

Machine Learning

Class Project

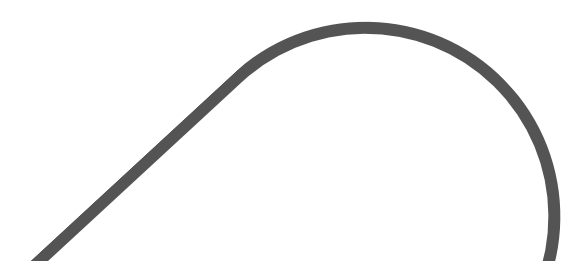
Presented by Abdallah Chidjou

Maximizing Marketing Efficiency with Predictive Modeling

A Case Study on Customer Term Deposits

KEY POINTS

- **Business Context :** The dataset use here is from a financial institution's marketing campaign for banking products.
- **Main Objective:** Predict the likelihood of a customer accepting a term deposit offer.



DATASET OVERVIEW

- **Categorical Feature:** Include the following; ('job', 'marital', 'education', 'default', 'housing', 'loan', 'contact', 'month', 'poutcome', 'y')
- **Numerical Feature:** Include the following; ('age', 'balance', 'day', 'duration', 'campaign', 'pdays', 'previous')

Target Variable:

The target variable here, is ('y'); indicates term deposit subscription (Yes/No).

This is what the dataset Header looks like

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
0	30	unemployed	married	primary	no	1787	no	no	cellular	19	oct	79	1	-1	0	unknown	no
1	33	services	married	secondary	no	4789	yes	yes	cellular	11	may	220	1	339	4	failure	no
2	35	management	single	tertiary	no	1350	yes	no	cellular	16	apr	185	1	330	1	failure	no
3	30	management	married	tertiary	no	1476	yes	yes	unknown	3	jun	199	4	-1	0	unknown	no
4	59	blue-collar	married	secondary	no	0	yes	no	unknown	5	may	226	1	-1	0	unknown	no
5	35	management	single	tertiary	no	747	no	no	cellular	23	feb	141	2	176	3	failure	no
6	36	self-employed	married	tertiary	no	307	yes	no	cellular	14	may	341	1	330	2	other	no
7	39	technician	married	secondary	no	147	yes	no	cellular	6	may	151	2	-1	0	unknown	no
8	41	entrepreneur	married	tertiary	no	221	yes	no	unknown	14	may	57	2	-1	0	unknown	no
9	43	services	married	primary	no	-88	yes	yes	cellular	17	apr	313	1	147	2	failure	no

Shape: (4521, 17)

Default

- Description: Indicates whether the customer has credit in default (yes or no).
- Significance: Customers with a history of defaults might be less likely to subscribe to additional financial products

campaign

- Description: Number of contacts performed during this campaign.
- Significance: More frequent contacts could increase awareness, but there may also be a saturation point where additional contacts become ineffective.

previous

- Description: Number of contacts made before this campaign.
- Significance: Indicates how persistent past attempts were, which may impact a customer's current decision.

Duration

- Description: Duration of the last contact in seconds.
- Significance: The length of contact has shown to be a strong indicator of whether a customer is interested in the product, with longer calls being more likely to lead to positive outcomes.

outcome

- Description: Outcome of the previous marketing campaign (e.g., success, failure, unknown).
- Significance: Past success or failure could influence a customer's perception of the product and their likelihood of subscribing.

y (Target Variable)

