72

CMSE 381 Final Report: Avengers Dataset

Jaclyn Frishcosy

Integrity Statement:

I, Jaclyn Frishcosy, understand that integrity is a big building block of a functional society. Because I care about what others think of me and my character, I commit to working hard for my grades, citing sources when necessary, and creating my own work, not taking someone else's. I acknowledge that I am aware of the MSU ethical standards for integrity.

Introduction/Problem Statement

The Avengers may be superheroes, but they are not invincible. In fact, many members of the Avengers have died. However, death does not seem to be permanent for these characters. In this report, I aim to analyze the statistics surrounding deaths of the Avengers from 1963 through 2015 using basic Python methodology and Support-Vector Machine modeling in order to better understand the Marvel universe and how the comic writers decide to kill off their characters.

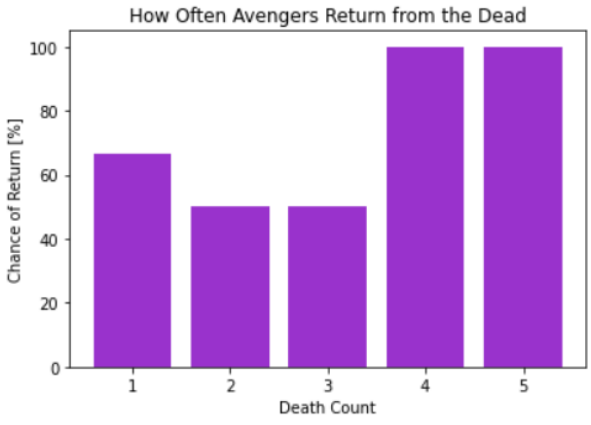
The dataset I will be working with is avengers.csv, which consists of 173 different Avengers and 21 variables for each Avenger. Some of these variables are not of use to the models or statistics, so they are omitted from the data as needed.

Related Work

The article, "Joining The Avengers Is As Deadly As Jumping Off A Four-Story Building" by Walt Hickey (https://fivethirtyeight.com/features/avengers-death-comics-age-of-ultron/), covers the Avengers data mentioned above. This article does not have any graphs, but it goes into some detail regarding the number of Avengers that have died and how many times each has died.

Hickey found that 69 of the 173 Avengers have died at least once within the time range of the data (1963 to 2015), which is equivalent to about 40% of Avengers. There was a total of 89 deaths, with 57 total resurrections, within these 69 Avengers.

Of the 69 who died, 46 characters returned from the dead – an astounding 66.7% recovery rate from their first death. However, as seen in the below bar graph, only 50% of characters recover from their second and third deaths.



The graph also shows that 100% of characters who die a fourth or fifth time get resurrected. This is true, but is an anomaly. The only character who died more than thrice from this dataset is Jocasta, who has never stayed dead. According to the Marvel wiki page for Jocasta, she is an android who has been put back together every time she has been destroyed ("died").

Additionally, based on the mortality and recovery rates for the Avengers, Hickey found that on average, an Avenger has died once every 7.15 months, with a permanent death every 19.88 months – slightly over one in every three deaths.

My Model

The statistics about the number of times Avengers have died are interesting, but I would like to know more. For example, does the amount of time a character has spent as an Avenger factor into how often they die? What about membership? Number of appearances? Gender?

To test some of these theories, I ran two different SVMs on the data, dividing the data by class. I started by splitting the data into training and testing sets, with the labels being membership.

****Note****: I will drop all columns past Death1 for this section due to the large amount of NaN values in the other death/return columns. Thus, this model will focus on only the first time dying. There are only 2 probationary members of the Avengers, which is not enough to form a proper fit, so I will be excluding those two members as well.

I then fitted a Support-Vector Machine model to the membership training data and tuned hyperparameters.

****Note****: I will provide the output of the cell used to train the model due to the long runtime. This is pasted directly from my Jupyter Notebook.

```
Fitting the classifier to the training set
Best estimator found by grid search:
SVC(C=100, class_weight='balanced', gamma=0.001)
Runtime 2122.3442752361298
```

Once the membership model was fitted, I then split the data into training and testing sets again, this time with the labels being gender. I fitted an SVM to this data as well and tuned hyperparameters.

Results

We now have SVM models for our data based on both membership and gender, but how accurate are our new models? I ran an accuracy test on both models via classification report, confusion matrix, and accuracy score.

For the membership model:

```
              precision    recall  f1-score   support

           0       0.97      0.91      0.94        35
           1       0.80      1.00      0.89         4
           2       0.60      0.75      0.67         4

    accuracy                           0.91        43
   macro avg       0.79      0.89      0.83        43
weighted avg       0.92      0.91      0.91        43

[[32  1  2]
 [ 0  4  0]
 [ 1  0  3]]
Accuracy: 0.9069767441860465
```

The model predicted 39 test points correctly, with three false positives and one false negative. This results in a 90.7% accuracy. Overall, the model is a very good fit. The number of false positives is a bit high for my liking, but it would function well in a real-world application.

For the gender model:

Unfortunately, due to the run time and a possible crash of my Python kernel, I was unable to finish fitting the gender SVM model, and since I couldn't restart the kernel for fear of my first model having to be rerun (which took over 35 minutes, as is seen by the printed runtime), I was unable to run its accuracy test. I tried everything I could think of to get my results, including re-running my code in a new notebook while the old one was still running, but alas, it did not work.

Discussion and Conclusion

Based on the results from my model, it does appear that membership level has an effect on whether a character dies or not. Considering that an android died and returned on five different occasions, I doubt that the writers of the comics care much about whether an Avenger is a full-fledged member or an honorary member; all superheroes are created equal.

Because of the aforementioned reasons, I have no conclusion to my gender model. If I were to make a guess based on the data, it is likely that gender does not have a strong effect on whether a character dies or not. I have seen a few Avengers movies, and gender equality does seem to be a focus; granted, the only movies I've seen were likely created after 2015.

I found that the statistics regarding percentage of revival were oddly rational. The probability of surviving the first death is exactly two-thirds, and the probabilities of surviving the second and third death are both exactly one half. It is possible that the Avengers comic writers purposefully try to commit to a certain number of deaths and revivals—the anomaly, of course, being Jocasta, as mentioned above.

****Note****: I noticed an hour before the deadline that I used the completely wrong model for the data. An SVM does not make sense for correlation. SVMs are used for classification. However, I

have no time to rewrite my entire report, so I must make do with what I have. I will make sure I'm ready for the final exam, and hopefully, I'll do well enough on that to pass the course.

Works Cited

Hickey, Walt. "Joining The Avengers Is As Deadly As Jumping Off A Four-Story Building." FiveThirtyEight, ABC News, 12 May 2015, fivethirtyeight.com/features/avengers-death-comics-age-of-ultron/.