**Performance Review – Alphabet Soup Application Success Predictor**

1. **Overview**

A deep learning model has been created for the nonprofit foundation Alphabet Soup.

The purpose of this model is to predict whether a funding application such as money required for product development, preservation, health care or community service is likely to be successful or unsuccessful. However, the model needs to be accurate and trustworthy once generalised and therefore a reviewal of the steps taken to create the model and it’s performance will be undertaken to help create transparency and evaluate the model created.

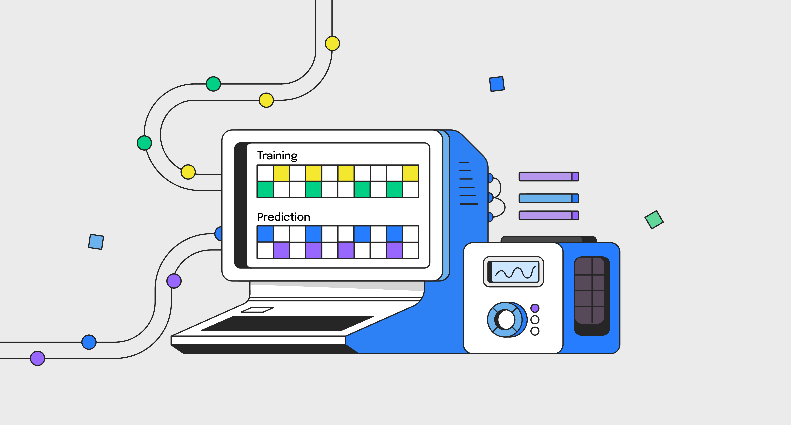
[(Ardis 2023)](https://www.programmersinc.com/application-modernization-benefits/)

1. **Results**:

Data Preprocessing

[(Admin 2023)](https://academy.maculogix.com/news-events/one-size-does-not-fit-all/)

* **Target variable:** IS\_SUCCESSFUL – a binary classification of if an application is likely to result in the money being used effectively [successful (1) or unsuccessful (0)].

[(Barkved 2022)](https://www.obviously.ai/post/the-difference-between-training-data-vs-test-data-in-machine-learning)

* **Features:** The variables which are fed into the model to aid prediction:
  + APPLICATION\_TYPE – The type of Alphabet Soup application
  + AFFILIATION – The affiliated sector of industry
  + CLASSIFICATION – Government organisation classification
  + USE\_CASE – Use case for funding
  + ORGANIZATION – Organization type
  + INCOME\_AMT – Income classification
  + SPECIAL CONSIDERATION – Special consideration for the application
  + ASK\_AMT – Funding amount requested

([iStock 2020](https://www.istockphoto.com/photo/white-puzzle-with-piece-that-does-not-fit-on-a-green-background-gm1278321226-377300039))

* **Removed variables:** Removed from the input data because they were not considered to be either a target or a feature within the optimized model:
  + EIN – Identification
  + NAME – Identification
  + STATUS – Active status

Compiling, Training, and Evaluating the Model

* **Neurons:** 11 neurons were used in the first hidden layer and 28 neurons were used in the second hidden layer.
  + A general rule states that 2-3 x more neutrons than the input features should be used. However, with 42 input features, it was preferable to avoid this rule whilst still achieving the desired accuracy as more computational resources would be required.
  + A Keras auto tuner was ran which trialled the possible hyperparameter combinations on the original model created. The suggested number of neurons in the results were used as guidance which returned a near respectable accuracy.
  + However, some changes were made to composition of the features within the optimized model, and due to time constraints another Keras autotuner was not ran. Therefore, to help stem efficiency on the optimised model, the number of neurons was increased in the second layer to help stem more communications between the computations which were taking place on each node due to the complexity of the input features.
* **Layers:** 3layers were used
  + There is an ongoing argument within the data scientist community that even the most complex interactions can be characterized by as few a 3 hidden layers.
  + 3 was sufficient in considering enough of the interactions between variables. Although some of the Keras autotuner results suggested to use more than 3 layers, with each additional layer comes more computational resources and therefore for a non-profit organization, 3 was adequate.
* **Activation Functions:** Layer 1 - Tanh, Layer 2 – Relu, Layer 3 - Sigmoid
  + Layer 1 – Tanh: The Tanh activation function was used as it can be used for classification or regression. It was able to take in the input features and transform the output to a range between -1 and 1. On the initial data set, it helps to classify data into one of two classes which would then be fed into the next layer where more complex patterns could be distinguished.
  + Layer 2 – Relu: The rectified linear unit activation function was used as it is ideal for modelling non-linear input data for classification or regression. All the -1 values from the first layer would be changed to 0 at this stage which still yielded an accuracy of 75% so this combination seemed to work well with the input features.
  + Layer 3 – Sigmoid: Generally, a less complex activation should be used on the output layer as the output is going to take the form of a simple binary classification. By this layer, only the most meaningful synapse communications would reach this layer. Therefore, the sigmoid activation function was selected due to its ability to simply predict probabilities and using them to categorise an s-curve.
* **Model Performance:**
  + Once optimised, the 75% target performance was met.
  + The initial model did not reach this 75% accuracy benchmark on the testing data, it instead returned an accuracy of 72.6%. Therefore, the following steps were taken to increase model performance:
    - A screen shot of a graph

      Description automatically generatedAfter plotting a box plot of all the ask amounts in the dataset, it was apparent that there were some potential outliers which could have been hindering the models ability to accurately predict whether an application would be successful. Therefore, any data points which exceeded the computed upper and lower bound were removed from the dataset.