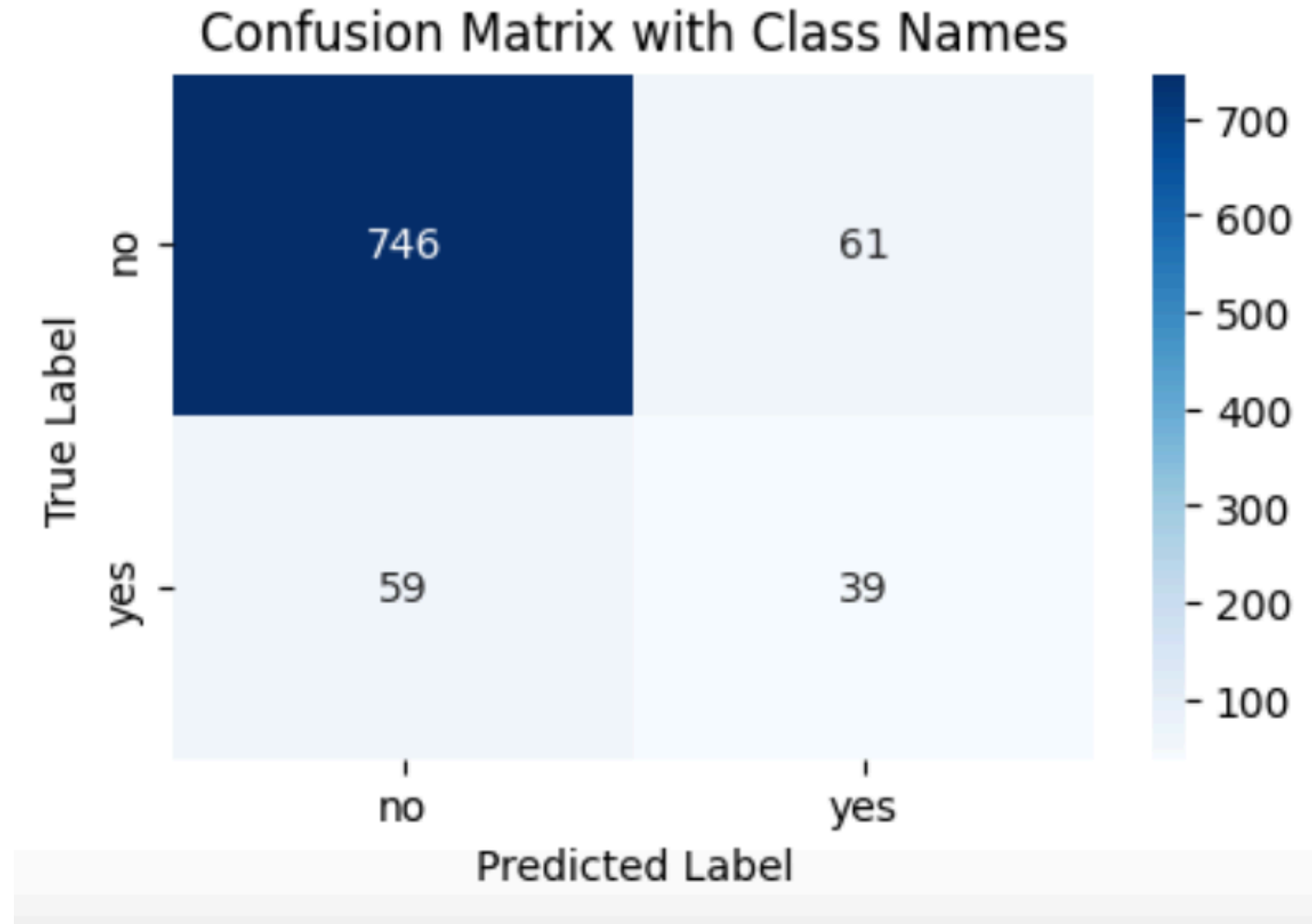
- Description: Whether the customer subscribed to a term deposit (yes or no).
- Significance: This is the target variable we want to predict using the other features in the dataset.

MODELING APPROACH

- Model Selection: Decision Tree chosen for interpretability and robustness.
 - I also put in use:
 - Logistic Regression
 - Random Forest
 - Support Vector Machine
 - K-Nearest Neighbors
 - Gradient Boosting
 - Neural Network
- Train-Test Split: "80-20 split for training and evaluation."

