



Predicting Death In Game of Thrones

By Megan O'Rorke & Lucas Chapin | GA Data Science 22 | Instructors: Sinan Ozdemir, Mars Joy, Imeh Williams

Project Overview

Goal: Predict likelihood a character is alive

Best model: Random Forest

Results:

Predicted correctly **82.3%**

Compared to a null accuracy of **74.6%**



"The truth is that people die, and people die in ugly ways, and even some of the good guys die, even people who are loved."

-George R.R. Martin

Data Set

Data set: 1946 characters, 375 columns

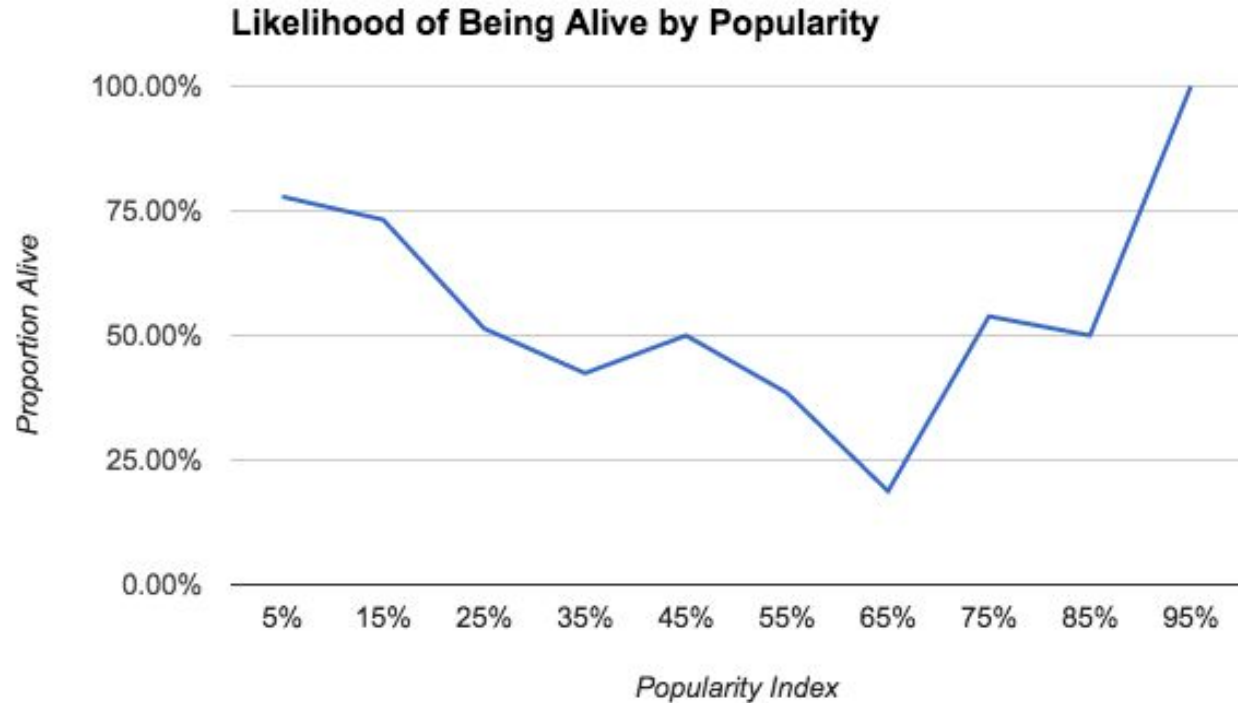
Source: Kaggle

Feature Columns: gender, book appearances, married, noble, number of dead relatives, popularity

