PREDICT THE EPIC SCI-FI UNIVERSE
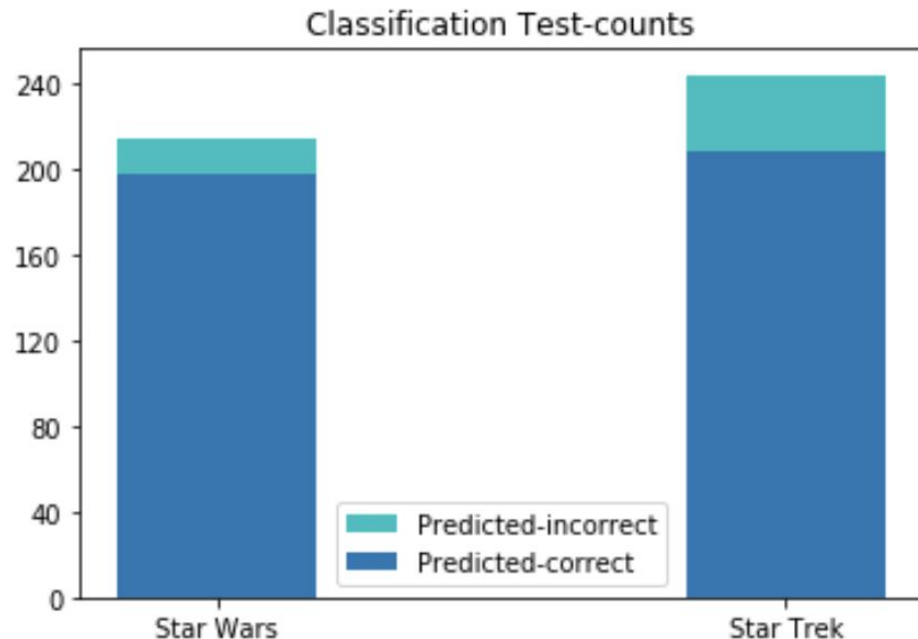
"How to skip SJW drama"

The Problem

Can a Sci-Fi fan support a franchise that has evolved past the tension between story and social justice?

Choice of predictive Models
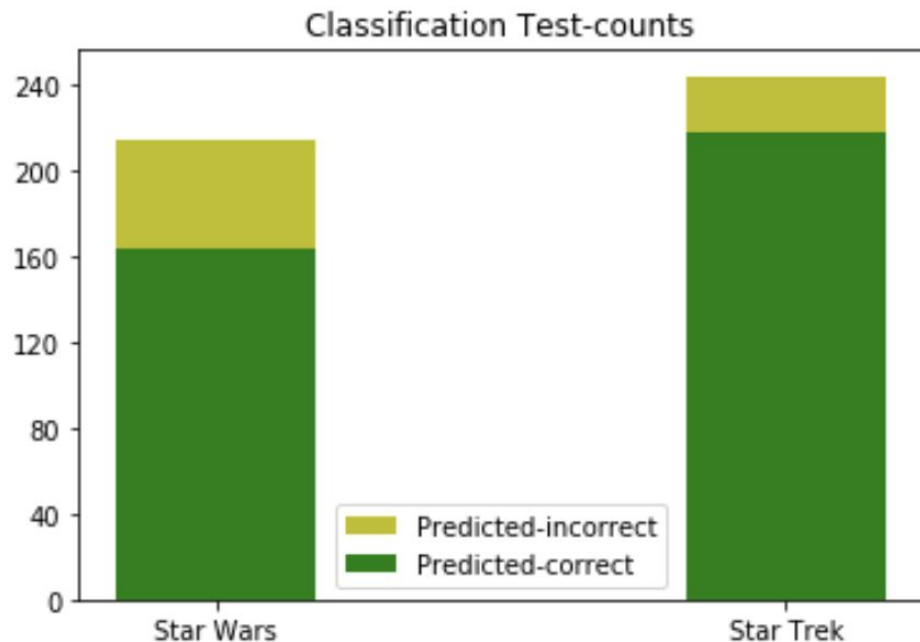
Model	f1-score
Logistic Regression	0.89
SVM	0.87
Naive-Bayes	0.86
Random Forests	0.83
KNN	0.47