MODEL EVALUATION

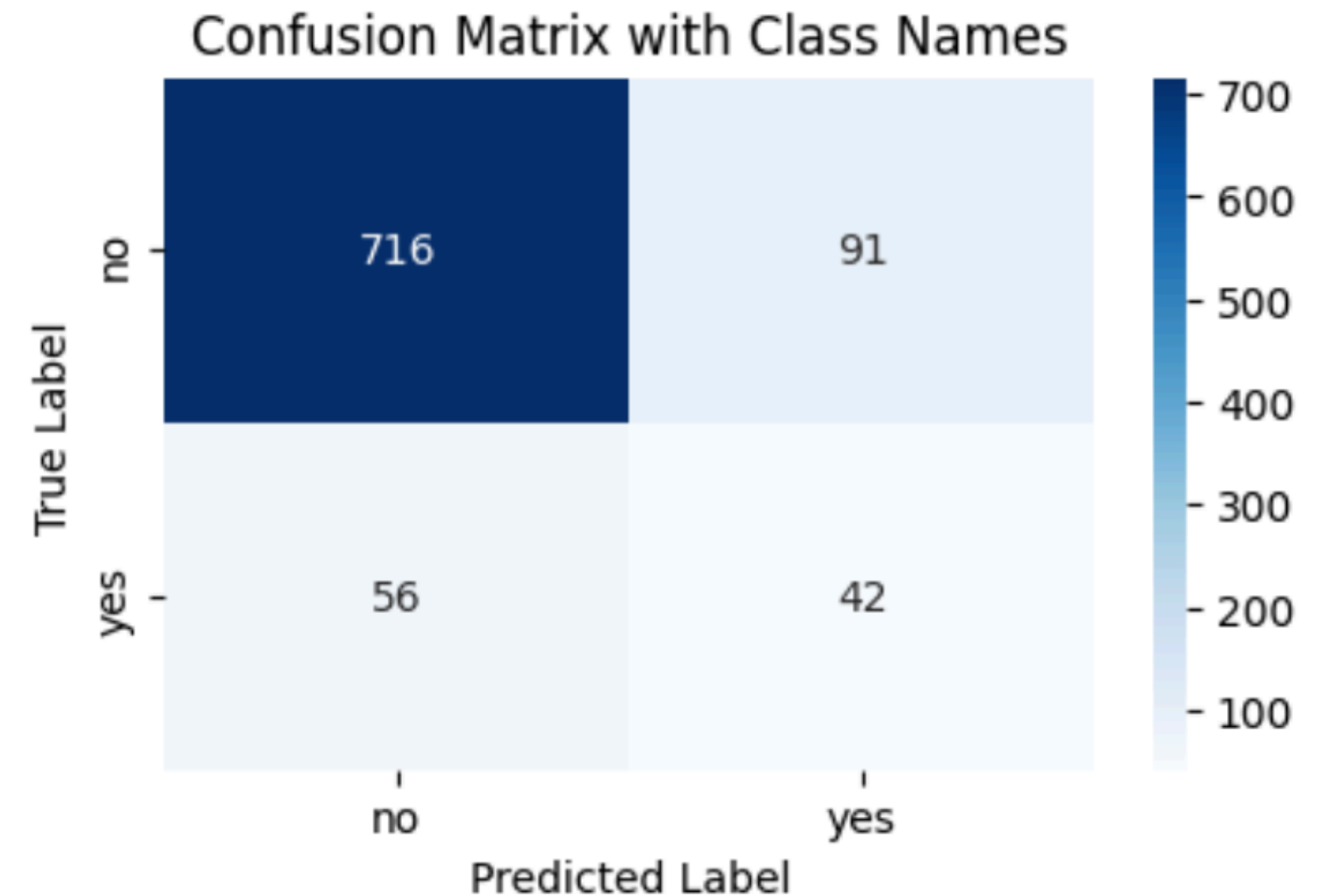
- Performance Metrics: Evaluated using accuracy, precision, recall, and F1-score.
- Key Results: Model achieved 86% accuracy but needs improvement for the minority class. (Decision Tree)



PERFORMANCE IMPROVEMENT

After applying SMOTE to handle the class imbalance, the accuracy level this time dropped to 84%, but looking at the previous confusion matrix, it shows some modest improvement.