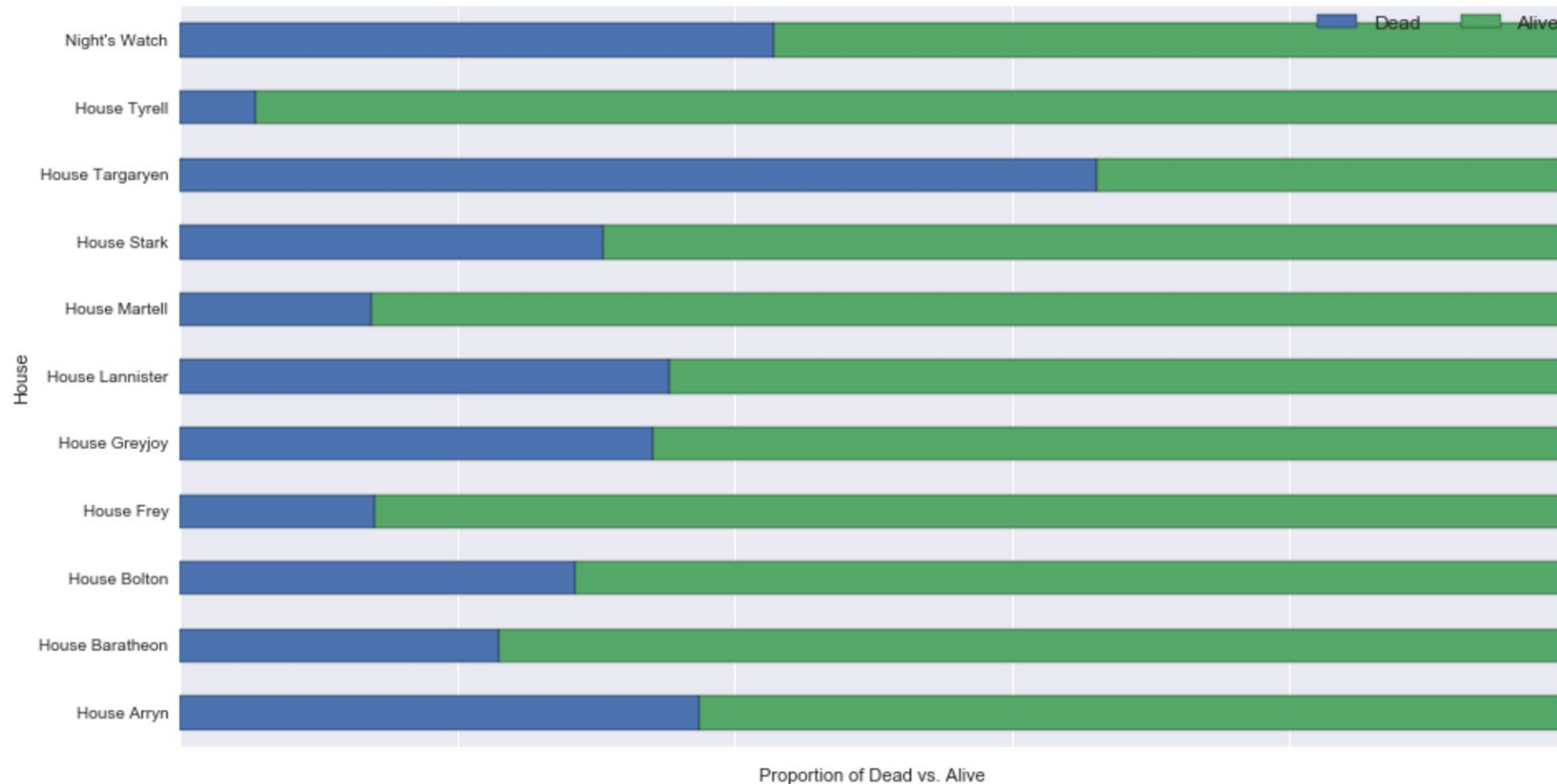
Calculated Feature Columns: shares name with house, culture dummy variables, age dummy variables (groupings)

