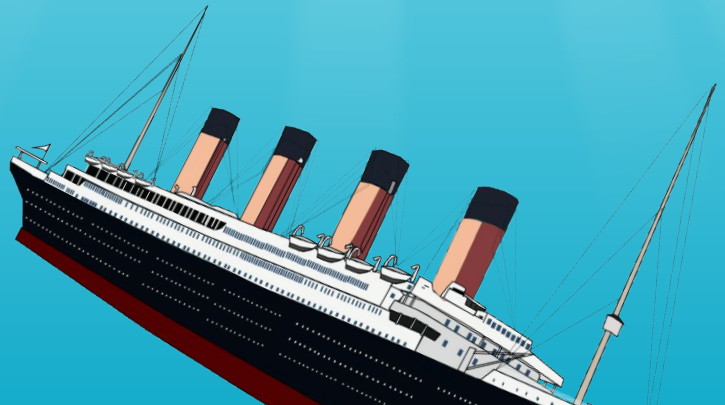
Sink or Swim:
Using Machine Learning to Predict Titanic Deaths

Layla Burgan and Clare Sakauye



Overview

- Looked at over 1000 passenger records from the RMS Titanic using a Kaggle dataset
- Tried multiple machine learning models and optimization methods to achieve desired accuracy
- Used our best model to predict the outcome of specific data points (Rose and Jack from the 1997 film Titanic)

The Data

Passenger Class (1 = 1st; 2 = 2nd; 3 = 3rd) Survival (0 = No; 1 = Yes) n

~~Name~~

Sex

Age

Number of Siblings/Spouses Aboard

Number of Parents/Children Aboard

~~Ticket Number~~

Passenger Fare (British pound)

Cabin embarked Port of Embarkation (C = Cherbourg; Q = Queenstown; S = Southampton)

