



Shark Tank

Machine Learning Model



Andrea Ceriati, Kathryn Lesh, Emily Sims, and Scott Zuckerbrow

Project Overview:

Purpose:

Analyze Shark Tank data to see what makes a pitch successful on the show.

Goals:

To train a Machine Learning Model to accurately predict if a business will get a deal on Shark Tank.

About the Data:

Seasons: 1 - 15

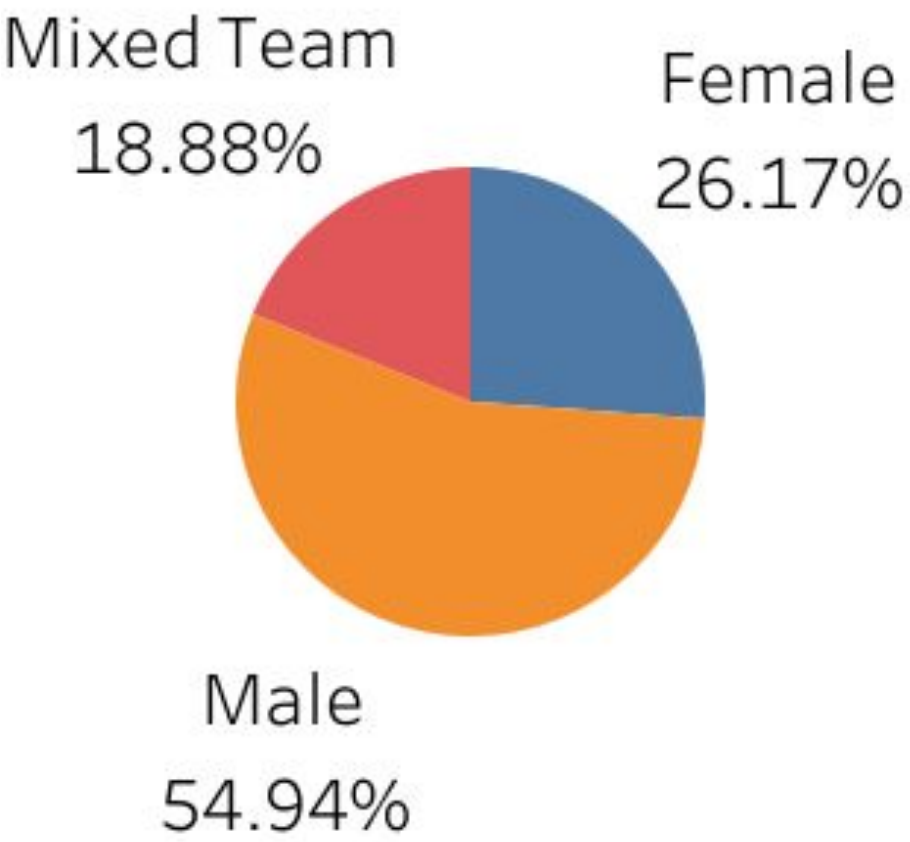
1345 data entries

51 columns

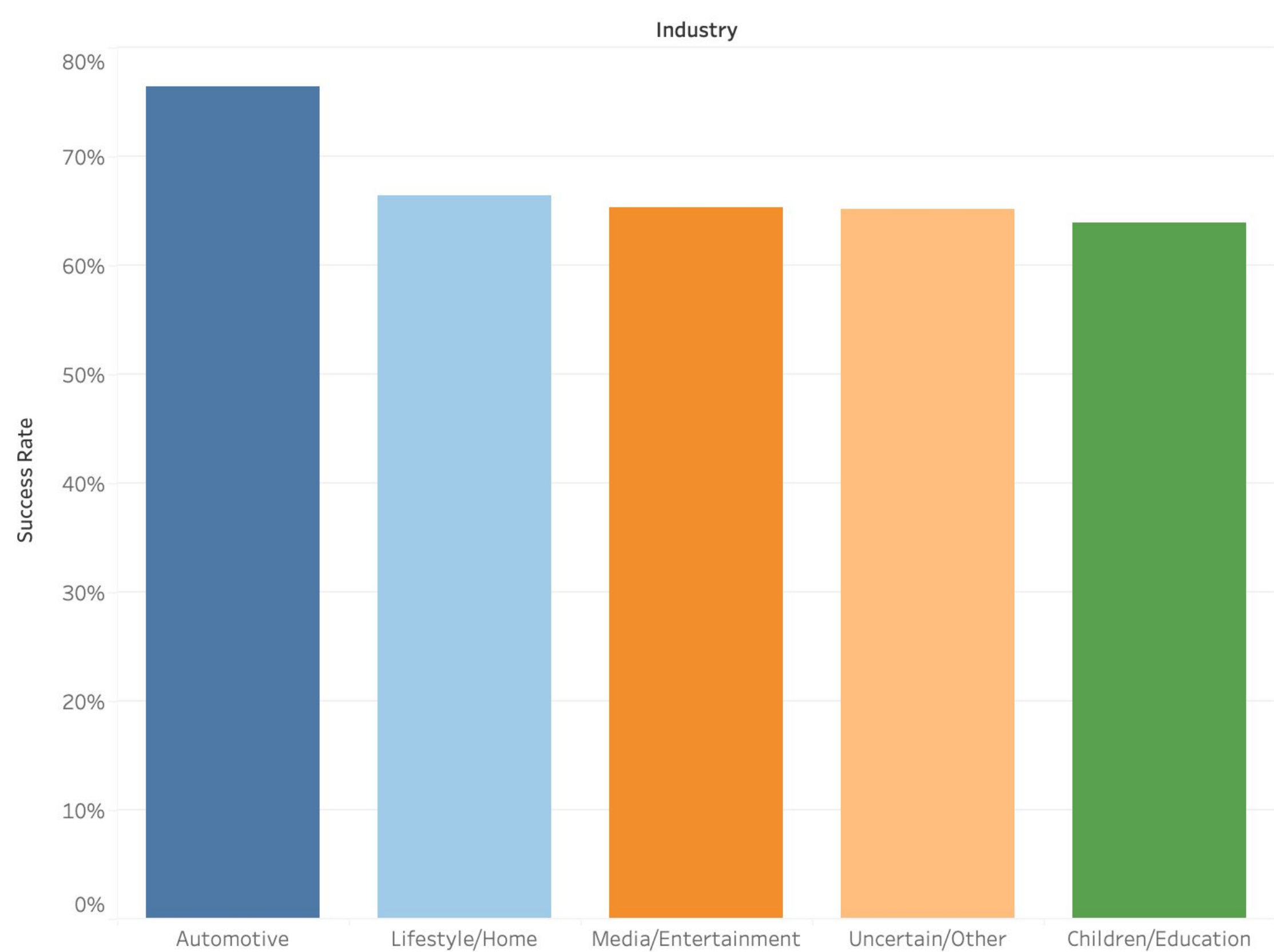
Startup Name	Episode Number	Pitch Number	Season Start	Season End	Original Air Date	Industry	Business Description	Company Website	Pitchers Gender	View
AvaTheElephant	1	1	2009-08-09	2010-02-05	2009-08-09	Health/Wellness	Ava The Elephant - Baby and Child Care	1	Female	
Mr.Tod'sPieFactory	1	2	2009-08-09	2010-02-05	2009-08-09	Food and Beverage	Mr. Tod's Pie Factory - Specialty Food	1	Male	
Wispots	1	3	2009-08-09	2010-02-05	2009-08-09	Business Services	Wispots - Consumer Services	1	Male	
legeFoxesPackingBoxes	1	4	2009-08-09	2010-02-05	2009-08-09	Lifestyle/Home	College Foxes Packing Boxes - Consumer Services	1	Male	

Cleaned dataframe sample:

About the Data:



