

PYTHON PROJECT

NAME : Neeraj Roop Venkatesan

UTA ID: 1001644243

Based on the top 50 movies from Imdb, Metacritics, Rotten Tomatoes the following inferences were made. I fetched the movie names, their ratings, reviews and genres from the 3 websites using BeautifulSoup python library and created a CSV file. I took all the reviews of top 50 movies and created three separate dictionaries of all unique words, except the stopwords, and the count of each word. By comparing the common **top 50 frequently used words** from the websites using a certain **similarity metrics**, it is found that IMDB and Metacritics have more words in common indicating that their usage of words (**Word style**) is more similar. But when the frequency is brought into comparison, they have the **similarity score of 88%**. IMDB and Rotten Tomatoes have least number of common frequent words but with **91% similarity score**. Metacritics and Rotten Tomatoes have **90% similarity score** with a decent amount of common frequent words. In my point of view, **the count of common words also influence the similarity score**.

SIMILARITY SCORE BASED ON COMMONLY USED WORDS IN REVIEWS

```
The similarity score between IMDB and metacritics is: 0.8874
Similar words are - ['family', 'whose', 'young', 'world', 'help', 'one', 'two', 'men', 'war', 'years', 'life', 'new']

The similarity score between IMDB and Rotten Tomatoes is: 0.9103
Similar words are - ['life', 'one', 'new', 'family']

The similarity score between Metacritics and Rotten Tomatoes is: 0.9074
Similar words are - ['family', 'stars', 'new', 'one', 'time', 'life', 'film', 'story']
```

By comparing the similarities in movies (**Top 50 from each websites**) across each of the websites, it is found that only IMDB and Metacritics have common movies. Rotten Tomatoes does not have any movie in common. So it is not possible to check if Rotten Tomatoes and other websites categorize the movies similarly based on the genre of the movies.

SIMILAR MOVIES BETWEEN TWO WEBSITES

```
The similar Movies between IMDB and Metacritics is: ['city lights', '12 angry men', 'parasite', 'casablanca', 'psycho', 'the godfather']
The similar Movies between IMDB and Rotten Tomatoes is: []
The similar Movies between Metacritics and Rotten Tomatoes is: []
```

By comparing the Genre of the common movies between IMDB and Metacritics, It is found that, 4 out of 6 common movies have exactly the same genres, while the other two is almost nearly similar. Getting a **similarity score of 77%**. I infer that by comparing more common movies, the similarity scores can increase. **But the scope of this project is limited to top 50 movies only**.

city lights	['Comedy', 'Drama', 'Romance']	['Drama', 'Comedy', 'Romance']
12 angry men	['Crime', 'Drama']	['Drama', 'Crime']
parasite	['Comedy', 'Drama', 'Thriller']	['Joon-hoBong']
casablanca	['Drama', 'Romance', 'War']	['Drama', 'Romance', 'War']
psycho	['Horror', 'Mystery', 'Thriller']	['Mystery', 'Thriller', 'Horror']
the godfather	['Crime', 'Drama']	['Drama', 'Thriller', 'Crime']

SIMILARITY SCORE OF GENRES OF COMMON MOVIES BETWEEN IMDB AND METACRITICS IS 0.7778