****

Merve Ozgul

Hult International Business School

A Song of Ice and Fire: How to Stay Alive?

**Key Issue**

Literature Consulting Group was hired by George R.R. Martin, to present actionable key insights and recommendations on character survival in Game of Thrones for the next book. A Song of Ice and Fire is a series of five epic fantasy novels where characters fight over the throne to rule the world and is set in a fictional world that has three discovered continents as Seven Kingdoms of Westeros, Essos and Sothoryos. The series involves about 2000 characters from 39 different cultures and from more than 300 houses, wars and bizarre and supernatural events. As the storyline gets more complex, and number of characters increase, it becomes arduous for characters to remain. Moreover, it becomes more challenging and tenser situation for George R.R. Martin to create a satisfactory next book, after the tremendous popularity of HBO series Game of Thrones Series and books (indiewire.com). Consequently, objective of the analysis is to predict which characters will live or die in the book series and provide recommendations to George R.R. Martin for character killing strategy.

**Findings from Exploratory Data Analysis**

* There are five books that are already published: A Game of Thrones, A Clash of Kings, A Storm of Swords, A Feast for Crows and A Dance with Dragons. Most of the characters were introduced to the readers with the first book A Game of Thrones and only 62% of those characters survive. First book has the lowest survival rate as it tells about the civil war going on between noble houses over the throne (www.shmoop.com). Additionally, some characters in the dataset did not appear in any of the five books. It might be due to the fact that they do not have a great influence over the important events in the book.
* Majority of the characters are male and death rate for them is greater than the female characters.
* House Targaryenis the only house that majority of the characters are dead (Figure 1). According to fandom.com, The Night’s Watch is the most crowded house, a military order that protects Westeros which has a high rate of death characters in our dataset. On the other hand, majority of characters from House Stark, House Frey and House Tyrell are alive.
* People from Northmen culture live in the North of Westeros, which is the largest region in the entire Westeros, along with other cultures as Andal, Cronnogmen and First Men. The ruler house of the North is House Stark (belongs to Northmen culture) according to gameofthrones.fandom.com.

**Key Insights from Prediction Models & Recommendations for Survival**

Four different classification models were used in order to predict which characters will die or live and come up with a survival strategy (Figure 2). Models The insights were driven based on the features that contributed to the model to learn most.

* Being popular is troublesome for characters to survive. Additionally, popularity has a strong positive correlation with the total number of deceased relatives of a character.
* Appearing in more than one book influences whether the character will die or live (Figure 3). If there are more events that the character involves, he/she is more likely to survive. Especially, appearance in the fourth book Feast for Crows is important to survive.
* Although it is hard to tell about life expectancy of characters as there is not enough information about the time concept in the novels, date of birth and age still have great influence on a character’s death.
* Region that character is from impacts whether the character will die or survive. Both people from Essos and Westeros have a high chance of survival, however living in Essos seems to provide a better chance of survival. This might be related to Essos being more developed than Westeros. Additionally, extreme winter conditions in Westeros and being close to the Wall might contribute to a weaker survival rate.

Based on these facts, the characters that should survive but in Westeros can escape to Essos before the long winters or any threat (i.e. Arya Stark ending up in Essos, after although she is originally from Westeros). Additionally, character might fight for his/her beloved ones to avoid the increase in the number of deceased relatives or send his/her relatives to a safer place if they are under danger. Finally, resurrecting a character can be an option, if he/she really has to live since it is a fictional world (i.e. HBO TV Series, Resurrection of Jon Snow). However, this can be a big wildcard and can be only applicable for the major characters.

**Limitations & Suggestions for Improvement**

* **Better Data Collection & ETL:** The quality of the data was poor, as it included vast amount of missing values. Better data collection and ETL process (Extract, Transform, Load) can lead to improved results.
* **Using other Ensemble Models:** One of the greatest challenges is predicting the number of dead characters correctly. This is due to the class imbalance, there are more characters alive than the characters who are dead. Using algorithms such as Gradient Boosting Machines (GBM) can improve the prediction performance.

