Movie Script Analysis

Presented By:
Faizan Mughl 20l-0939
Iqra 21l-6212

Contents

Research Article

The emotional arcs of stories are dominated by six basic shapes

Presented in 2016 by John and peter. Filtered dataset of around 1300 fictional books stories were used. Methodology used was mathematical and computational techniques like SVD

Problem Statement

Movie Script Analysis: Unveiling the Six Universal Plots

The project seeks to offer a compelling argument backed by systematic analysis and critical evaluation!

And the most important thing: investigate and substantiate the claim that all stories,can be distilled into a core set of six archetypal plots and can be used to build a movie recommender afterwards.

Description of Data set

Data Collection

The data has been scraped using selenium from IMDb and IMSDb website and is a data of top 3 movies from 1998-2022

Data columns

Columns that has been used are of movie script,genre,running time,movie year,movie title.

Columns used

Movie name ,year, movie script

Preprocessing Techniques

1
Converting text to lower case (script col)

2
Removing line breaks. Preventing disruption in analysis

3
Removing punctuation marks.

4
Removing stopwords and numbers.

5
TF IDF : cosine similarity.

Models used and Why?

K-means

- Identification of Common Emotional Arcs
- Efficient and Interpretable Grouping

BERT

- Capturing Contextual Relationships
- Understanding Nuanced Emotional Content

Which one was better?

K-means

- K means had silhouette score of 0.7.
- It produced clusters that are relatively better separated
- Better performance
- Required few seconds for execution.

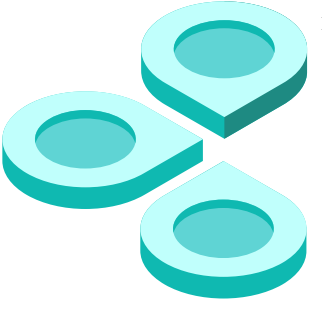
BERT

- It had silhouette score of 0.04.
- BERT might capture more nuanced emotional context, while K-means might offer clearer delineation of distinct clusters.
- Required high computational power.
- Required 2 hours for execution.

Project uniqueness

Deep learning model
BERT model is used.

Research Based
This proves the research by an efficient approach that has been implemented.



Future usage
This could be presented as a basis for further sentimental analysis.

Thanks!

Does anyone have any questions?