Game of Throne Project

Jiwei Li

Introduction:

There are approximately 2,000 characters in *A Song of Ice and Fire*. Then this book series was made into HBO series Game of Thrones Game of Thrones which is about some constant duel for the kingship fight in the land of Westeros between some important and powerful families. In this dataset, we have 26 columns which contain numeric, binary and categorical variables. And also, there are more than half missing values in some columns. In this passage, we will use feature engineering, variable selection, and model development technique to find out some important factors to predict which characters in the series will probably live or die and give data-driven recommendations on how to survive in Game of Thrones.

Key Insight:

Based on the model summary and random forest importance plot , the top 7 most important factors that affect survive rate in this dataset are age , date of birth, whether appear in book4, whether part of Targaryen house, whether part of Night's Watch house ,popularity, number of dead relations and whether is belong to Valyrian culture . Through the analysis, we can know that except the "book4" all these factors are negative to the "isAlive". What's more, House Targaryen follow the rules of consanguineous marriage so that there will be a high risk of disease in this family and it can lead to premature death. In terms of the Night's Watch house, most of the members have miserable fate and their only task is to guard the cold border wall

against the North barbarians and horrible monsters. So, they are in danger every day and more likely to die. Valyrian is the almost ruined culture and is the cradle of Targaryen house so it also decreases the chance of surviving in this case. Based on the violin plot, it indicate that character who appear in book4 are more likely to survive. One more interesting thing is that a noble character who is married have more opportunity to survive. At last I find two characters age and date of birth are wrong. After searching on the website, I modify them correctly. After obtaining the right age data, I find that characters who have very high age are more likely to survive but characters with very low age are more likely to die. Because the name and S. No have no use to help us predict the column "isAlive" so I just drop them in my analysis. For feature engineering, I convert culture, house and title into binary variables and set the threshold to make it .Also I create new columns which is related to the survival rate .

Recommendations (improving the probability to survive):

1. Be sure to strive to appear in Book4
2. Keep away from Targaryen house, Night's Watch house and Valyrian culture. Try to be a character without these three attributes .
3. Don't be very popular among the series of these books. The low popularity, the higher amount of surviving characters.
4. Try to be a noble and married character

My best model is Gradient Boosted Machines. The AUC score is 0.84. My test score is 0.8923.