~~Lifeboat body~~

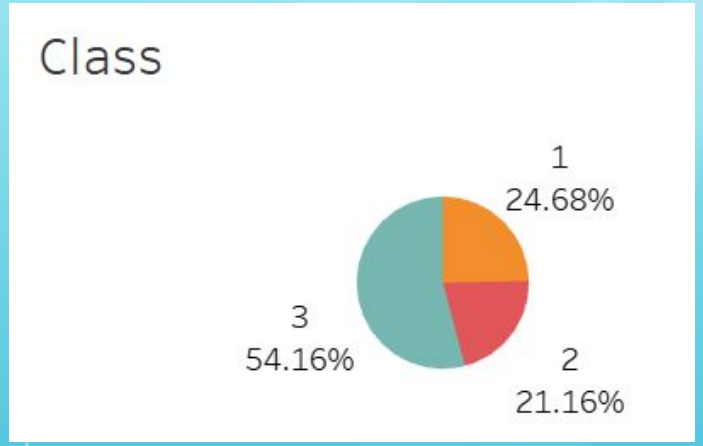
~~Body Identification Number~~

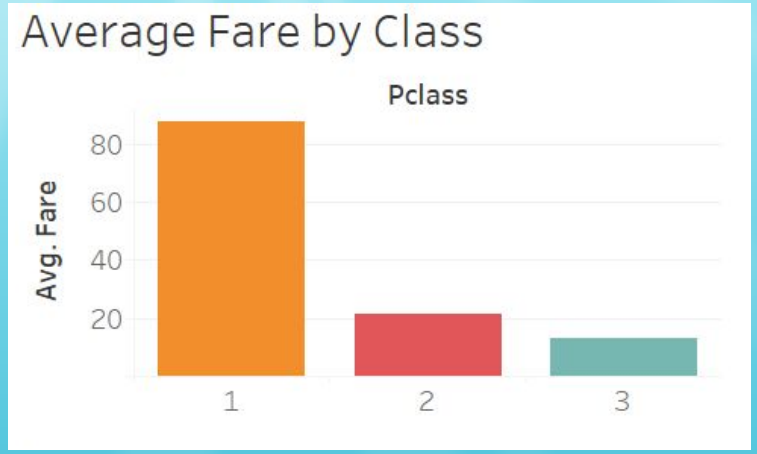
~~Home/Destination~~

Several irrelevant columns were dropped

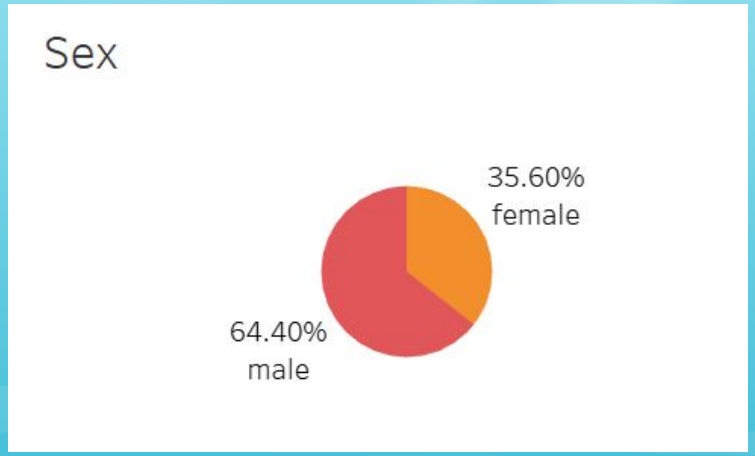
Titanic Demographics

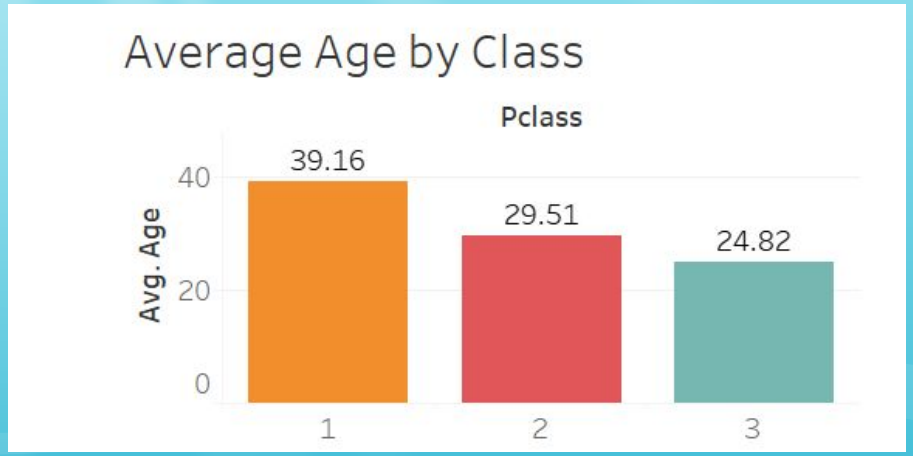
Who was on the Titanic?





Who was on the Titanic?





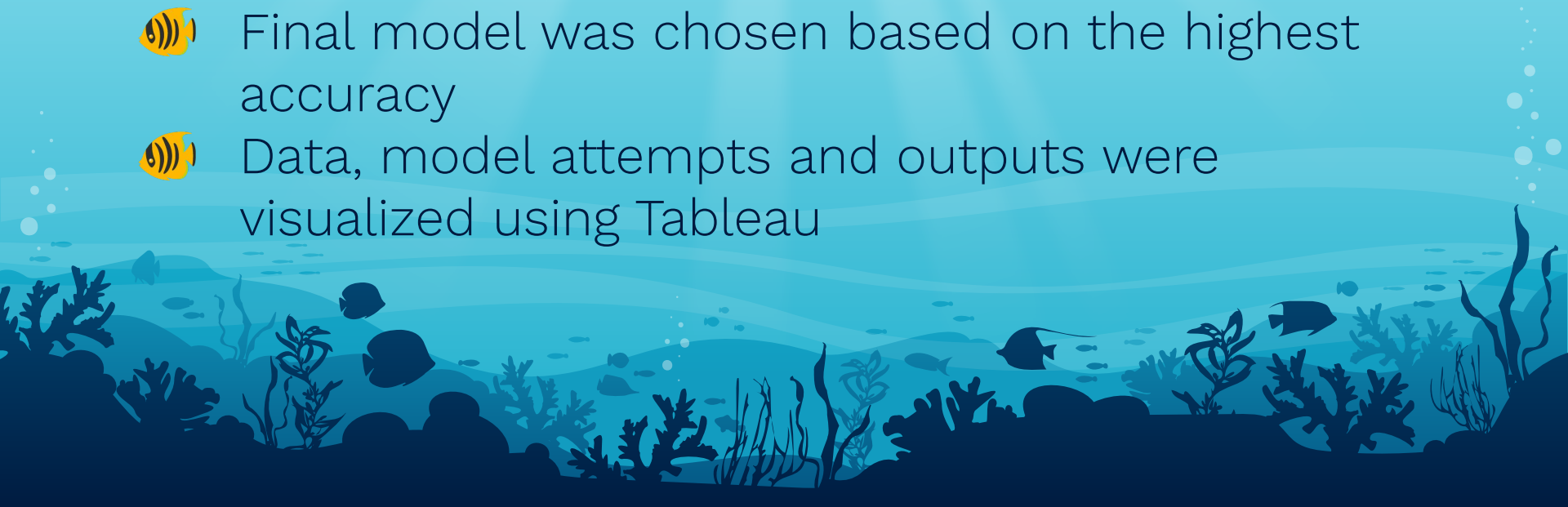
How much did a ticket cost?

