ELECTRIC DREAM MACHINE

Our adventures in *SSSSSPPPPAAAAAACCCEEE*

OUR COMPETITION

Frame Story:

SPACESHIP TITANIC!

An evil wormhole blackhole asteroid iceberg sunk the space ocean liner cruise ship, and half the passengers teleported!

Question:

WhY DiD SoMe PaSsEngERs TeLePort ANd OthErs Didn'T!?

Problem:

Make a **binary classification** model from passenger information that can predict whether or not a passenger will be teleported

About the Data

Training size: 800kb

Test size: 400kb

Num instances training (m): 8694

Num instances testing (m): 4277

Teleported/Not teleported: Roughly 50/50

Params: n=8?

- ★ Passenger Name (string)
- ★ Home Planet (enum)
- ★ Destination Planet (enum)
- ★ Cryosleep (bool)
- ★ Age (int)
- ★ VIP (bool)
- ★ Cabin (string/int/bool)
- ★ Money Spent (enum/double)

Initial Results from Logistic Regression

78.4% accuracy

Decision Tree

77.4% Accuracy

Random Forest

79.2% accuracy

Naive Bayes

76.0% accuracy (Gaussian)

61.9% accuracy (Multinomial)

Multilayer Perceptron

80.1% accuracy

Top 3 Configurations

Multi-Layer Perceptron
- 80.13% Average CV Accuracy
- 81.14% Final Dataset Accuracy

Random Forest
- 79.17% Average CV Accuracy
- 80.28% Final Dataset Accuracy

Logistic Regression
- 78.43% Average CV Accuracy
- 79.29% Final Dataset Accuracy

Top Performer

Multi-Layer Perceptron
-   80.13% Average CV Accuracy
-   81.14% Final Dataset Accuracy

takes more factors into account than merely presence and absence, like naive bayes.

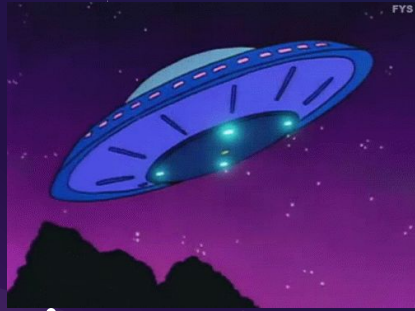Combs back and forth, redistributing weights and improving itself.

However, we had one last trick up our sleeves...

Super mega ultimate ensemble

An ensemble of our top 3 configurations

Takes predictions from all 3, returns the majority vote

How did we do?

★

★ Placement: #335/2572
★ Accuracy: 80.71%

Tips and tricks

Dividing our "Cabin" attribute into 3 separate attributes

- More useable data

PassengerID

- Identifying Group number