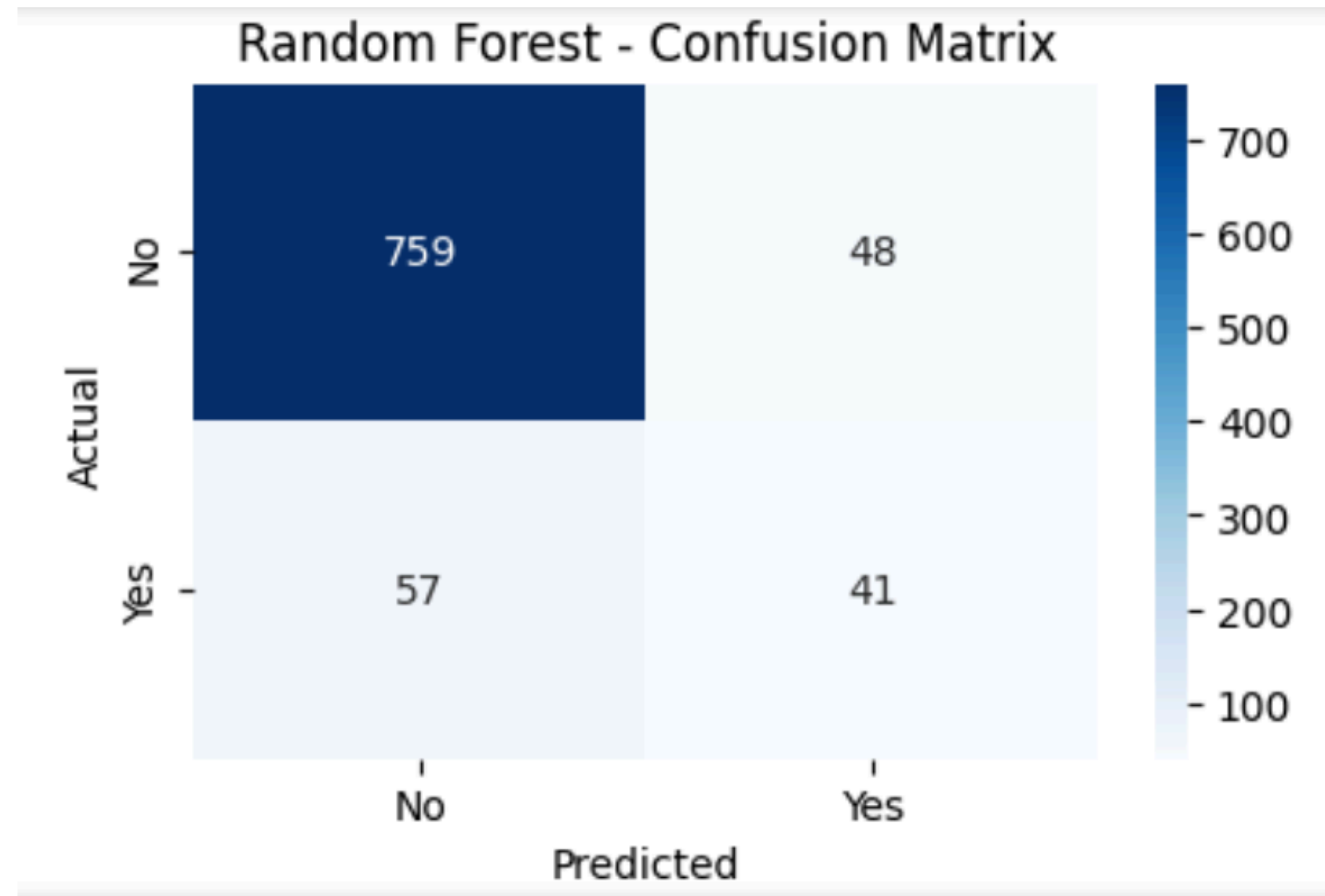
(Decision Tree)



PERFORMANCE IMPROVEMENT

After use of several models, I was able to get a slightly better accuracy with Random Forest.

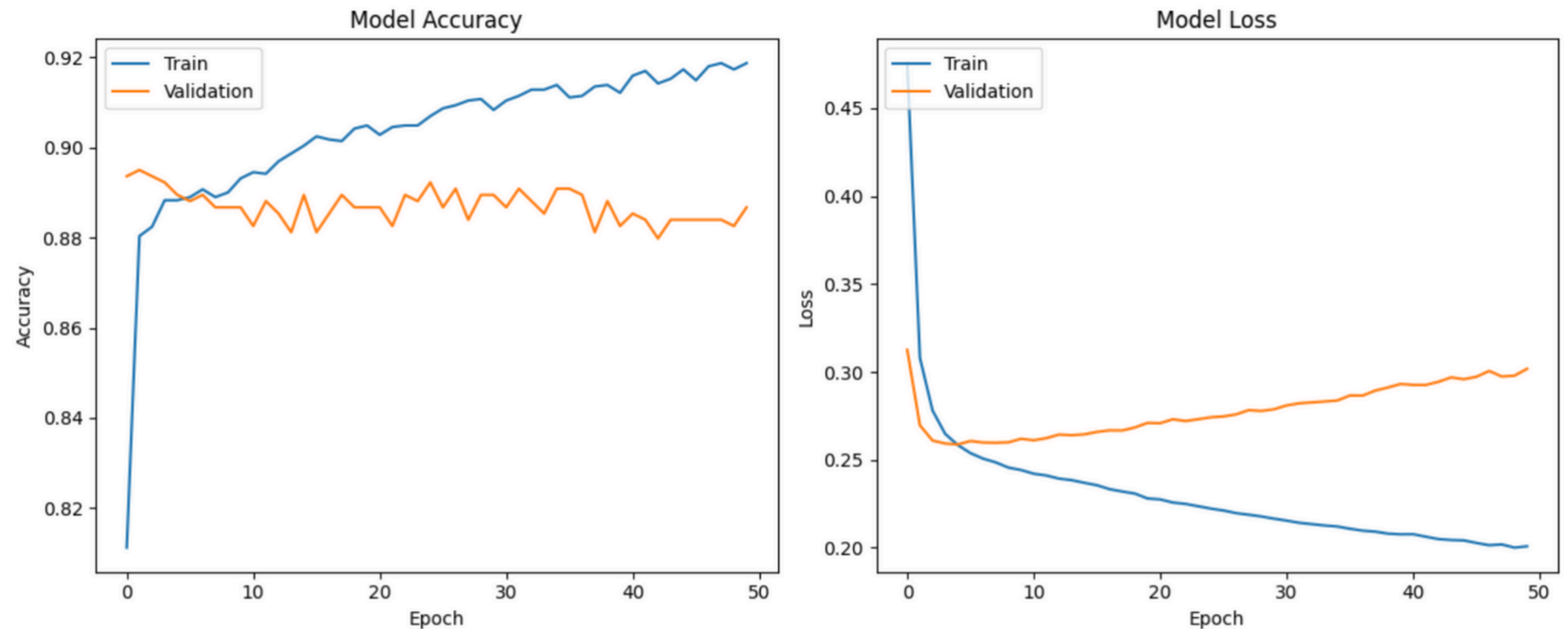
- Key Results: Model achieved 88% accuracy