Response Variable: isAlive

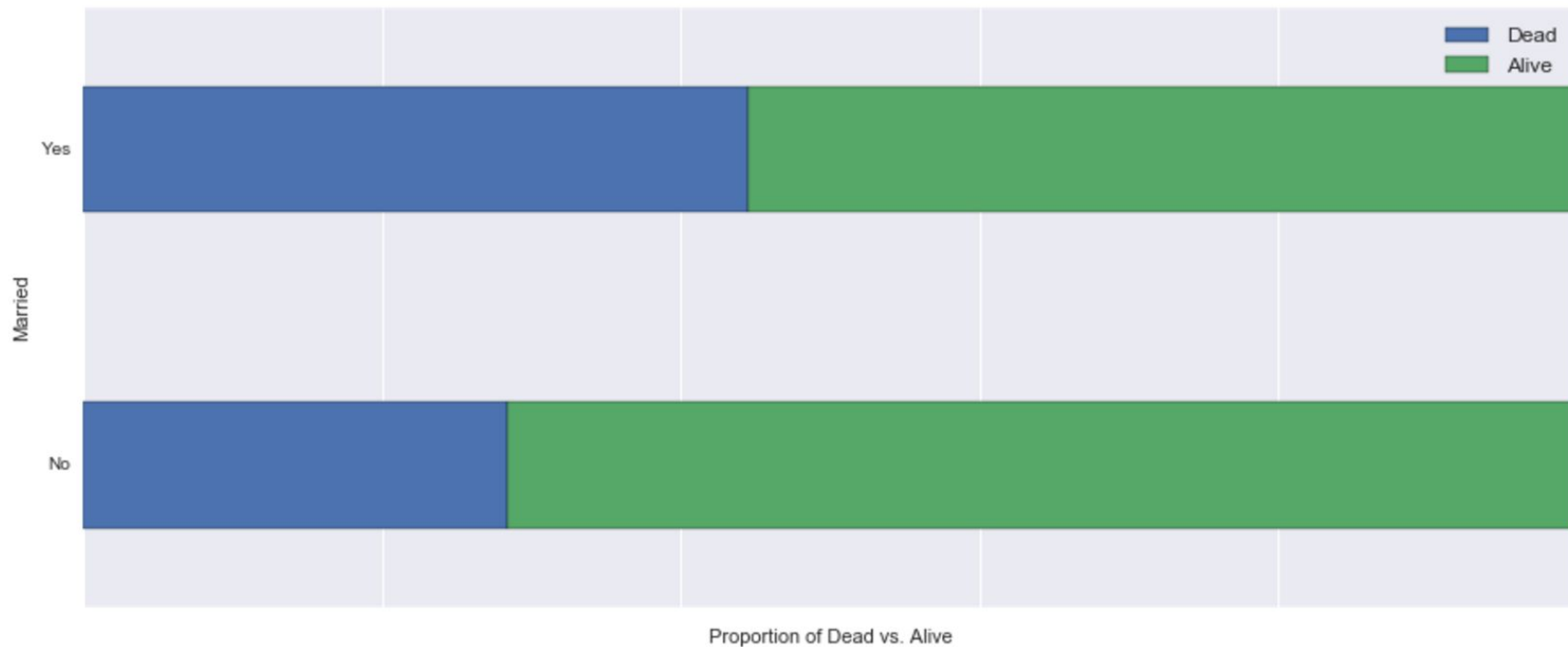
Exploratory Analysis



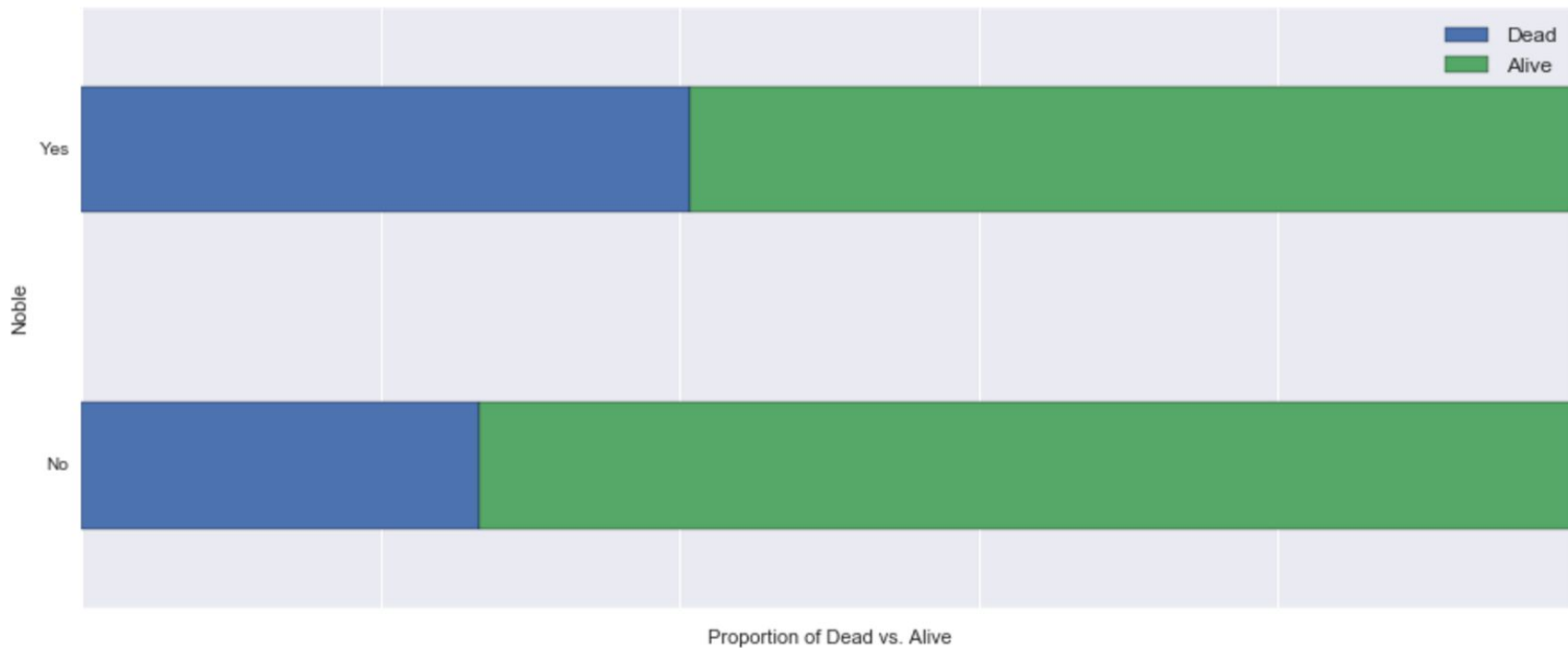