Class	Price in 1912 ($)	Price in 2022 ($)
First Class	$4,350	$133,132
First Class	$150	$4,591
Second Class	$60	$1,834
Third Class	$35	$1,071

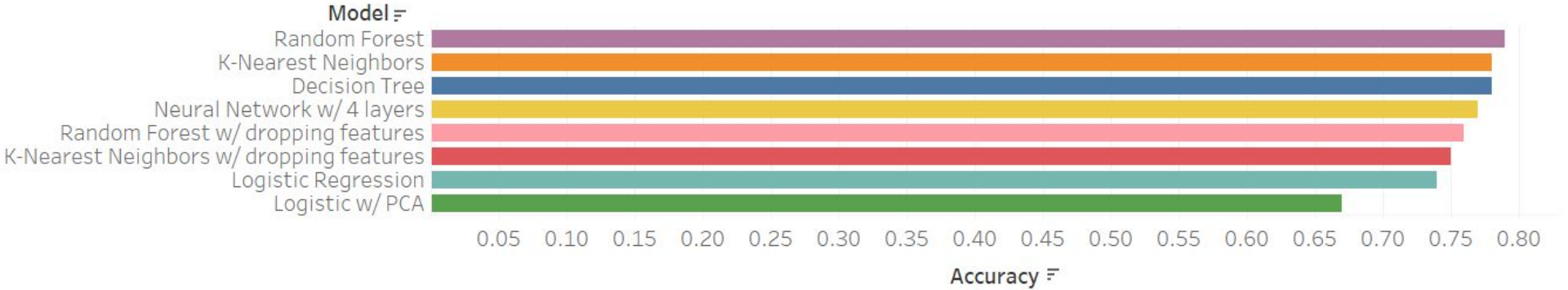
Picking a Model

Methods Used

- Data was cleaned and explored using Pandas
- Scikit Learn and Tensorflow were used for our various model attempts and optimizations
- Final model was chosen based on the highest accuracy
- Data, model attempts and outputs were visualized using Tableau