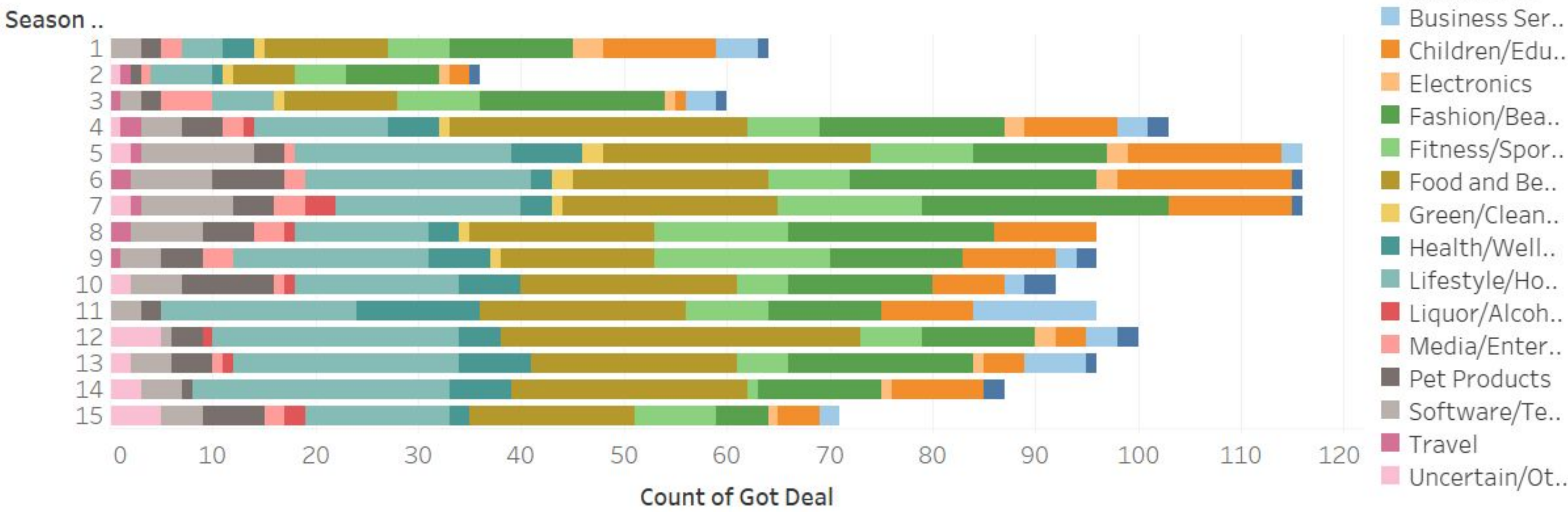
Gender of all pitchers



Top 5 industries with the highest success rate

About the Data:

Industry Success by Season



Initial Feature Analysis

- **Objective:** Understand the relevance of feature for prediction
- **Exploratory data analysis:**
 - Visualized the data
 - Chi-Square Testing
 - Model Testing

- **Selected Features:**

○ Original Ask Amount	○ Pitchers Gender_Female
○ Original Offered Equity	○ Pitchers Gender_Male
○ Has_Website	○ Pitchers Gender_Mixed Team

Model Development

- **Logistic Regression:**
 - Simple and interpretable
 - Suitable for binary classification tasks
- **K-Nearest Neighbors (KNN):**
 - Non-parametric nature
 - Effective in complex decision boundaries
- **Deep Neural Network (DNN):**
 - Capable of identifying complex patterns
 - Captures hierarchical features

Model Optimization:

- **Sequential Feature Selection:**

- **Approach:** Iteratively incorporate or omit features based on their impact.
- **Scoring Metric:** F1 score.
- **Benefits:** Optimize model performance while mitigating overfitting risks.

- **Hyperparameter Tuning:**

- **Approach:** Leveraging Keras Tuner for systematic tuning.
- **Key Hyperparameters:** Activation functions, layer count, neuron count per layer.
- **Benefits:** Enhance model optimization and generalization to new instances.

Logistic Regression Findings:

Used sequential
feature selector to
evaluate the most
relevant features
first

```
Logistic Regression:
Selected Features: ['Original Ask Amount', 'Original Offered Equity', '
Pitchers Gender_Female', 'Pitchers Gender_Male', 'Pitchers Gender_Mixed
Team']

Number of Features: 5

              precision    recall  f1-score   support

Did Not Get Deal      0.59      0.08      0.13        133
      Got Deal      0.62      0.97      0.75        204

      accuracy                   0.61        337
      macro avg      0.60      0.52      0.44        337
      weighted avg      0.60      0.61      0.51        337

Confusion Matrix:
[[ 10 123]
 [  7 197]]
=====
```

KNN Findings:

Accuracy: 62%

```
=====
K-Nearest Neighbors:
Selected Features: ['Original Ask Amount', 'Original Offered Equity', '
Pitchers Gender_Female', 'Pitchers Gender_Male', 'Pitchers Gender_Mixed
Team']

Number of Features: 5

              precision    recall  f1-score   support

Did Not Get Deal      0.50      0.23      0.32         133
   Got Deal           0.63      0.85      0.72         204

    accuracy                   0.61         337
   macro avg              0.56      0.54      0.52         337
  weighted avg              0.58      0.61      0.56         337

Confusion Matrix:
[[ 31 102]
 [ 31 173]]
=====
```

Neural Network Findings:

Model 1:

Accuracy: 63.8% Loss: 65.7%

Confusion Matrix:

```
[[ 27 106]
 [ 16 188]]
```

Model 2: Hyperparameter Tuning

Accuracy: 62.6% Loss: 65.5%

Confusion Matrix:

```
[[ 44  89]
 [ 37 167]]
```



Uses & Future Applications

Design a model to predict which shark will make a deal

Expand on this model to predict which business ideas a particular shark will invest in.

Could be expanded to include data from other investment or crowdfunding platforms

Kickstarter, GoFundMe, or Indiegogo to evaluate the most successful proposals and predict success.

Conclusion

Best Model: Neural Network

Accuracy: 62% Loss: 66.1%

Other Factors:

- Traits such as charisma, poise and confidence
- Professional appearance
- Previous sales
- Other investors
- The quality of the product/service





**Any questions
sharks?**

*Thank you for
listening!*