**Appendices**

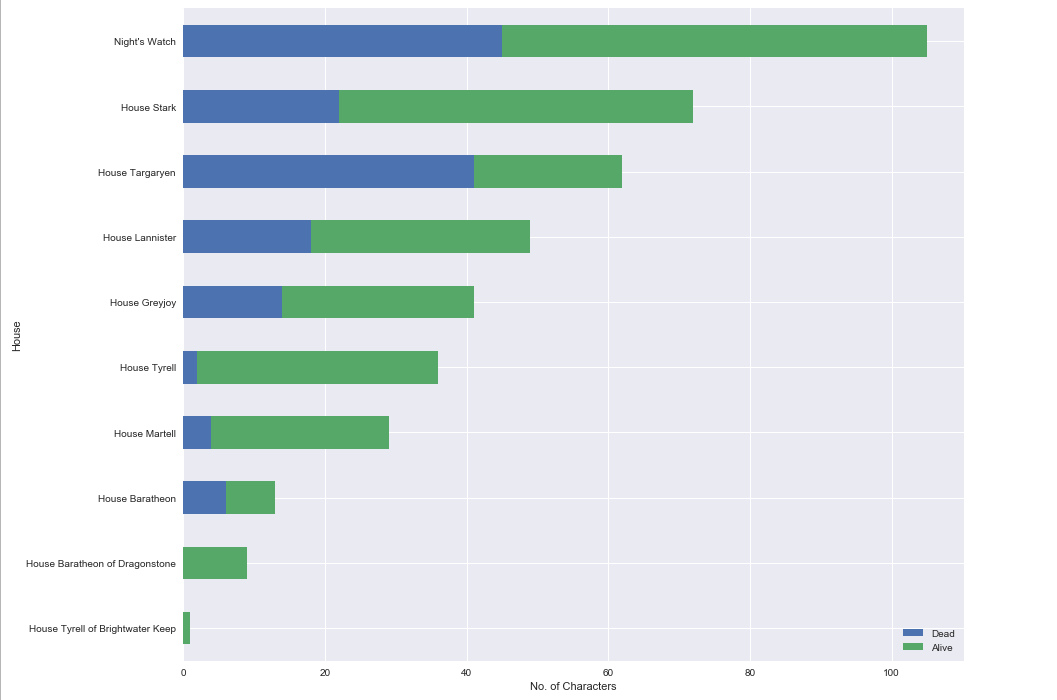
Figure 1: Houses & Survival Rate

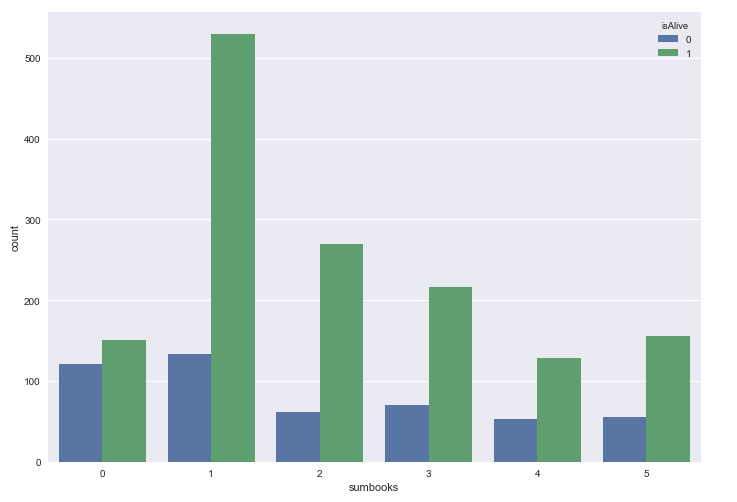
Figure 2 – Methodology & Model Performances

**Methodology**

Predicting which characters will survive is a binary classification problem. The steps taken consist of data exploration, data preparation, model building and prediction analysis. Additional features related to culture based on regions, great houses and total appearance in books during the data preparation process. Three single classification models (Logistic Regression, KNN Classifier, Decision Classifier Tree) and one ensemble model (Random Forest) were used to predict whether character will live or die. Models’ performances were evaluated based on Area Under the Curve (AUC), confusion matrix, cross validation scores and accuracy, as these metrics measures the number of true predictions for dead or alive characters.