Logistic Regression - Part 1

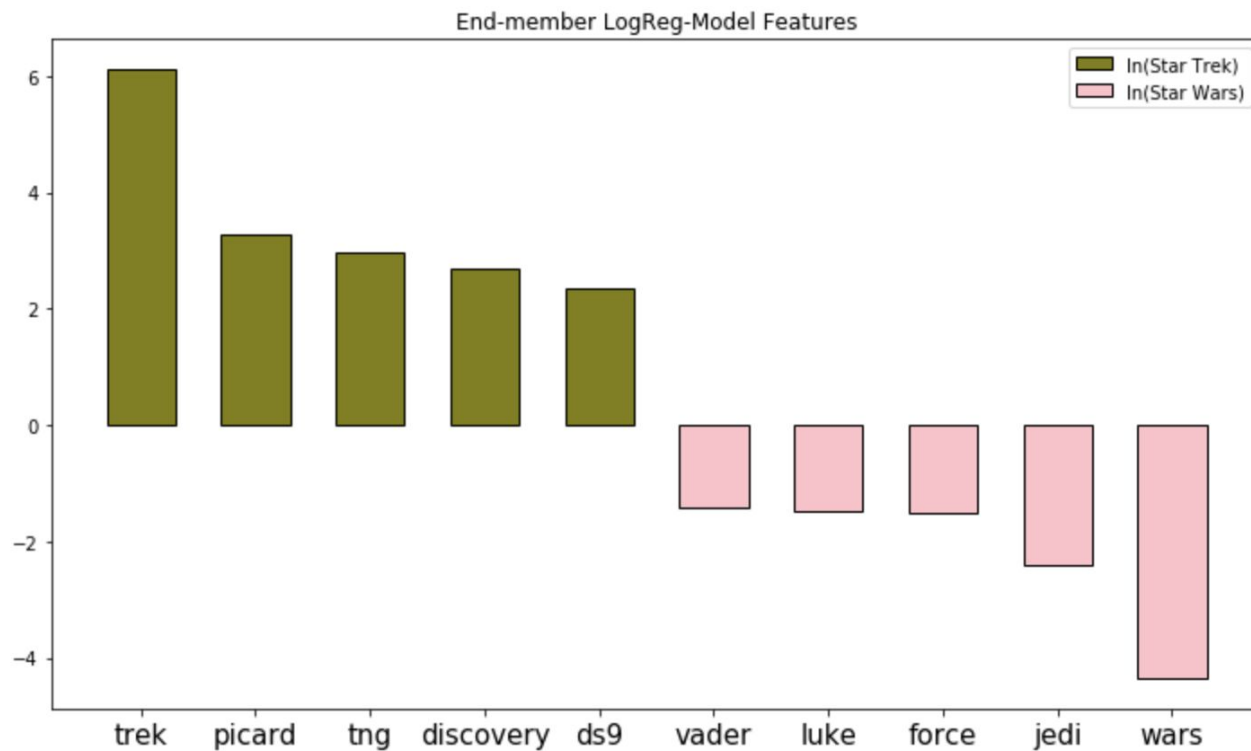This model has the fewest errors in