Data Analysis In “Game of Thrones”

Leyi Wang	001493704
Xiao Liu	001472735
Yan Lyu	001493900
Yitong Liu	001494304

Background

"Game of Thrones" is a wonderful TV series adapted from the novel "Song of Ice and Fire" of George RR Marti. This TV series is centered around the struggles between the nine families, with its complex political landscape, a large number of characters and deaths of characters. It has shaped thousands of characters with vivid performance and different personalities.

However, due to cultural, geographical and linguistic differences, the audience who first watched Game of Thrones (such as us) often could not figure out different characters and could not understand the background of the story. After analyzing the dataset of Game of Thrones in Python and coming to some general conclusions, the audience who are new can have a better understanding of this TV series.

Moreover, as a fanatic, we can also get a deeper understanding of the play from the following analysis: for example, to establish different models to predict the death of the characters, we can have an experience as a scriptwriter; to analyze different battles according to the model built, we can further understand the motivation of the characters in the play. In this way, completing such a project is a perfect combination of interest and learning.

Summary

1. Demonstrate the character relationship of the characters in “Game of Thornes”
2. Predict the death of a character in “Game of Thornes”
3. Analyze the battles in “Game of Thornes”

Specific Aims

(1) Data preprocessing

Now, we have the dataset of wars between different countries and death characters. To better build model, we firstly preprocess the data.

Preprocessing operation includes recuperating missing data, classifying data, dividing training sets and test sets, feature scaling, and dimensionality reduction, and so on.

(2) Data analysis

Before we formally build the model, we need to have a general understanding of what the data is.

In the dataset character-deaths.csv and character-predictions.csv: the age distribution of the dead characters, the country distribution, and the relationship with the survivor can give a better predictor of the factors that influence the death of characters.

In the dataset battles.csv: It’s easier to know which countries often contradictions and which countries have had a peaceful and friendly relationship.

(3) Building a model

We plan to use the four algorithms- naive Bayes, linear regression, random forest, and decision tree to predict the death of characters.

(4) Model comparison

According to the four models built above, compare the pros and cons of each model and select the best one.

Raw Data

battles.csv:

name, year, battle_number, attacher_king, defender_king, attacker_1, attacker_2, attacker_3, attacker_4, defender_1, defender_2, defender_3, defender_4, attacker_outcome, battle_type, major_death, major_capture, attacker_size, defender_size, defender_size, defender_commander, summer, location, region, note,

character-deaths.csv:

Name, Allegiances, Death Year, Book of Death, Death Chapter, Book Intro Chapter, Gender, Nobility, GoT, Cok, FfC, DwD