For this work I build a model for a nonprofit, to select the applicant who have the best chances to get the support they need.

The model reaches 77% of accuracy on predict who could be a good candidate.

The columns contained in the data are:

```
EIN
NAME
APPLICATION_TYPE
AFFILIATION
CLASSIFICATION
USE_CASE
ORGANIZATION
STATUS
INCOME_AMT
SPECIAL_CONSIDERATIONS
ASK_AMT
IS_SUCCESSFUL
```

In the first and second model reached 72%. In these models the columns “NAME” and “EIN” were discarded. In the first model were used three layers with the following configuration:
First hidden layer, 7 nodes.
Second hidden layer, 14 nodes.
Third hidden layer, 21 nodes.

As an experiment the number of nodes were adjusted as follows:
First hidden layer, 8 nodes.
Second hidden layer, 16 nodes.
Third hidden layer, 25 nodes.

After this change the accuracy of the model didn’t improve so the approach changes to consider the column ‘NAME’ in the model.

The third try improved from 72% to 77.73% of accuracy. In this model we included the column “NAME.

```
# 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 - 0s - loss: 0.4805 - accuracy: 0.7773 - 335ms/epoch - 1ms/step
Loss: 0.48050156235694885, Accuracy: 0.7772594690322876
```

With the difference in the accuracy, it can be assumed that “NAME” data is important for the improvement in the accuracy of the model, but it need to review the overall distribution of the data. Because the agglomeration on the data could have an impact in the overall outcome of the model.

In summary, the data contained in the dataset is important for the correct functionality of the model. In the other hand, it could be a good idea check the correlation between the “NAME” column and the “IS_SUCCESFUL” column to gain a better understanding of the model’s functionality.