

AI Internship Project

Movie Reviews

by: Aarshika Kanjani, Ronak Pai, Angus Bharucha

Project Description

Idea:

The movie end game was one of marvel's most famous movies.

It was most famous as it was the ending of the avengers faze. However, this movie is not liked by everyone, some people like it and others don't,

so let's analyze the movie reviews to find out what people feel about it.

The idea behind this project is to give people recommendations based on the movies they watch and reviews of the people

who have watched that particular movie.

Problem statement:

Criticizers and supporters of the marvel fan base input their reviews about the movie.

The problem we are trying to solve is to see or view the percentage of people who liked or didn't like the film and test

the machine to see how accurate it is.



Data Description

Data set:

- Username
- Content
- Score (out of 3)
- Gender (1= Female, 2= Male, 3= Rather not say)-
- Date

Why is Gender important?

Gender is important because it depends on the viewer or the movie, for example, boys might prefer marvel movies more than girls as it involves

things that come to their liking, this can affect the data because there might be more guys who prefer the movie over girls.

Why is the date important?

The date is important as it gives an idea of when the person has watched the movie either at the time of the release

or a few months or years down the line. It would show the interest of the person and how engrossed he or she is in the movie.

Methodology

- So far we have collected the relevant data needed to train and test the model, now we need to create the model.
- To build a movie review model, we first need an algorithm that will act as the foundation from which the program can be coded.

These are the algorithms that we implemented in this machine learning project:

1) Linear Regression:

Using linear regression, its analysis is used to predict the value of a variable based on the value of another variable. This can be useful to others as the output can predict future decisions in movies.

The graphs that we are going to be using:

- 1) Pie Chart
- 3) Bar Graph
- 4) Scatter Graph

Libraries:

- 1) Pandas — For handling structured data
- 2) Scikit Learn — For machine learning
- 3) NumPy — For linear algebra and mathematics



Training:

Now we can begin with the actual method as to how we are going to train the model.

We have our data collected and organized, we first need to separate the data into train and test.

We do this using the `train_test_split` function, which will split all the data values into variables such as:

- `x_train`
- `x_test`
- `y_train`
- `y_test`

The parameters used in the function will be the predictor variable and target (predicted feature) variable. Now we import `.naive_bayes` (our classifier algorithm) from `sklearn` to train the model Use the `.fit` function - `x_train`, `y_train` as parameters

Training accuracy:

We can find how accurate our model is after training, using `.accuracy_score` from `sklearn`.

How will our project solve our problem?

Our project will solve the problems of future viewers who want to watch Marvel movies as our model will predict the ratings and reviews.

The output(predictions) of our model will help future viewers to decide whether to watch the upcoming movies. Our visuals that we have implied

in the project will help decide for people whether they would really want to watch an upcoming marvel movie or not, as we are using endgame as our example

this can also be implied for other future movies as well.



References

- <https://www.kaggle.com/code/nalkrolu/marvel-movie-dataset-eda-sql-and-plotly>
- <https://thecleverprogrammer.com/2022/04/12/tiktok-reviews-sentiment-analysis-using-python/>
- <https://www.upgrad.com/blog/data-structures-algorithm-in-python/#:~:text=What%20are%20algorithms%20in%20Python,guid%20the%20writing%20of%20algorithms.>
- <https://www.guru99.com/python-tutorials.html>

Conclusion

Our project's main aim is to predict the movie reviews and identify the target audience of a certain genre of movies we chose. This can help the movie producers and directors to understand the appropriate audience and accordingly target and segment. This will increase valuable viewership of the movies which indefinitely benefits the stakeholders along with the producers of the movies. Another merit of our project is that it can be edited and worked upon in the future as well as our project will become more accurate and enhanced with more movie reviews and inputs which can be worked upon in the future.

—