Five tips to boost survive chance by five times in the world of game of throne

Binxiao Pan

Machine Learning Assignment 2

March 19, 2019

Abstract

Life is hard and dangerous in the world of Game of Throne (GOT). Every few characters in GOT die for natural cause. The purpose of this report is to find factors could help audiences increasing the chance to survive in case they may shift into the parallel universe of GOT.

Key Findings

The top 5 factors affect survive rate in the GOT world are date of birth, whether show up in book-4, whether part of Targaryen house, popularity, number of dead relations and whether part of valyrian culture. Since characters cannot change their birthdate, they should focus on rest four factors to increase the chance of survive.

Data Cleaning & Feature Engineering

This research excludes variables with missing value percentage more than 80% due to the lack of creditability. The S.No column is removed because it is manually labeled column which wouldn't drive any insights. To predict the variable of isAlive, it doesn't make sense to include dateOfBirth and age since we can directly get the answer through sum them up. This paper excludes both variable since the missing data rate is high and author trying to generate more general insights rather than the difference of generations. For feature engineering, we convert title, culture and house into binary variables like bi_house_house frey which indicate whether the character is in frey house. To reduce bias, we only cover groups with size larger than five.

Modeling

After data cleaning and feature engineering, we filter the variables by its correlation with isAlive column and only keep the top variables with correlation more than our benchmark. (set as 0.1 and increase gradually to see how each variable contribute to the accuracy of model). The benchmark we set finally is 0.17 then get 5 key variables left: book4_A_Feast_For_Crows, bi_house_house Targaryen, popularity, numDeadRelations, bi_culture_valyrian.

```
In [157]: got_df.groupby(['filt','isAlive']).nunique()['name']
Out[157]:
filt   isAlive                        isAlive                     1.00
False  0            470                book4_A_Feast_For_Crows     0.27
       1           1028                bi_house_house targaryen   -0.17
True   0             25                popularity                 -0.18
       1            423                numDeadRelations           -0.19
Name: name, dtype: int64                bi_culture_valyrian        -0.21
                                       Name: isAlive, dtype: float64
```

Recommendations

To increase the chance of survive, author of report highly recommends my audience to follow the following recommendations: 1) Do whatever possible to show up in book_4; 2) Do not be in Targaryen house. If in it already, leave as soon as possible; 3) Reduce your popularity. It is more likely to survive by being a nobody than somebody; 4) Don't let anger overwhelm you. Revenge is one of the reasons you can die in the conflicts. Even sounds cruel, but statistically it makes sense to quit revenge to avoid death. 5) Do not be in Valyrian culture.

Generally, characters follow all these five tips in our datasets (show in book_4, not in either Taragayen house or Valyrian culture, less than population median in term of popularity and dead relation amount) show huge decrease in death rate, which is 5.6% (25 out of 448) compare to 31.3% (470 out of 1498) of the rest characters of datasets.