Game of Thrones: prediction model with machine learning in python

Cheng CHENG

Summary

This dataset contains approximately 2,000 characters from the HBO series *Game of Thrones*. I'm going to build a classification machine learning model based on this dataset, in order to predict which characters in the series will live or die. As a result, interesting findings and insights will be shared with data-driven recommendations on how to survive in Game of Thrones.

There are over one thousand missing values in several features, as well as misclassifications of 'male'; for example, "Ser" is the title given to knights, who are exclusively male in that setting, so it is not a gender-neutral title there. ("Knighthood") I tried to impute 'male' feature with its assuming correlation with 'title' feature, but result didn't show significant improvement, therefore I ignored this issue given its not significant substance. Exploratory data analysis is conducted, facilitating feature treatment and engineering. This is not a perfect dataset, I abandoned some good 'features' for I don't want to assign biased value to missing values.

Initially, I used Logistic Regression classifier, of which the predictions result is included in the final excel file for comparison purpose. My final model is Random Forest Classifier, of which the highest mean AUC value after cross-validation is 0.790.

Insights & Recommendations

In general, the earlier a character shows up in the show, the more likely this character is going to die. Male characters are less likely to live; Popular characters are more likely to die; Those who have dead relations are more likely to die.

For those who have incomplete profile in terms of age, mother, father and heir are more likely to live, presumably because they are not popular main characters who are more involved in dangerous life-threatening plots. But within the small group of people who have heir, they have a high chance to live if their heir is alive.

The Valyrian culture group is associated with high risk, so are people from House Targaryen.

Survival tips:

Stay low key, popularity comes with risk. But if you are a main character, make sure your heir safe and sound.

Avoid showing up in book1: A Game of Thrones, but if possible, do show you face in book4: A Feast for Crows.

Be in a family with most relations alive, protect your family.

Stay vigilant if you're from Valyrian or House Taragaryen, or if you're male.

Reference:
"Knighthood." *Game of Thrones Wiki.*http://gameofthrones.wikia.com/wiki/Knighthood