predicting whether a post-title is talking Star Wars or Star Trek. (11.3%)
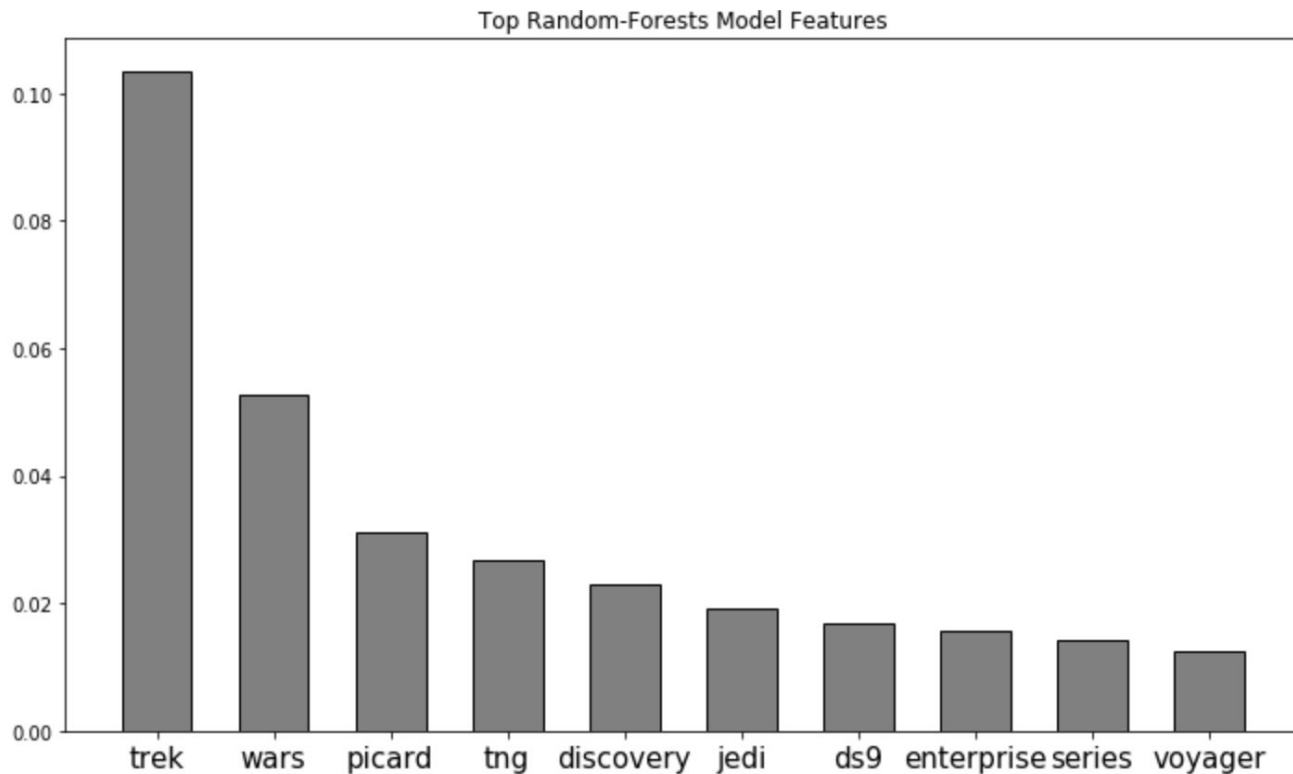