Exploratory Analysis



Exploratory Analysis



Exploratory Analysis

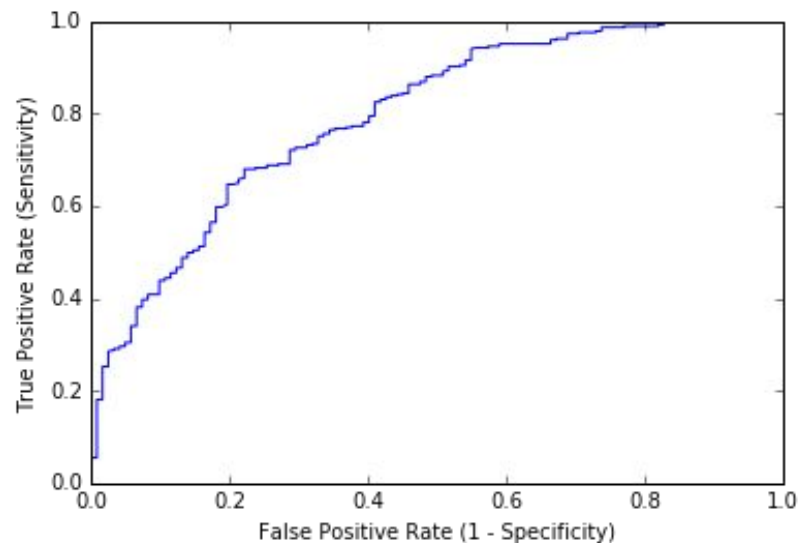
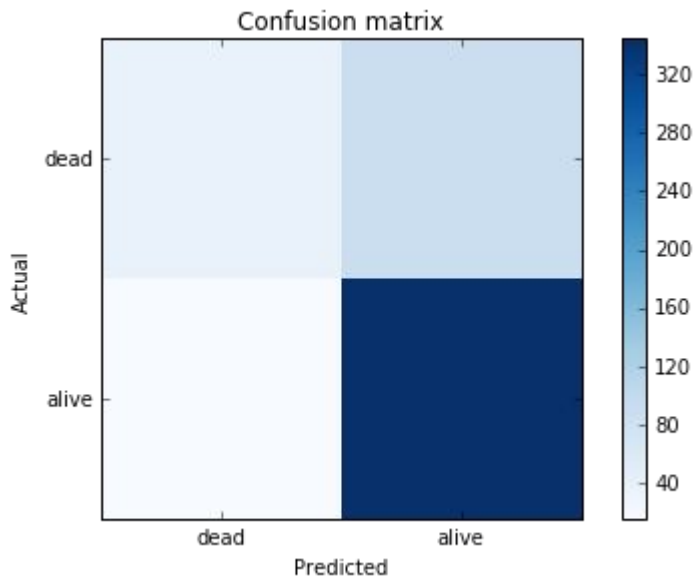