**Model Performances**

|  |  |  |
| --- | --- | --- |
| Model Name | Scores | Comment |
| Logistic Regression | **﻿Training Accuracy Score** 0.799  **Testing Accuracy Score:** 0.7692  **Training AUC Score** 0.6386  **Testing AUC Score** 0.5828  **Confusion Matrix:**  ﻿[[ 10 40]  [ 5 140]]  **Cross validation scores:**  ﻿**Average:** 0.795  Minimum: 0.772  Maximum: 0.818 | Logistic Regression was an overfitted model. Both AUC and accuracy scores are really low. The model overestimates the number of people that are going to be alive.  Note: Oversampling instead of stratified sampling gave lower scores. |
| KNN Classifier | **With all features**  **Training Accuracy Score** 0.7613  **Testing Accuracy Score:** 0.7487  **Confusion Matrix:**  ﻿[[ 23 27]  [ 8 137]]  **Cross Validation:**  **﻿Average:**  0.822  Minimum: 0.815  Maximum:0.826  **With selected features**  **﻿Training Score:** 0.8749  **Testing Score:** 0.8564  **Training AUC Score:** 0.8013  **Testing AUC Score:** 0.779  **Confusion Matrix:** [[31 19]  [ 9 136]]  **﻿Average:** 0.819  **Minimum**: 0.806  **Maximum:** 0.831 | KNN Classifier tend to overfit. Although it performs better than logistic regression, it is not fully reliable. Scaling the variables might improve the model performance.  After selecting some features, model has improved and there has been a decline in the number of falsely alive predicted characters (Decrease from 40 to 19) without compromising from the correct predictions for alive characters. |
| Tree Model | ﻿**Training Score:** 0.8372  **Testing Score:** 0.8256  **﻿Training AUC Score** 0.7427  **Testing AUC Score** 0.7241  **Confusion Matrix:** [[ 30 20]  [ 22 123]]  **Cross Validation:**  **﻿Average:** 0.808  Minimum: 0.789  Maximum:0.819 | Tree model returns high scores of accuracy and AUC. However, model tends to overfit. Number of correct predictions for alive characters could be higher. |
| Random Forest | ﻿**Training Score** 0.7858  **Testing Score:** 0.7795  T**raining AUC Score** 0.7883  **Testing AUC Score** 0.7862  ﻿**Confusion Matrix:** [[ 40 10]  [ 33 112]]  **Cross Validation:**  ﻿**Average:** 0.792  Minimum: 0.787  Maximum: 0.801 | Instead of running the random forest on stratified sample, I used the class\_weight = ‘balanced’ to handle the class imbalance. Tuned vs. Not tuned model can be compared with a ROC curve to decide which one is better off. |

Figure 3 – Total Appearance in Books & Survival Rate

References

A Wiki of Ice and Fire. (n.d.). Retrieved March 18, 2019, from https://awoiaf.westeros.org/

Cooper, G. F. (2018, August 15). Some Game of Thrones TV deaths won't happen in books. Retrieved March 18, 2019, from https://www.cnet.com/news/george-r-r-martin-some-game-of-thrones-tv-deaths-wont-happen-in-books/

Game of Thrones Wiki. (n.d.). Retrieved March 18, 2019, from https://gameofthrones.fandom.com/wiki/Game\_of\_Thrones\_Wiki

Ramesh, S. (2015, January 25). Why don’t the people of Westeros migrate to Essos during the long winters. Retrieved March 18, 2019, from https://www.quora.com/Why-dont-the-people-of-Westeros-migrate-to-Essos-during-the-long-winters

Sharf, Z., & Sharf, Z. (2018, November 12). George R.R. Martin Says He's Struggling to Finish the Next 'Game of Thrones' Book Because the Show Is So Popular. Retrieved March 18, 2019, from https://www.indiewire.com/2018/11/george-rr-martin-game-of-thrones-struggling-winds-of-winter-1202019930/

Shmoop Editorial Team. (2008, November 11). A Game of Thrones. Retrieved March 16, 2019, from https://www.shmoop.com/game-of-thrones-book/