

Analysis of Survival Rates aboard the RMS Titanic

GA Final Presentation July 15th - Aditya Rao



What Happened in 1912?

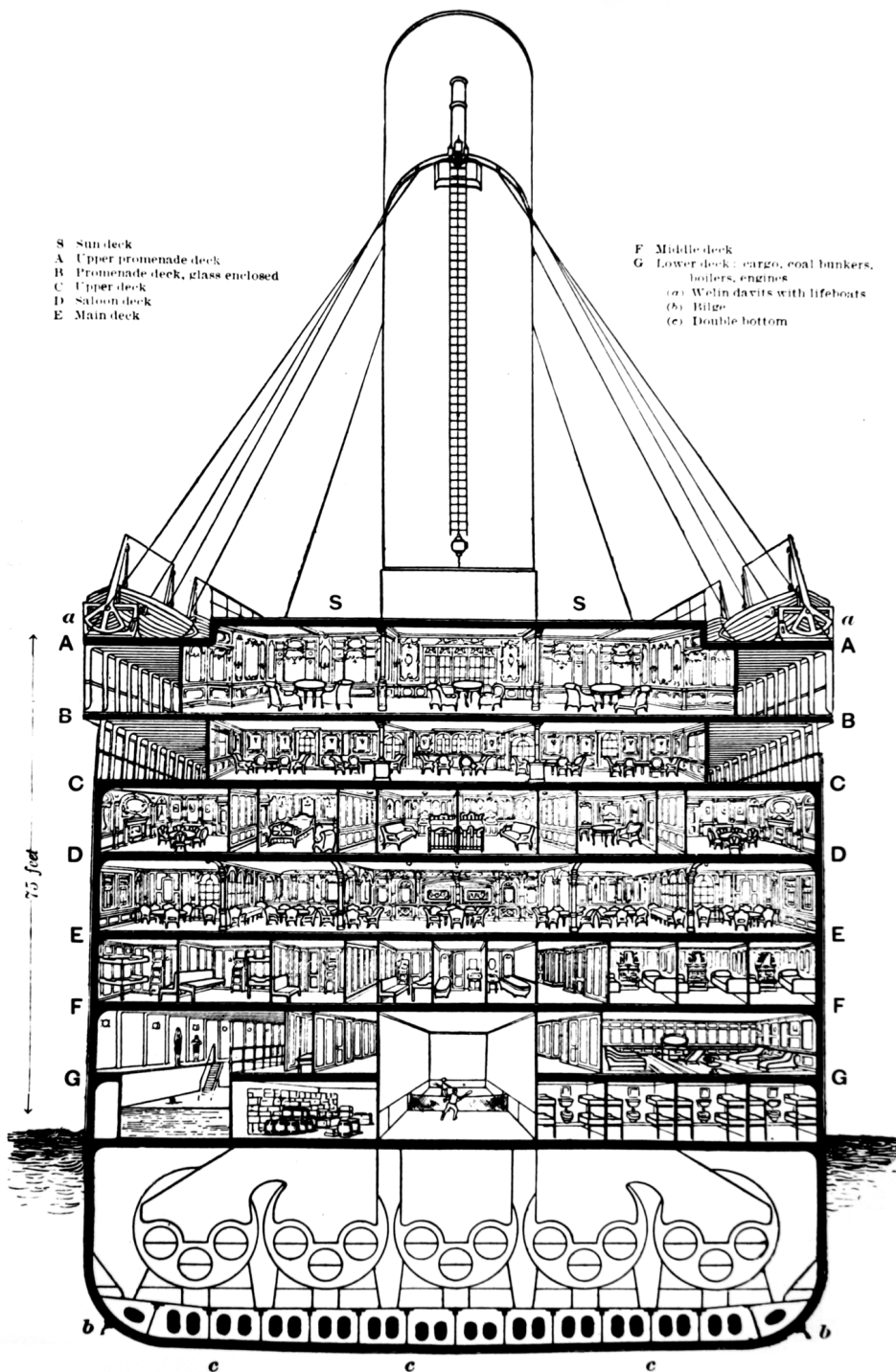


- RMS Titanic set sail on 10th April 1912
- Picked up passengers from 3 ports - Southampton, Cherbourg, Queenstown
- Crashed into an iceberg and sank on 15 April 1912

Cross Section Cutaway of the Titanic

S Sun-deck
A Upper promenade deck
B Promenade deck, glass enclosed
C Upper deck
D Saloon deck
E Main deck

F Middle-deck
G Lower-deck: cargo, coal bunkers, boilers, engines
(a) Winch davits with lifeboats
(b) Bilge
(c) Double bottom



- Decks A - G
- Deck C - 1st Class
- Deck F - 2nd Class
- Deck G - 2nd and 3rd Class

Passenger Cross Section

- 1317 People total
- 324 in First Class
- 284 in Second Class
- 709 in Third Class
- 869 (66%) Male
- 447 (34%) Female
- 107 Children aboard

What are we looking for?