Classification Test-counts

Random Forests - Part 1

Classification Test-counts

This model is better at picking Star-Trek post-titles than Star-Wars post-titles.

Logistic Regression -Part 2

End-member LogReg-Model Features

Random Forests - Part 2

Top Random-Forests Model Features
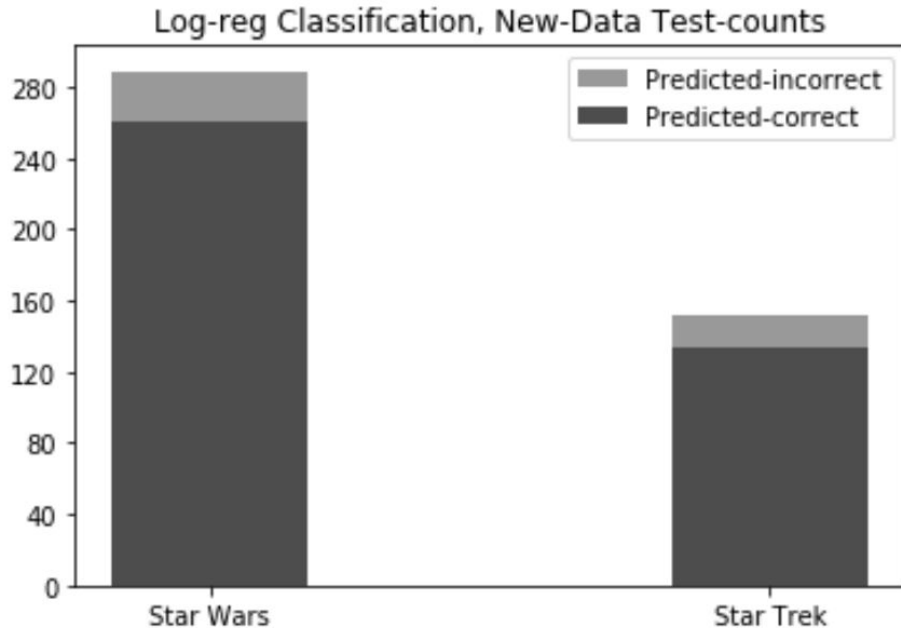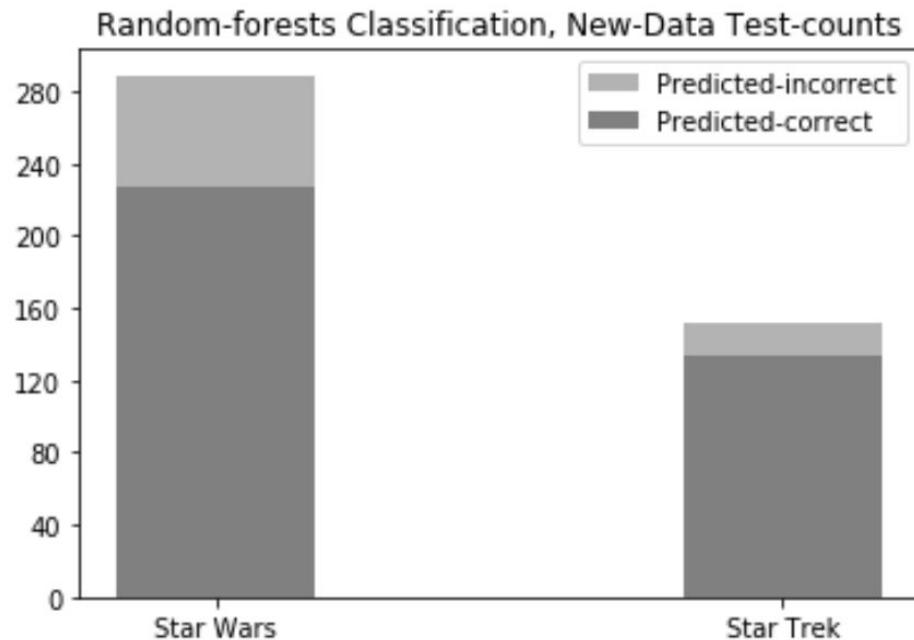
Logistic Regression - Part 3

The logistic-regression model is even good at picking out Star-Wars post-titles from another day. (89.4% accurate)

Log-reg Classification, New-Data Test-counts

- Predicted-incorrect
- Predicted-correct

Star Wars · Star Trek

Random Forests - Part 3

The random-forests model works well for Star-Trek post-titles from another day, despite being in the unbalanced minority.

Random-forests Classification, New-Data Test-counts

The conclusion

Yes!

You can identify Star-Wars-related media titles with almost 90% accuracy.

(Star-Trek language is that distinct.)

Last word

"Those who can, do Star Trek,

Those who can't, do Star Wars."

- William Shatner -