

Would You Have Survived the Titanic?



Cynthia Wang



Objective
Data
Models
Web App Demo

Objective

- ▼ Build a binary classifier that predicts the survival of passengers on the Titanic.
- ▼ Explore the factors that influenced survival of passengers in catastrophic events.

Data

Name
Age
Gender
Fare
Passenger Class

Title
Age Group
Group Size
Individual Fare
Family Size

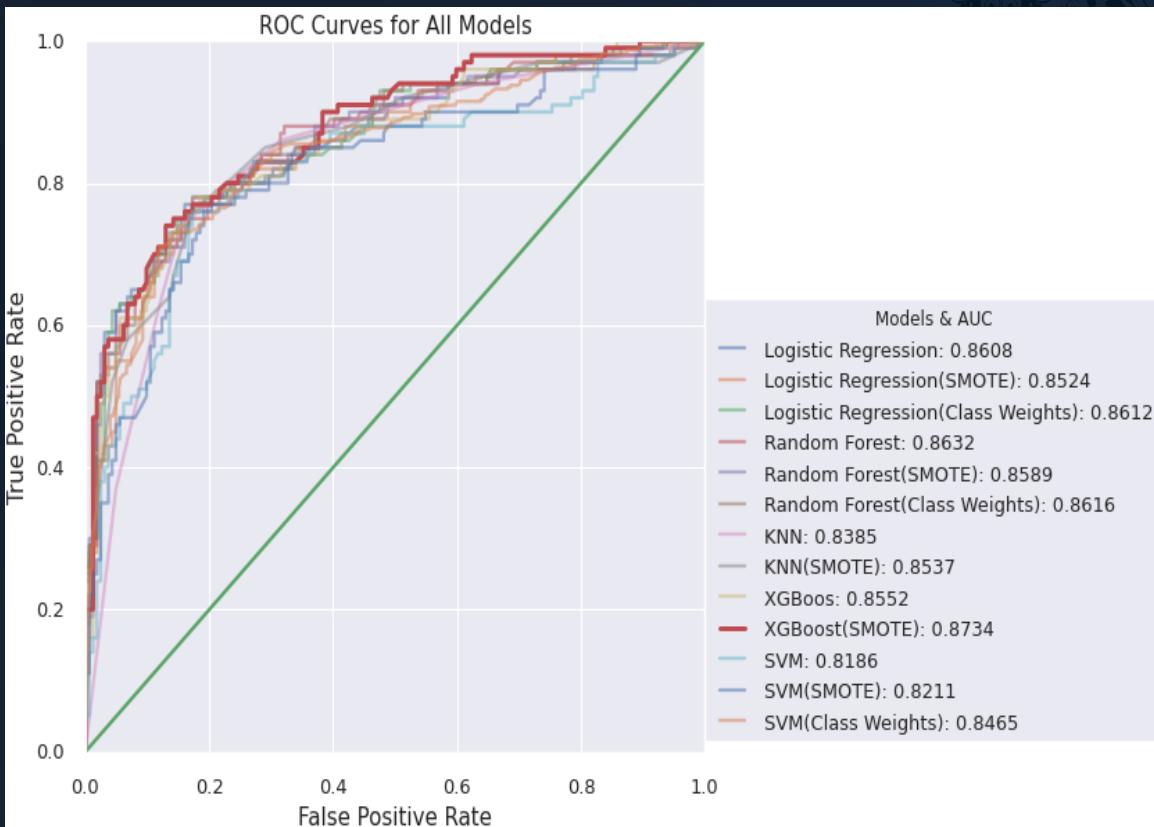


Survival Status

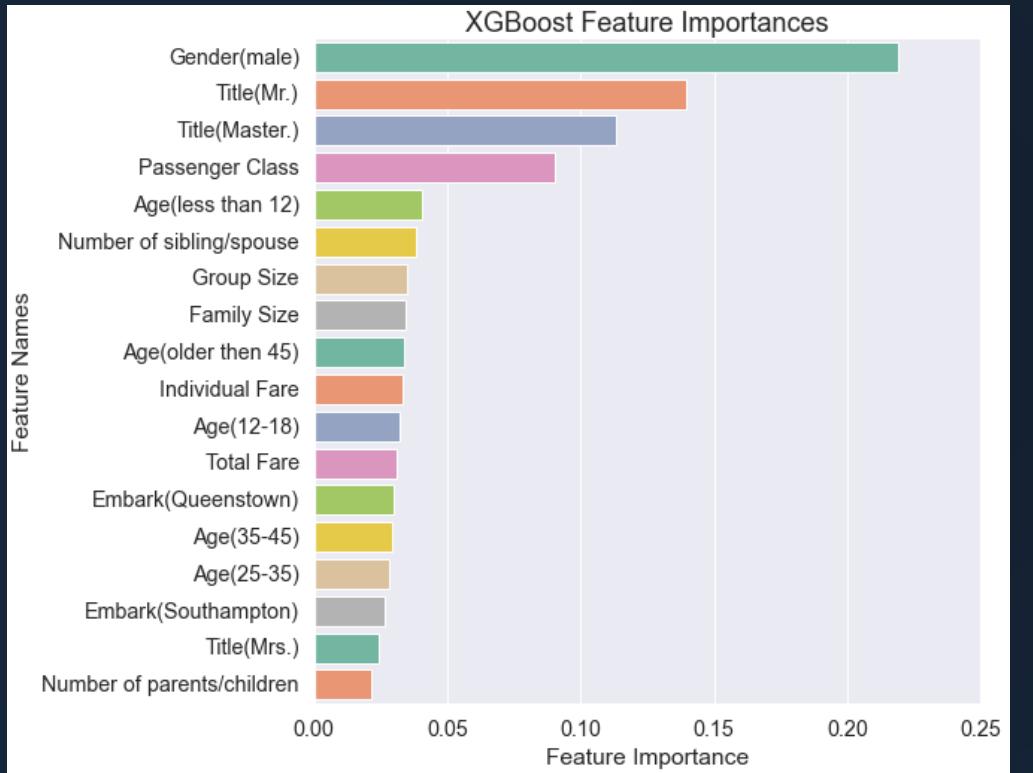
Siblings/Spouses/Parents/Children

Models

	CV F1	Test F1
KNN	73.36%	74.37%
Logistic Regression	74.51%	75.36%
SVM	74.72%	74.51%
Random Forest	77.08%	74.51%
XGBoost	77.10%	74.51%



Models



What features are important?

- Gender

- Title

- Passenger Class

- Age

Findings & Future Direction

What increased the chances of survival on Titanic?

- ▼ Female
 - Married female passengers had higher chances of survival
- ▼ Children
 - ▼ Especially for those under 12
- ▼ First Class

Look at other datasets about similar catastrophic accidents to understand what factors affect people's survival under extreme circumstances

The Titanic Survival Predictor

<https://titanic-survival-predictor1128.herokuapp.com/>

Would You Have Survived the Titanic?

What's your age?

Please select

What's your gender?

Please select

What's your title?

Please select

Where are you embarking?

Please select

What passenger class are you in?

Please select

How much is your ticket price (\$)? *

Are you traveling with your siblings/spouse? If yes, how many?

Are you traveling with your parents/children? If yes, how many?

How many people are you traveling with in total (excluding yourself)?