There are 4 questions we want to try and solve:

- Is there a predominant set of criteria that affected the people who survived?
- Does the data set support the hypothesis?
- Is there a machine learning technique that captures this?

What does our data set look like?

- Socio-economic status
- Name
- Sex
- Age
- Price of ticket
- Cabin(s)
- Port of embarkation
- Whether they had Siblings or Parents on board

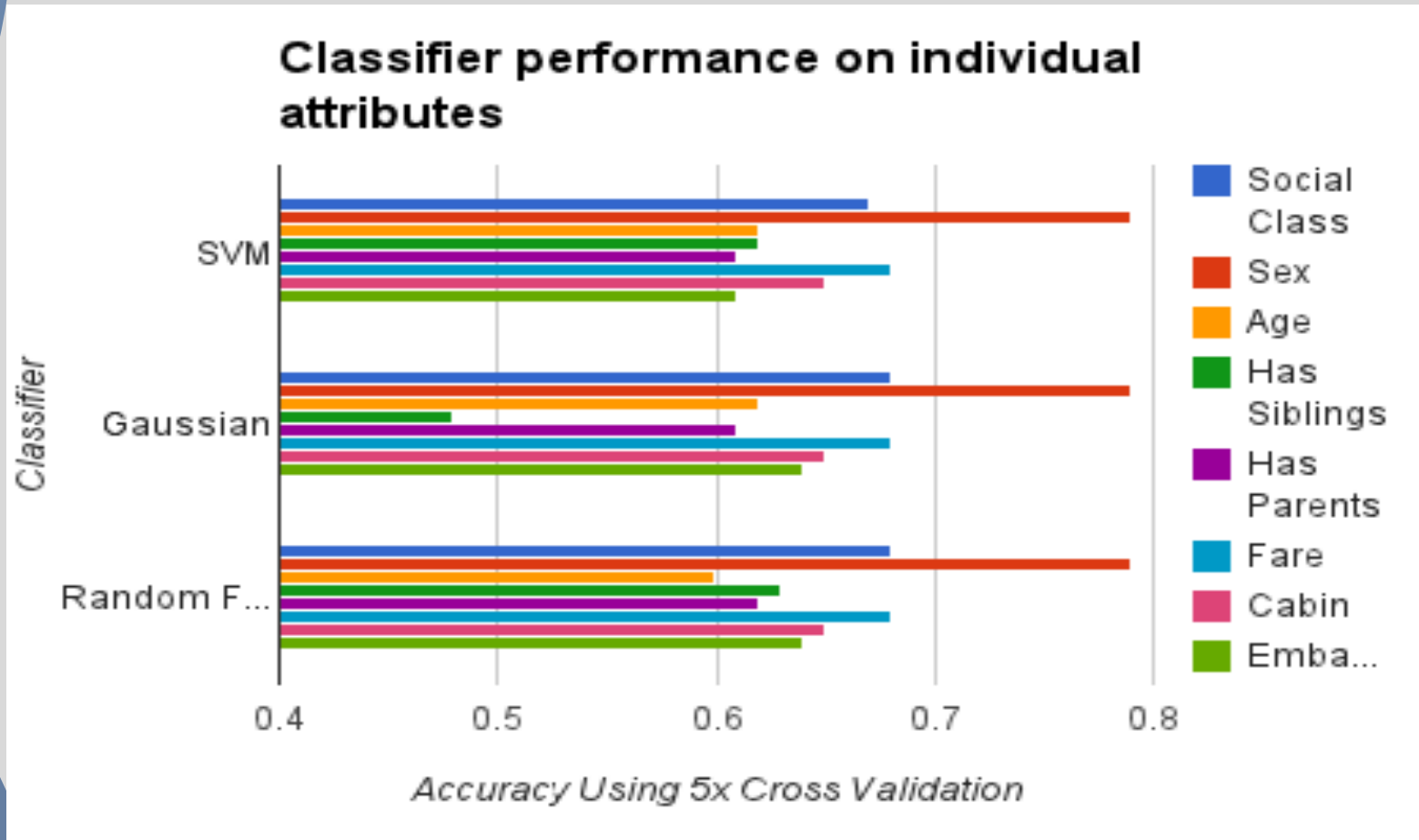
Factors biasing the output variable

- Tradition ("Women and Children First!")?
- Insufficient Number of life-boats on board?
- Most passengers in third class?
- Anybody care about siblings or their parents?