Removed

* + - A screenshot of a computer

      Description automatically generatedA screenshot of a computer

      Description automatically generatedOn the original model, STATUS was considered as a feature. However, this column was dropped upon optimization to reduce noise as the machine learning algorithm appeared to be learning bad habits from this feature.

OPTIMIZED

ORIGINAL

* + - A screenshot of a computer code

      Description automatically generatedA screenshot of a computer

      Description automatically generatedThe number of values in the ‘other’ bin was proportionately increased for the CLASSIFICATION feature. On the original model, any values which represented less than 1% of all applications were binned into an ‘Other’ bin. However, this was increased to 3% in the optimized model.

OPTIMIZED

ORIGINAL

Please note that as the ask amount outliers were dropped, the total number of applications classified as ‘other’ reduced in the optimized model but proportionally was increased by 2% from the original model.

* + - The unoptimized model has 21 neurons in the second hidden layer, this was increased to 28 in the optimized model.
    - A screen shot of a computer

      Description automatically generatedWithin the unoptimized model, the first layer has an activation of relu, however, this was changed to tanh in the first layer of the optimized model.

OPTIMIZED

ORIGINAL

* + - The number of epochs was deliberately not increased due to the amount of input features, if the epochs was significantly increased, it ran the risk of overfitting.

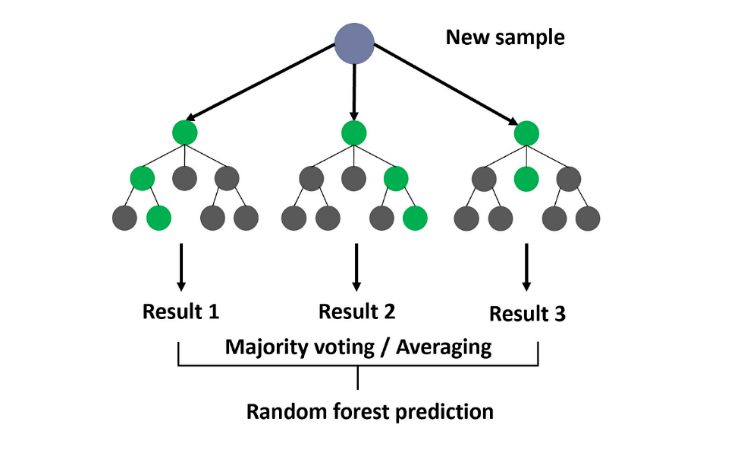
OPTIMIZED

ORIGINAL

1. **Summary**:

Overall results

* Overall, the optimised model will correctly predict whether a funding application will be successful or not 75% of the time; out of every 100 applicants, the model would not be able to correctly predict 25 application’s success level correctly.
* This final result has marginally met the requested level of accuracy but still has room for improvement.
  + The models architecture has been trialled across many suitable instances for the context, and to further improve the current model further would require many more layers which comes with associated costs due to the computer resources they requires.
  + The features could be explored further, finding ways to alter such as reinventing the binning structure so it does not contain another category for ‘other’ values but instead bins each unique value into broader groups could potentially assist the model to reach a higher accuracy.
  + Further features may be required, for example a breakdown of how the funding budget will be used may be much more insightful whilst training the model compared to the some of the current features.

Recommendation

[(Yehoshua 2023)](https://medium.com/@roiyeho/random-forests-98892261dc49)

* An alternative to the produced deep learning model would be the utilisation of the random forests supervised machine learning ensemble method.
  + Firstly, as the data set is labelled, this supervised technique would be beneficial.
  + Our data had 42 input features, random forests could help solve this classification problem because of it’s ability to sample the data and build several smaller, simpler decision trees. Although each tree is a weak learner as it is only trained on a small piece of original data, together, they can form a strong classifier.
  + This model could be good at predicting success because it is robustness to outliers which this data set was found to have and also to nonlinear data.
  + Random forests can run efficiently on large databases where it can keep 000’s of input variables. However, to help aid the model, it is possible to obtain feature importance which could really help when creating the model as the less important features could be omitted.

**References**

Ardis, J. (2023, August 15). 8 Benefits of Application Modernization. Programmersinc. <https://www.programmersinc.com/application-modernization-benefits/>

Admin, J. P. L. (2018, August 31). One Size Does Not Fit All - What Raising Kids and Practice Workflow Have in Common. MacuLogix Academy. <https://academy.maculogix.com/news-events/one-size-does-not-fit-all/>

Inc. (2020). IStock. <https://www.istockphoto.com/photo/white-puzzle-with-piece-that-does-not-fit-on-a-green-background-gm1278321226-377300039>

Barkved, K. (2022). Wwobviously.ai. <https://www.obviously.ai/post/the-difference-between-training-data-vs-test-data-in-machine-learning>

Yehoshua, D. R. (2023, July 17). Random Forests. Medium. <https://medium.com/@roiyeho/random-forests-98892261dc49>

‌