Logistic Regression

Accuracy: 78.2% compared to null accuracy of 74.6%

Sensitivity: 93.1% | Specificity: 38.7%

AUC: 76.7%

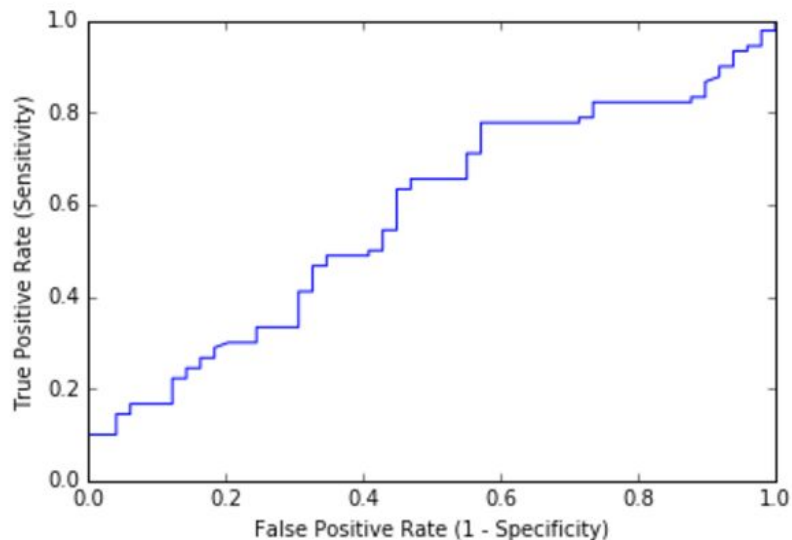
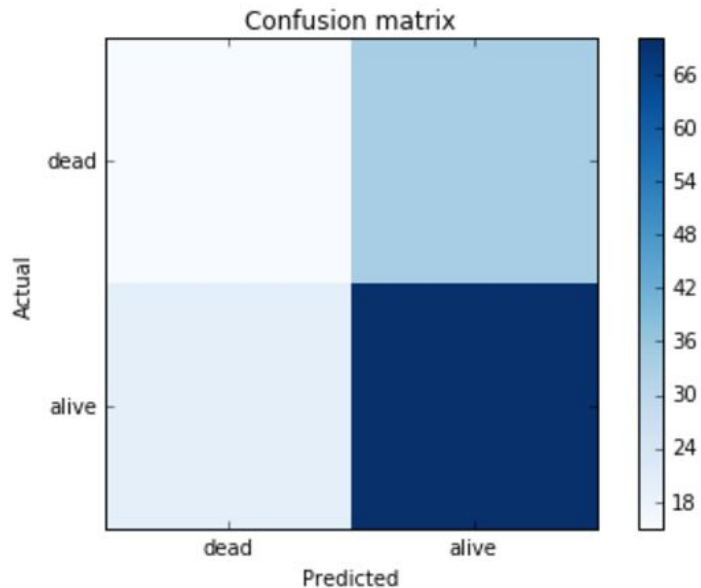