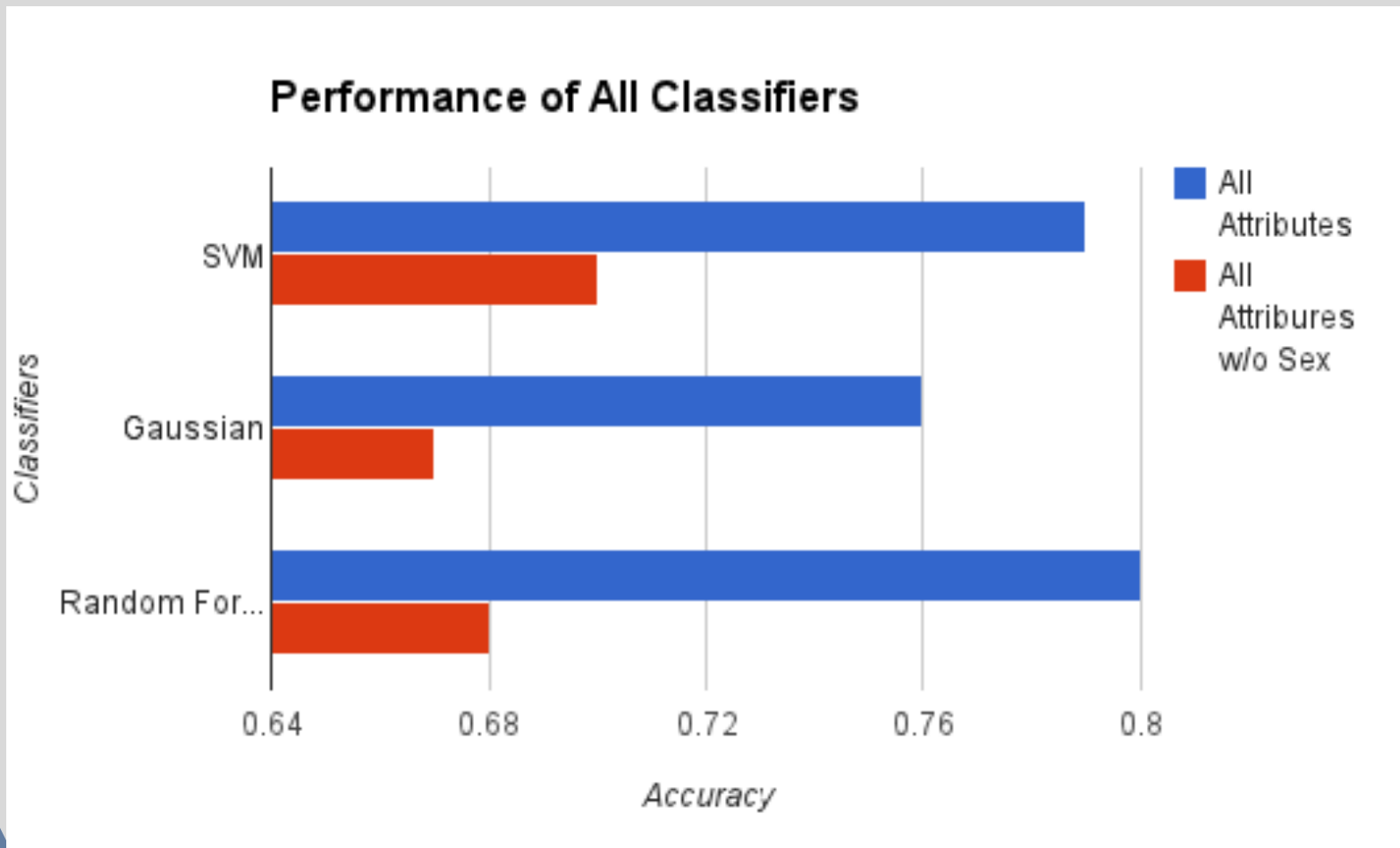
Type of problem that we're trying to solve

- This is a classification problem (who survived, who didn't).
- Analysis was done using 5x cross-fold validation and several classifiers:
- SVMs
- Gaussian Naive Bayes
- Random Forests.

Was there a major determinant of survival?



Which Classifier fared best?



Conclusion

- Sex was the single biggest determining factor
- All models uniformly degraded when sex was removed from the list of variables under consideration.
- There was no single attribute that affected the outcome more than the sex of the passenger.
- Random Forests and SVMs did the best job of correctly predicting outcome.

Some points to ponder from a more recent perspective

- Captain went down with his ship, getting people off till the very last minute
- Insufficient preparation by the crew
- Insufficient number of life boats
- Titanic was not carrying its full complement of passengers.