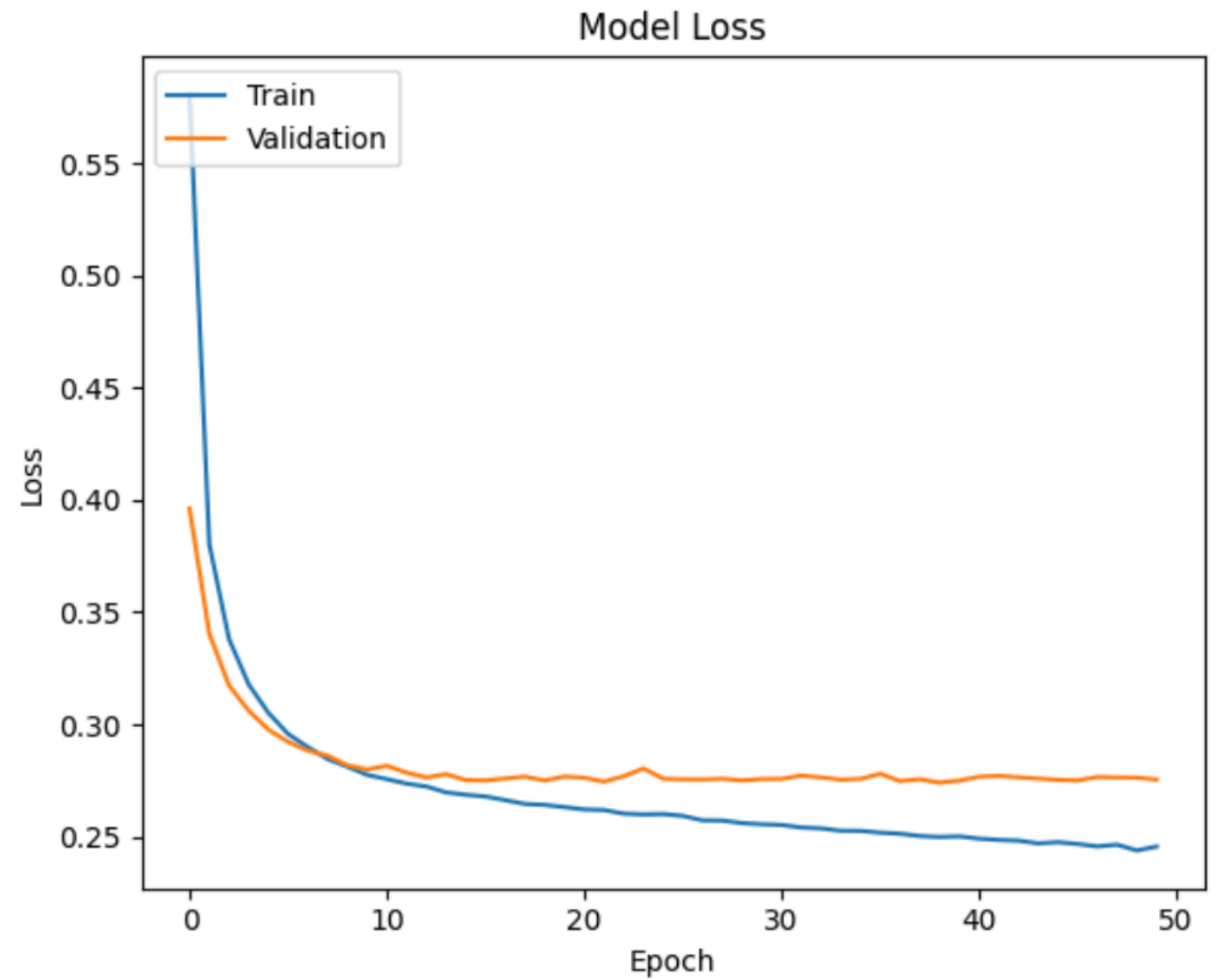