Model Attempts and Accuracy

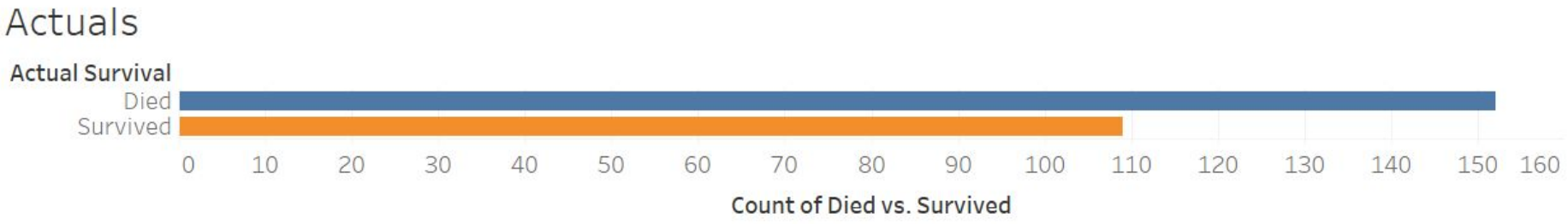


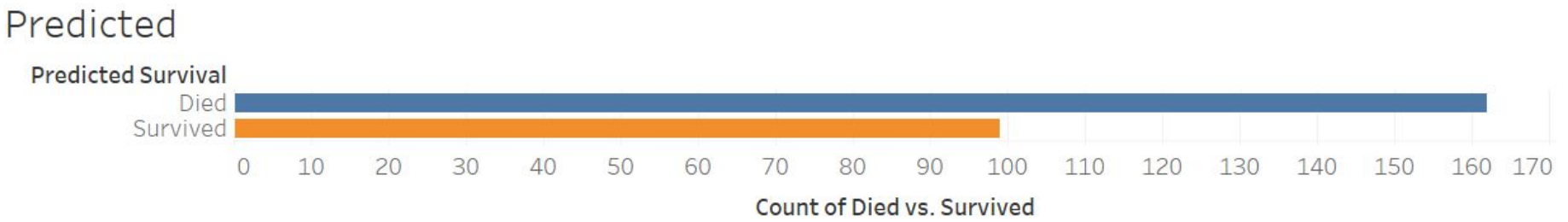
Random Forest

- The random forest model was our best model overall, with 79% accuracy
- We attempted model optimization by dropping features but there was no positive effect on accuracy
- Our random forest used 1000 trees (estimators)