Logistic Regression - Top 9

Accuracy: 75.8% compared to null accuracy of 68.1%

Sensitivity: 77.7% | Specificity: 30.6%

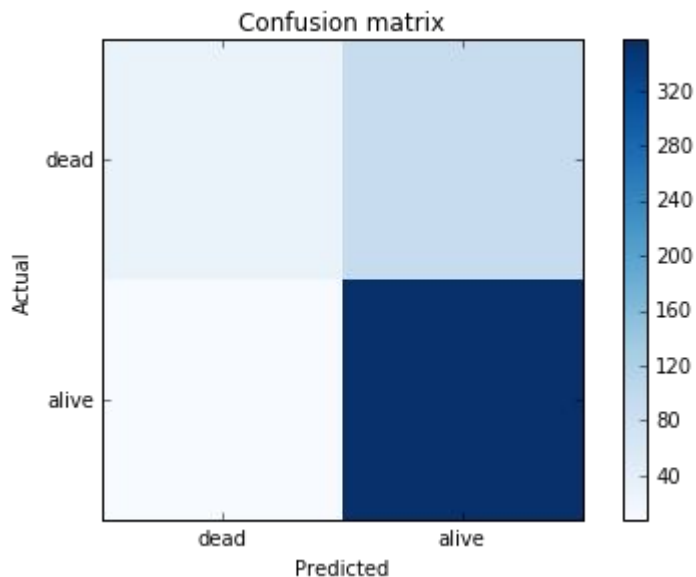
AUC: 57.7%



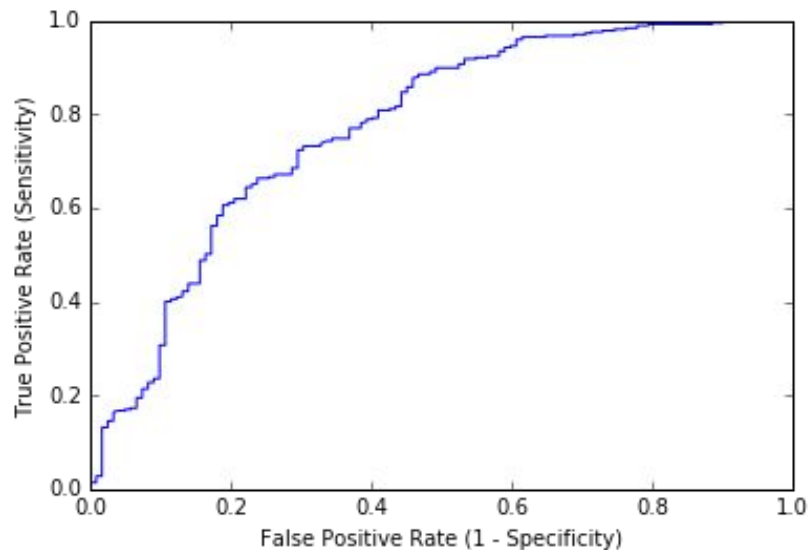
Naive Bayes

Accuracy: 78.0% compared to null accuracy of 74.6%

Sensitivity: 96.4% | Specificity: 29.0%



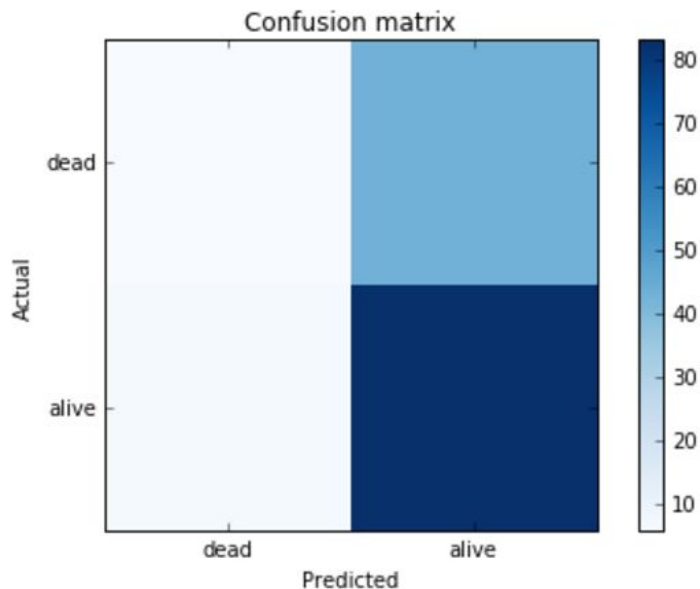
AUC: 73.1%



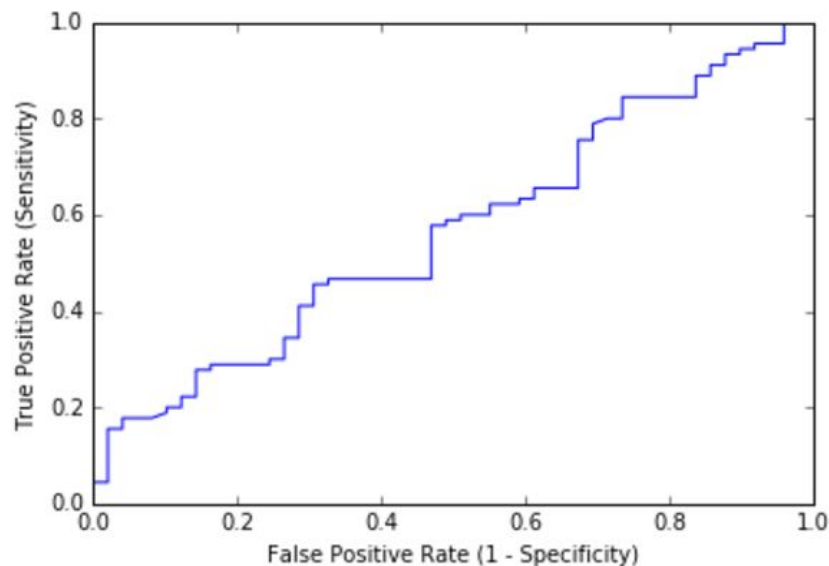
Naive Bayes - Top 9

Accuracy: 64.0% compared to null accuracy of 64.7%