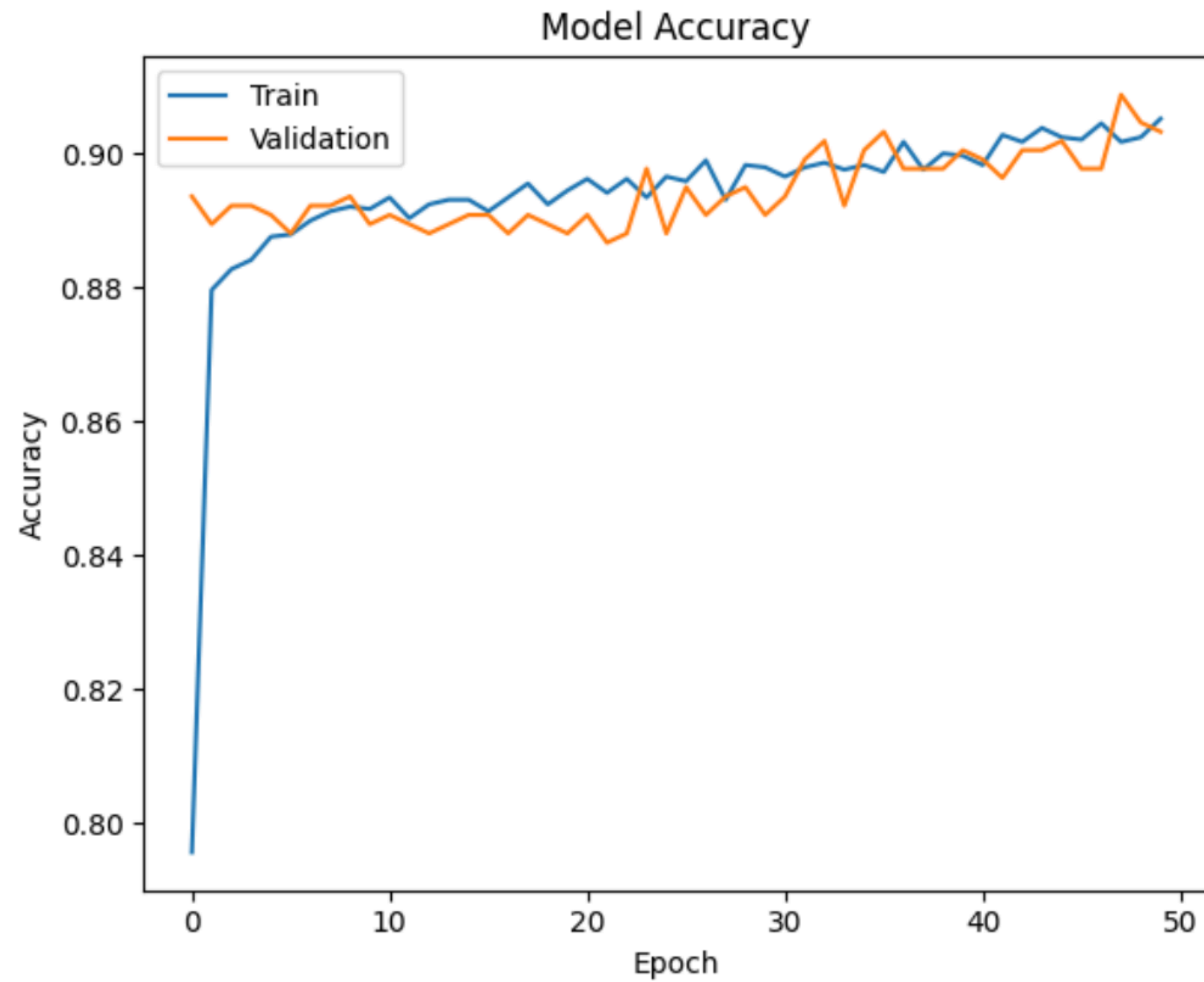
ADJUSTMENT