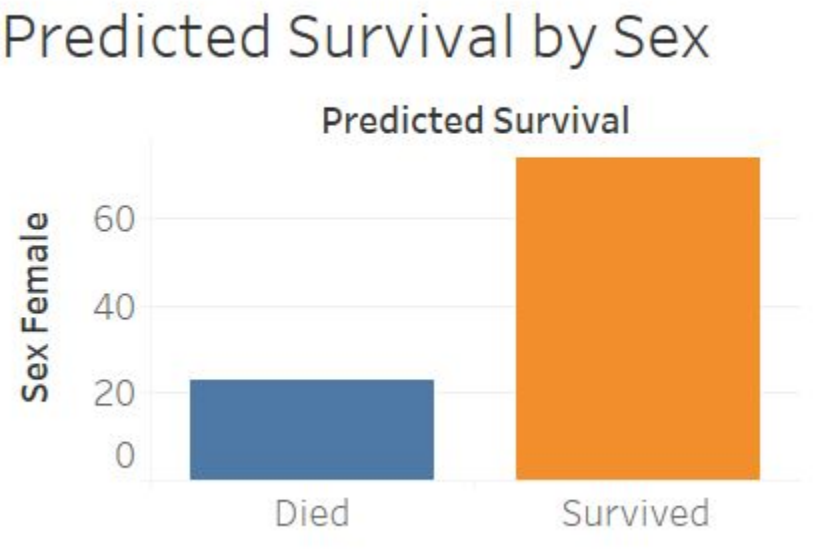
Analyzing Model Results

Actual vs. Predicted Survival

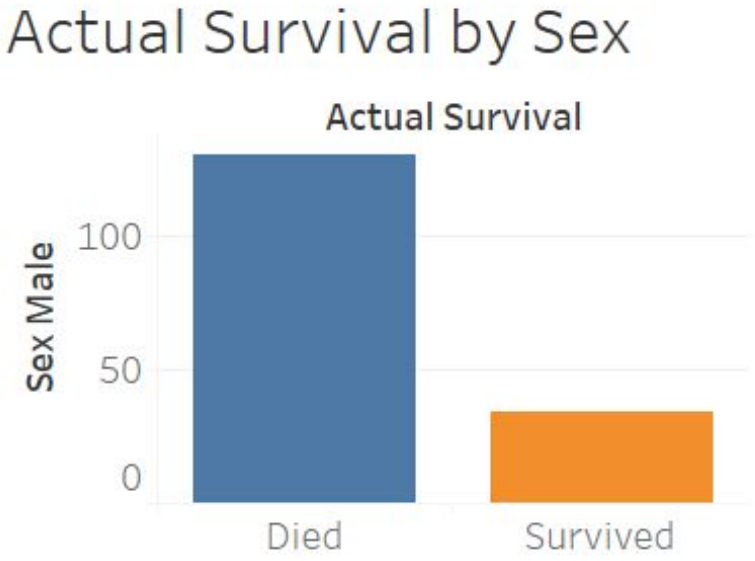


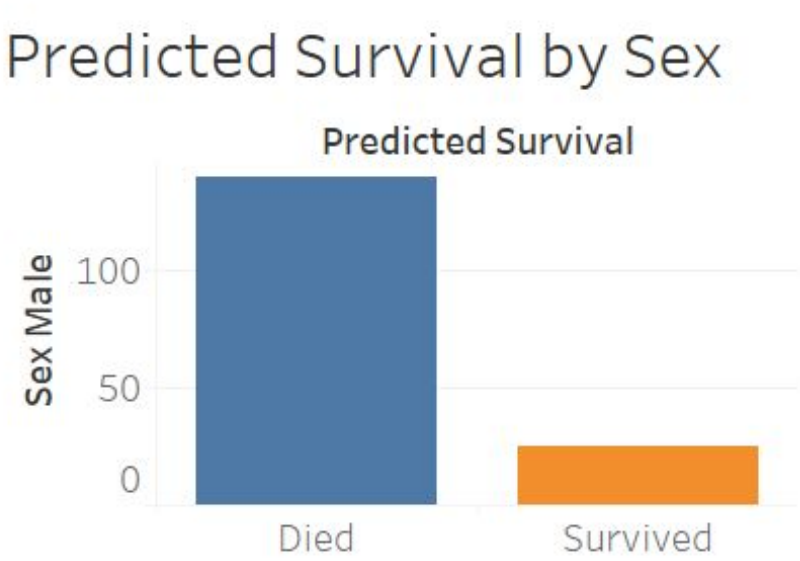


Actual vs. Predicted Survival by Sex (Female)

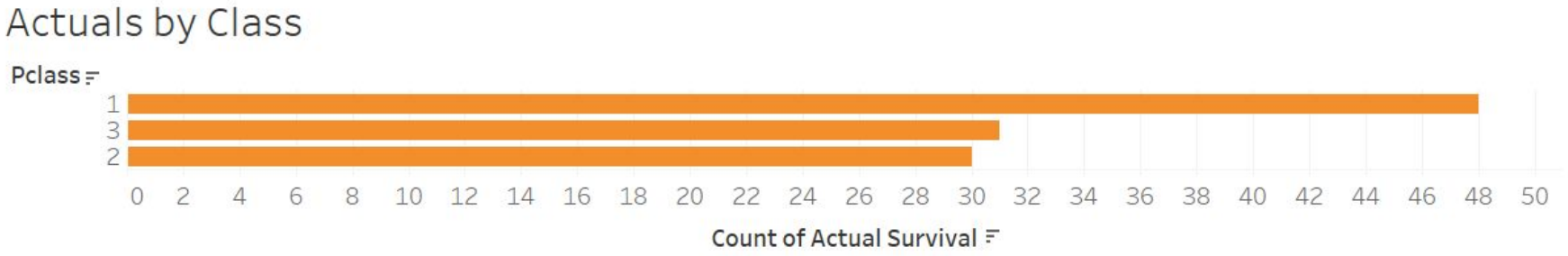


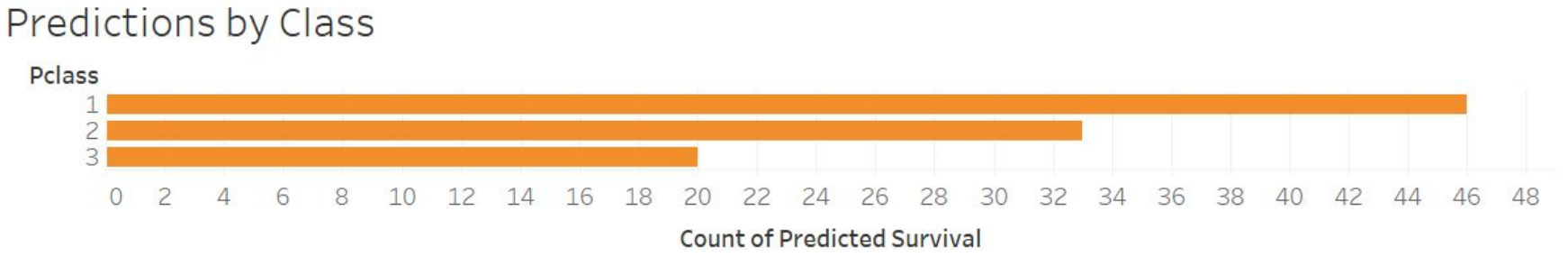
Actual vs. Predicted Survival by Sex (Male)

