Overview

The project's objective is to develop and optimize a deep learning model to predict the success of charitable donations. By analyzing various features related to organizations, the model aims to determine the likelihood of a donation's success, thereby assisting in making informed funding decisions.

Data Preprocessing

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

The target variable for the model is the success of a donation, typically represented as a binary outcome indicating whether a donation was successful or not.

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

Features include various attributes of the organizations seeking donations, such as:

- Organization's financial information
- Demographic details
- Historical donation data
- Other relevant characteristics that might influence donation success
- - What variable(s) should be removed from the input data because they are neither targets nor features?

Any identifiers or non-informative columns that do not contribute to the predictive capability of the model should be removed, such as:

- Unique IDs
- Non-informative categorical variables
- Columns with a high percentage of missing values

Compiling, Training, and Evaluating the Model

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

In the optimized notebook, there were 3 hidden layers, the first using 80 neurons, then 30 and 10, respectively.
Relu was used for the hidden layers and sigmoid was used for the output layer. Relu activation helps with non-linear relationships while sigmoid gives a probability output.

- Were you able to achieve the target model performance?

The optimized model proved to be more accurate than the initial model going from roughly 72% to 79% accuracy.

Initial:

```
268/268 - 1s - 2ms/step - accuracy: 0.7248 - loss: 0.5642
Loss: 0.5641876459121704, Accuracy: 0.724781334400177
```

Optimized:

```
268/268 - 1s - 3ms/step - accuracy: 0.7897 - loss: 0.4615
Loss: 0.4614519476890564, Accuracy: 0.7897375822067261
```

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

To increase the model performance, more relevant features were selected. New features were created through combinations or transformations, and parameters were adjusted.

Initial:

	EIN	NAME	APPLICATION_TYPE	AFFILIATION	CLASSIFICATION	USE_CASE	ORGANIZATION	STATUS	INCOME_AMT	SPECIAL_CONSIDERATIONS	ASK_AMT	IS_SUCCESSFUL
0	10520599	BLUE KNIGHTS MOTORCYCLE CLUB	T10	Independent	C1000	ProductDev	Association	1	0	N	5000	1
1	10531628	AMERICAN CHESAPEAKE CLUB CHARITABLE TR	T3	Independent	C2000	Preservation	Co-operative	1	1-9999	N	108590	1
2	10547893	ST CLOUD PROFESSIONAL FIREFIGHTERS	T5	CompanySponsored	C3000	ProductDev	Association	1	0	N	5000	0
3	10553066	SOUTHSIDE ATHLETIC ASSOCIATION	T3	CompanySponsored	C2000	Preservation	Trust	1	10000-24999	N	6692	1
4	10556103	GENETIC RESEARCH INSTITUTE OF THE DESERT	T3	Independent	C1000	Heathcare	Trust	1	100000-499999	N	142590	1

Optimized:

```
# Drop the non-beneficial ID columns, 'EIN' and 'NAME'.
application_df = application_df.drop(["EIN", "STATUS", "SPECIAL_CONSIDERATIONS"], axis=1)
application_df.head(2)
```

	NAME	APPLICATION_TYPE	AFFILIATION	CLASSIFICATION	USE_CASE	ORGANIZATION	INCOME_AMT	ASK_AMT	IS_SUCCESSFUL
0	BLUE KNIGHTS MOTORCYCLE CLUB	T10	Independent	C1000	ProductDev	Association	0	5000	1
1	AMERICAN CHESAPEAKE CLUB CHARITABLE TR	T3	Independent	C2000	Preservation	Co-operative	1-9999	108590	1

Initial:

Model: "sequential_2"

Layer (type)	Output Shape	Param #
dense_6 (Dense)	(None, 80)	3,520
dense_7 (Dense)	(None, 30)	2,430
dense_8 (Dense)	(None, 1)	31

Total params: 5,981 (23.36 KB)
Trainable params: 5,981 (23.36 KB)
Non-trainable params: 0 (0.00 B)

Optimized:

Model: "sequential_3"

Layer (type)	Output Shape	Param #
dense_9 (Dense)	(None, 100)	39,600
dense_10 (Dense)	(None, 30)	3,030
dense_11 (Dense)	(None, 10)	310
dense_12 (Dense)	(None, 1)	11

Total params: 42,951 (167.78 KB)
Trainable params: 42,951 (167.78 KB)
Non-trainable params: 0 (0.00 B)

Summary

The deep learning model was developed to predict the success of charitable donations by analyzing various organizational features. Through various steps and methods, the model was optimized to a near 7% increase in accuracy. The Random Forest model was applied to handle interactions effectively because it is less prone to overfitting with proper tuning. The Random Forest model resulted in an accuracy score of about 77%, which although is not higher than the optimized model, is still higher than the initial one.

```
Random forest model accuracy: 77.19
```