We can clearly see, looking at the accuracy and loss graph from the Neural Network that the model is overfitting

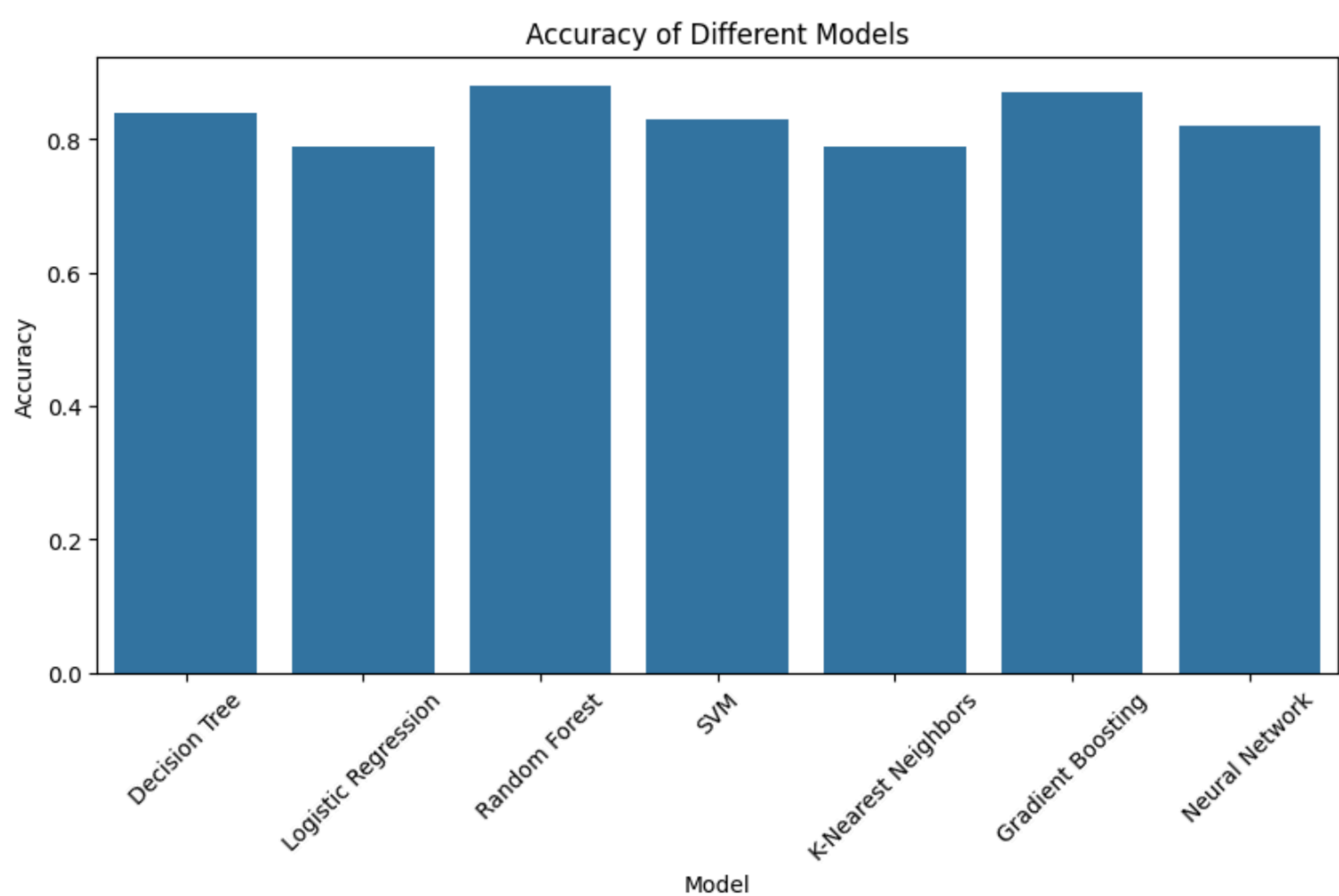
Key Results:
Model achieved
80% accuracy



We can add regularization or dropout layers for possible improvement.