Sensitivity: 92.2% | Specificity: 12.2%



AUC: 56.6%

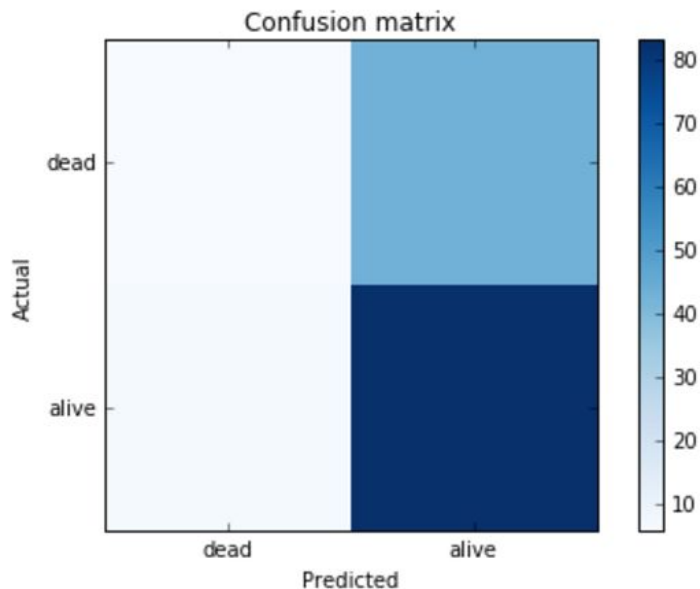




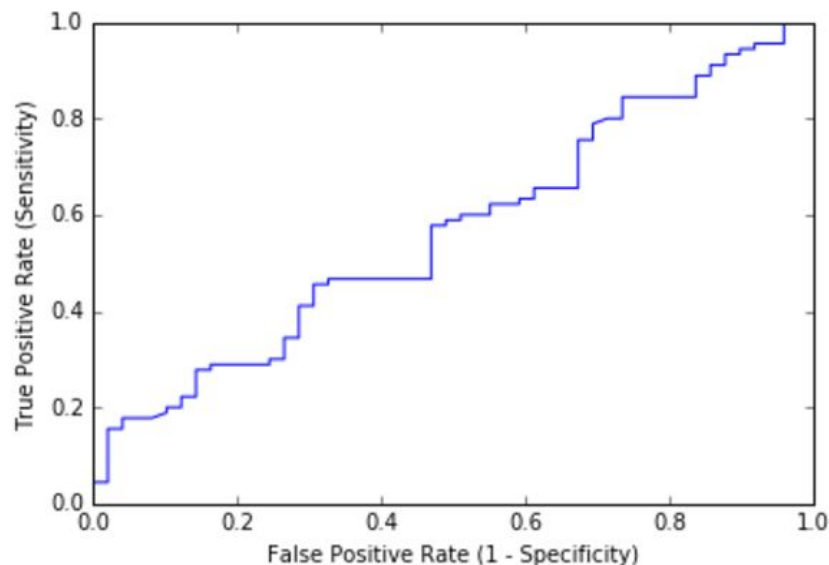
Decision Tree Classifier - Top 9

Accuracy: 64.0% compared to null accuracy of 64.7%

Sensitivity: 92.2% | Specificity: 12.2%



AUC: 56.6%



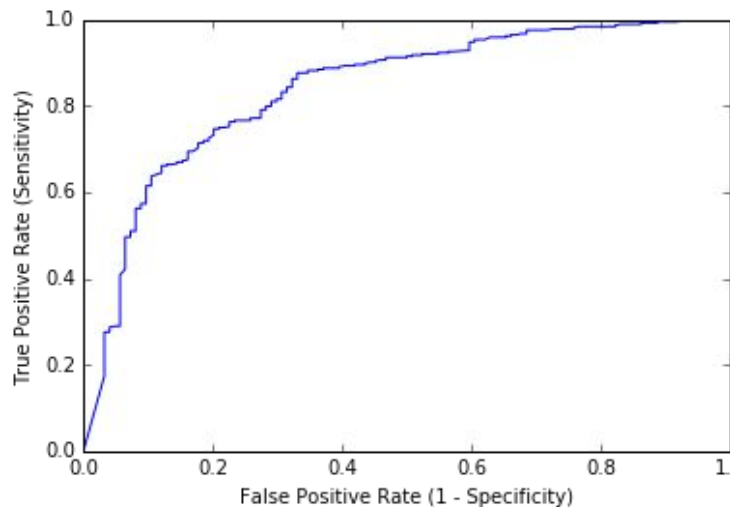
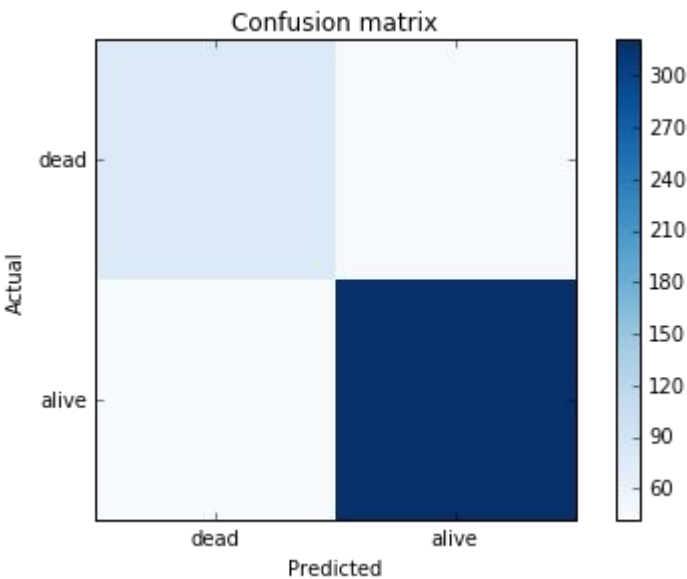
Random Forest

Accuracy: 82.3% compared to null accuracy of 74.6%

(of predicted dead, 65% actually dead. Of predicted alive, 88% actually alive)

