the model are **EIN** and **NAME**—Identification columns
* **APPLICATION\_TYPE**—Alphabet Soup application type
* **AFFILIATION**—Affiliated sector of industry
* **CLASSIFICATION**—Government organisation classification
* **USE\_CASE**—Use case for funding
* **ORGANIZATION**—Organisation type
* **STATUS**—Active status
* **INCOME\_AMT**—Income classification
* **SPECIAL\_CONSIDERATIONS**—Special considerations for application
* **ASK\_AMT**—Funding amount requested
  + **What variable(s) should be removed from the input data because they are neither targets nor features?**

The two variables that were removed from the input data were NAME and EIN as they only provided ID information and nothing that would help to determine if a funding application was successful or not.

* **Compiling, Training, and Evaluating the Model**
  + **How many neurons, layers, and activation functions did you select for your neural network model, and why?**

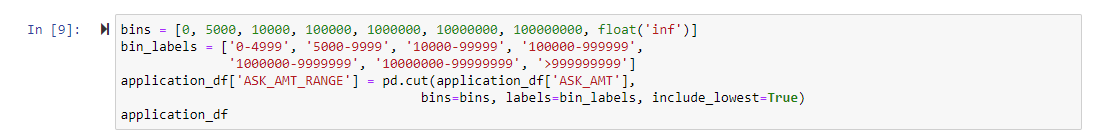
For the final model hyperparameter tuning was performed on the training data to determine the best activation function, number of neurons in the first and hidden layers and the number of hidden layers. The top model was chosen and that had 5 hidden layers, 26 neurons in the first layer and 15 neurons in the hidden layers. The activation was ‘relu’ and the output was sigmoid and one neuron as we wanted a binary output, was the application successful yes or no.

* + **Were you able to achieve the target model performance?**

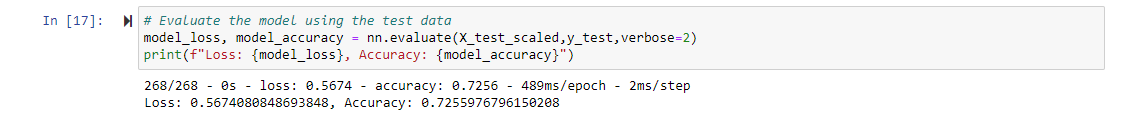
No the target model performance was not achieved. The highest accuracy achieved was 0.730962.

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

The first step taken to increase the model performance was to bin the ASK\_AMT column as this had over 8000 unique values. So the ASK\_AMT column was split into 7 bins in the ASK\_AMT\_RANGE column and the ASK\_AMT column was dropped:



Using the dame parameters as in the starter file the model was defined and trained but produced a lower accuracy of 0.7256.



So the original ASK\_AMT column was reinstated and the range column was removed. Next to optimise the model hyperparameter tuning was carried out to find the optimal activation, number of layers and number of neurons for the model:

A screenshot of a computer program

Description automatically generated

The best model parameters were then used to create another model using the training data and the model was assessed to have 0.730962 accuracy after 100 epochs which was an improvement. This same model was then tested using 150 EPOCS which decreased the accuracy to 0.7307.

A screenshot of a computer program

Description automatically generated

The model was then trained with 50 EPOCS and the accuracy was analysed using the testing data which produced a lower accuracy of 0.7275 so the optimisation was halted here.

1. **Summary**: Summarise the overall results of the deep learning model. Include a recommendation for how a different model could solve this classification problem, and then explain your recommendation.

In summary this analysis failed to produce a model that was able to predict the outcome of a funding applicants success that was over 75% accurate, therefor it is not recommended that this model be used in practice as it could yield moderate results. Whilst the model did improve with the addition of more hidden layers the improvement was small and so it would be recommended that another model be used to answer this question. A Random Forrest model could be better used to solve this problem. This model is a lot more user friendly using decision trees and combining them to make predictions, this makes it easier to visualize the different importance of each of the features, they also require less training time and so would be faster to optimize and alter. Random trees also don’t assume that the data is linear and can be used for non-linear data. Using a model like random trees would give a lot more insight into what the data was portraying which could then be used to make more informed decisions when trying to optimise the neural network.