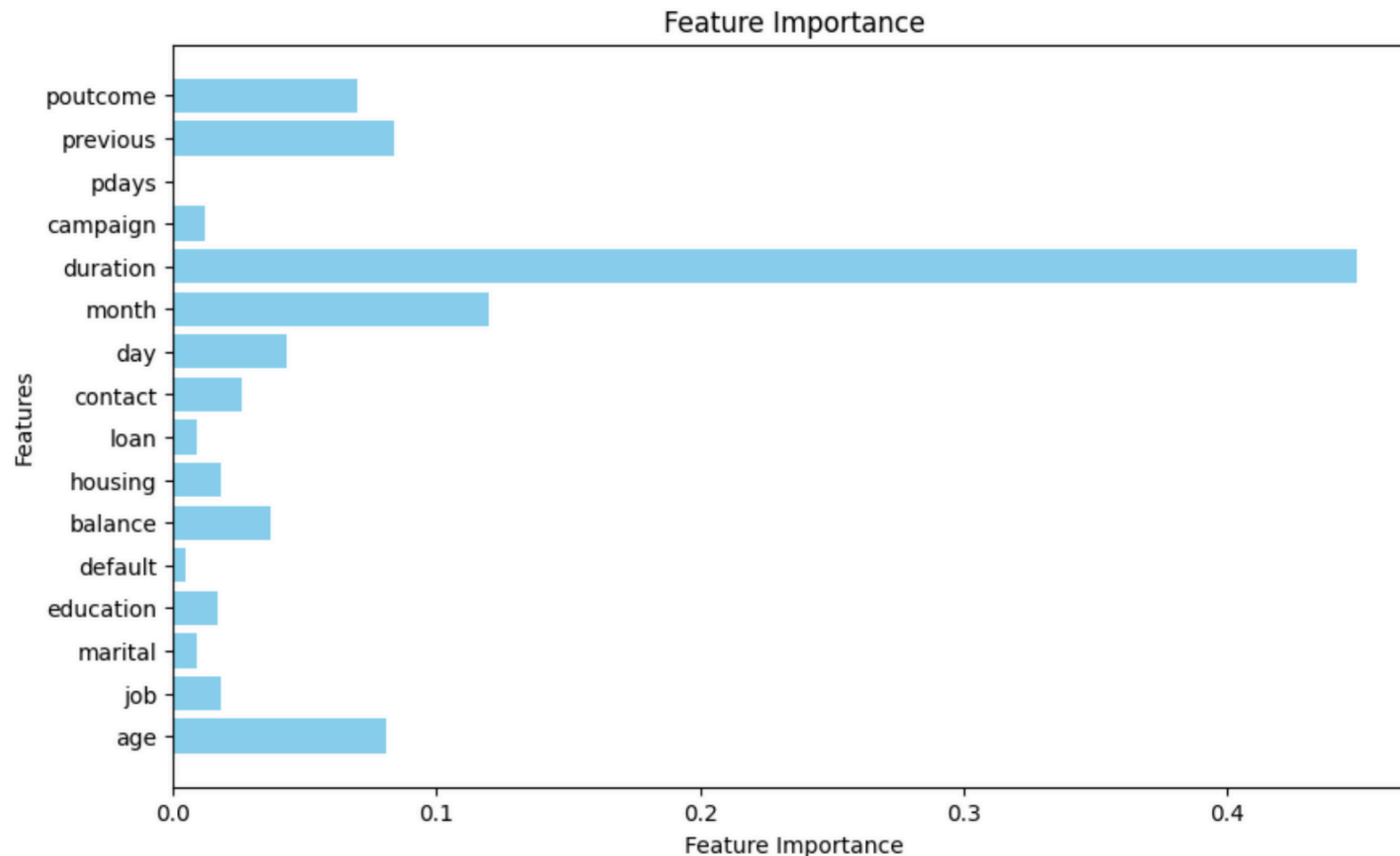


After regularization the model improved slightly with an accury 82%



INSIGHTS

Insights from the Data: Features like 'duration' (the length of contact) seem highly influential in determining customer response, suggesting that longer contact times correlate with a higher likelihood of success. It means that the customer is showing interest.



CONCLUSION

Processing the dataset and using several models, I obtained promising result; but have room for improvement. The bank should focus its marketing efforts on customers most likely to subscribe, reducing costs and improving the efficiency of campaigns.

(Customer showing interest)

**THANK
YOU**