Sensitivity: 88.4% | Specificity: 64.5%

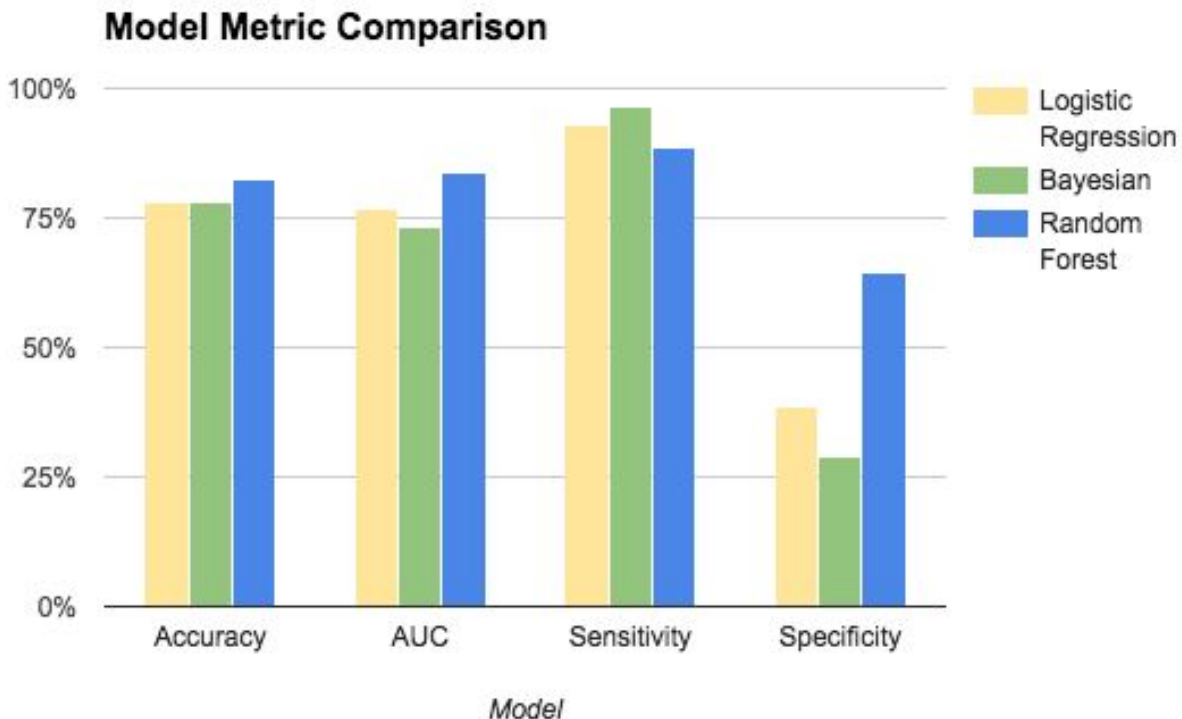
AUC: 83.9%



feature	importance
popularity	0.350557
book4	0.085490
book5	0.044841
isNoble	0.039672
male	0.039564
book2	0.039156
name_in_house	0.039068
book3	0.037651
book1	0.036715
isMarried	0.027985
is_middle_aged	0.026377
is_old	0.025393
boolDeadRelations	0.025031
Northmen	0.025006
Valyrian	0.021170
is_young	0.019720
Free Folk	0.018930
Ironborn	0.016358
Dornish	0.007670
Rivermen	0.006404
Ghiscari	0.006401
Reach	0.005150
Westermen	0.005082
Dothraki	0.004328
Tyrosi	0.004310

Overall Model Comparison

- Random Forest is the clear winner
- While Random Forest edges other models in Accuracy and AUC, Specificity is the real standout
- We care about this most because we want to predict the true negatives (who is going to die)!



Logistic Regression: Which Characters Might Die Next?

Name	Predicted Likelihood of Death
Daenerys Targaryen	82.7%
Tysha	80.2%
Walder Frey	78.1%
Ardrian Celtigar	78.0%
Aegon Targaryen (son of Baelon)	72.5%
Tommen Baratheon	72.0%
Stannis Baratheon	72.0%
Jaime Lannister	70.4%
Bryen	69.0%
Ayrmidon	67.2%
Bran Stark	66.8%
Tyrion Lannister	66.8%



Random Forest: Which Characters Might Die Next?

Name	Predicted Likelihood of Death
Collio Quaynis	69.3%
Toregg	67.2%
Stannis Baratheon	67.2%
Tommen Baratheon	67.2%
Steffon Varner	65.7%
Zei	64.0%
Crawn	63.9%
Parmen Crane	63.6%
Werlag	63.5%
Mother Mole	59.3%
Addam Frey	59.0%
Otho Bracken	59.0%



Future

Publish findings on Kaggle

Update model with additional data

Make a model for TV series