



# Classification Modeling Case Study

Machine Learning

ELIZAVETA CHAPLYGINA

SAN FRANCISCO, 2019

Content

Summary ..... 2

Key Insights ..... 2

Bibliography ..... 4

## Summary

“Game of Thrones” is a highly rated American TV series, an adaptation of “A Song of Ice and Fire” by George R. R. Martin. The phenomenon of such popularity of the show can be explained by the well-crafted characters of the books and human psychology. Historical, supernatural and mystical elements are combined with a complex plot where it is hard to guess the next event: will the bad characters be punished, who wins the power struggle and, more importantly, who dies next?

The purpose of this analysis is to predict if the character is dead or alive based on 24 different variables such as the title, house to which the character belongs to, which book he or she was mentioned, age, popularity rate, etc.

## Key Insights

Our data consists of 1946 characters, almost 75% of which are marked ‘alive’ by the end of book 5: “A Dance with Dragons”. Around 62% are men and 212 characters appeared in all of the books. Another noticeable fact about the data is a big amount of missing values for such variables as date of birth, title, whether parents are alive and culture.

In the external research done previous to working with the data it was found that the house name is a potentially good predictor of the likelihood of survival. If the reader would follow the plot, it is possible to track which houses of Westeros are currently at war and in danger of dying altogether. Most of the House Targaryen is dead, while the House Frey is safe for now. Night’s Watch is a military order and is in a risky position of wildling attacks and dangerous expeditions. Belonging to those houses could serve good predictors of the likelihood to survive/die.

In terms of the relations between the character’s popularity and survival our data does not always follow common trends: the most noticeable death gaps are seen for non-popular characters (rate 0-0.013) and extra popular (rate 0.73-0.86). Recommendation in this case would be to seek the fame as the likelihood to die if the character is not popular at all is slightly higher than in the middle.

It is interesting to notice some of the heroes mentioned in the dataset have not appeared in any of the books (272 characters). It could mean that the reader knows that they exist (e.g. old kings or next book heroes) but there is no storyline for them in the books.

The last noticeable key factors are age or date of birth (if mentioned). The motivation to use these data is to predict who dies simply because they are old.

Note:

**AUC** value is 0.809

## Bibliography

[https://en.wikipedia.org/wiki/Game\\_of\\_Thrones#Cultural\\_influence](https://en.wikipedia.org/wiki/Game_of_Thrones#Cultural_influence)

<https://www.statista.com/topics/4187/game-of-thrones/>

<https://www.hbo.com/order/game-of-thrones/>