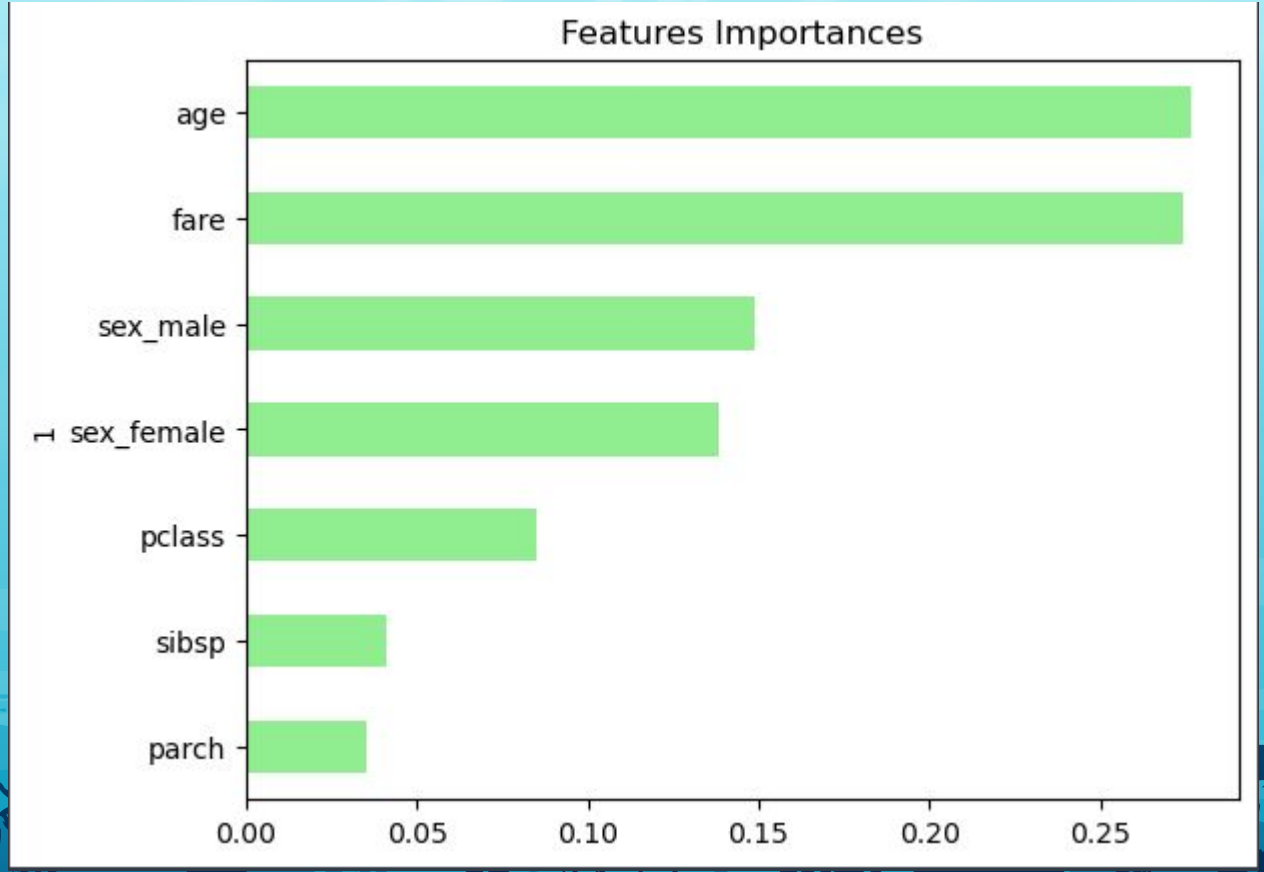




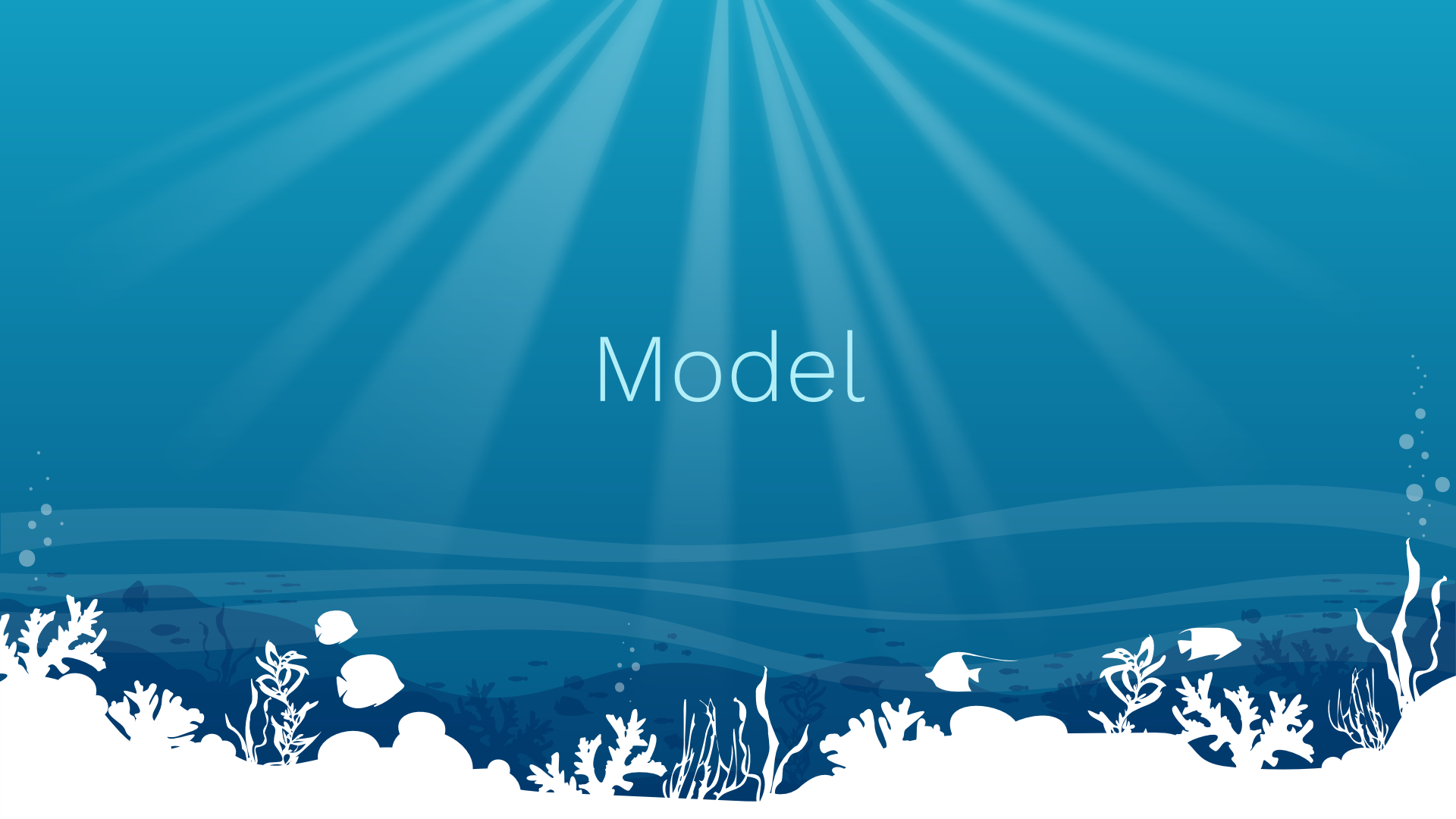
Actual vs. Predicted Survivors by Class






Features Importances
age
fare
sex_male
sex_female
pclass
sibsp
parch
0.00
0.05
0.10
0.15
0.20
0.25

Model



Any volunteers?

Info needed:

- Class (1,2,3)
- Age
- Siblings/Spouses
- Parents/Children
- Fare
- Sex (M/F)
- (Queenstown,Cherbourg, or Southampton)

Typical Fares in 1912 GBP:

- First class: 80-500 GBP
- Second Class 20- 70 GBP
- Third Class: 4-15 GPB

Findings

- Overall, our model accurately predicted that young 1st class, women were most likely to survive
- Using feature importance, age, fare and sex contributed most to passenger survival
- Random forest was the most effective in predicting survivors at a 79% accuracy
- Our model successfully predicted Jack and Rose's fate