* The median ticket prices for First Class, Second Class, and Third Class passengers are \$600, \$250, and \$150 respectively (in current US dollars). Even though individual ticket prices are different, feel free to use this information to estimate your survival likelihood!

Your probability of survival:

0%



Thanks!

Appendix

	Model	Test AUC	Test F1:	Threshold(train,test)	Train AUC	Train F1	TrainCV F1
0	Logistic Regression	0.860772	0.751220	0.5	0.857401	0.748718	0.739240
1	Logistic Regression (SMOTE)	0.852438	0.753623	0.5	0.854241	0.750630	0.745139
2	Logistic Regression (ClassWeights)	0.861204	0.728111	0.5	0.856946	0.750605	0.737767
3	Logistic Regression (ThresholdAdjustment)	0.860772	0.758621	[0.3858858858858859, 0.5043043043043044]	0.857401	0.754217	0.000000
4	KNN	0.838488	0.743719	0.5	0.917734	0.789272	0.733623
5	KNN (SMOTE)	0.853735	0.745098	0.5	0.895790	0.764780	0.732693
6	KNN (ThresholdAdjustment)	0.838488	0.743719	[0.4009009009009009, 0.4009009009009009]	0.917734	0.789272	0.000000
7	Random Forest	0.863241	0.737374	0.5	0.965446	0.874346	0.756532
8	Random Forest (SMOTE)	0.858920	0.745098	0.5	0.964426	0.875481	0.770870
9	Random Forest (ClassWeights)	0.861574	0.738916	0.5	0.966810	0.879795	0.763750
10	Random Forest (ThresholdAdjustment)	0.863241	0.743719	[0.4790790790790791, 0.4923923923923924]	0.965446	0.880519	0.000000
11	XGBoost	0.855185	0.743719	0.5	0.952210	0.847854	0.763522
12	XGBoost (SMOTE)	0.873364	0.745098	0.5	0.942065	0.825921	0.771019
13	XGBoost (ThresholdAdjustment)	0.855185	0.752475	[0.35015015015015016, 0.4538538538538539]	0.952210	0.855377	0.000000
14	SVM	0.818611	0.745098	0.5	0.823422	0.746770	0.747226
15	SVM (SMOTE)	0.821142	0.734300	0.5	0.844214	0.743003	0.743146
16	SVM (ClassWeights)	0.825772	0.720379	0.5	0.846520	0.733417	0.736662
17	SVM (ThresholdAdjustment)	0.818611	0.745098	[0.5603603603603604, 0.43003003003003004]	0.823422	0.751643	0.000000

