IMDB Movie Recommender Based on Storylines

AISHWARYA.S.M

APPROACH

This project extracts movie data from IMDb (2024) using Selenium, focusing on movie names and storylines. The collected storylines are pre-processed and analyzed using Natural Language Processing (NLP) techniques such as **TF-IDF** and **Count Vectorizer**. To identify similarity, **Cosine Similarity** is applied, enabling the system to recommend movies with storylines most similar to a given input. An **interactive Streamlit interface** allows users to enter a storyline and receive the **top 5 recommended movies**, complete with similarity scores and summaries.

DATA PREPROCESSING

Data Collection:

Scraped IMDb (2024) movie titles and storylines using Selenium; saved the data to `genre_movies.csv`.

Preprocessing:
Cleaned text using regex (removed non-alphanumeric characters), converted to lowercase, and ensured all text was in string format.

NLP Techniques:
Applied tokenization and lemmatization, then converted tokens back to string. Used TF-IDF and Count Vectorizer to represent storylines numerically.

Similarity Calculation:
Used Cosine Similarity to measure the closeness between movie storylines.

Recommendation System:
Suggested top 5 similar movies based on the input storyline.

User Interface:
Developed an interactive Streamlit app for real-time movie search and recommendations.

Top 5 similar movies based on the input storyline.

Anora — *Score: 1.000*

F* Marry Kill** — *Score: 0.116*

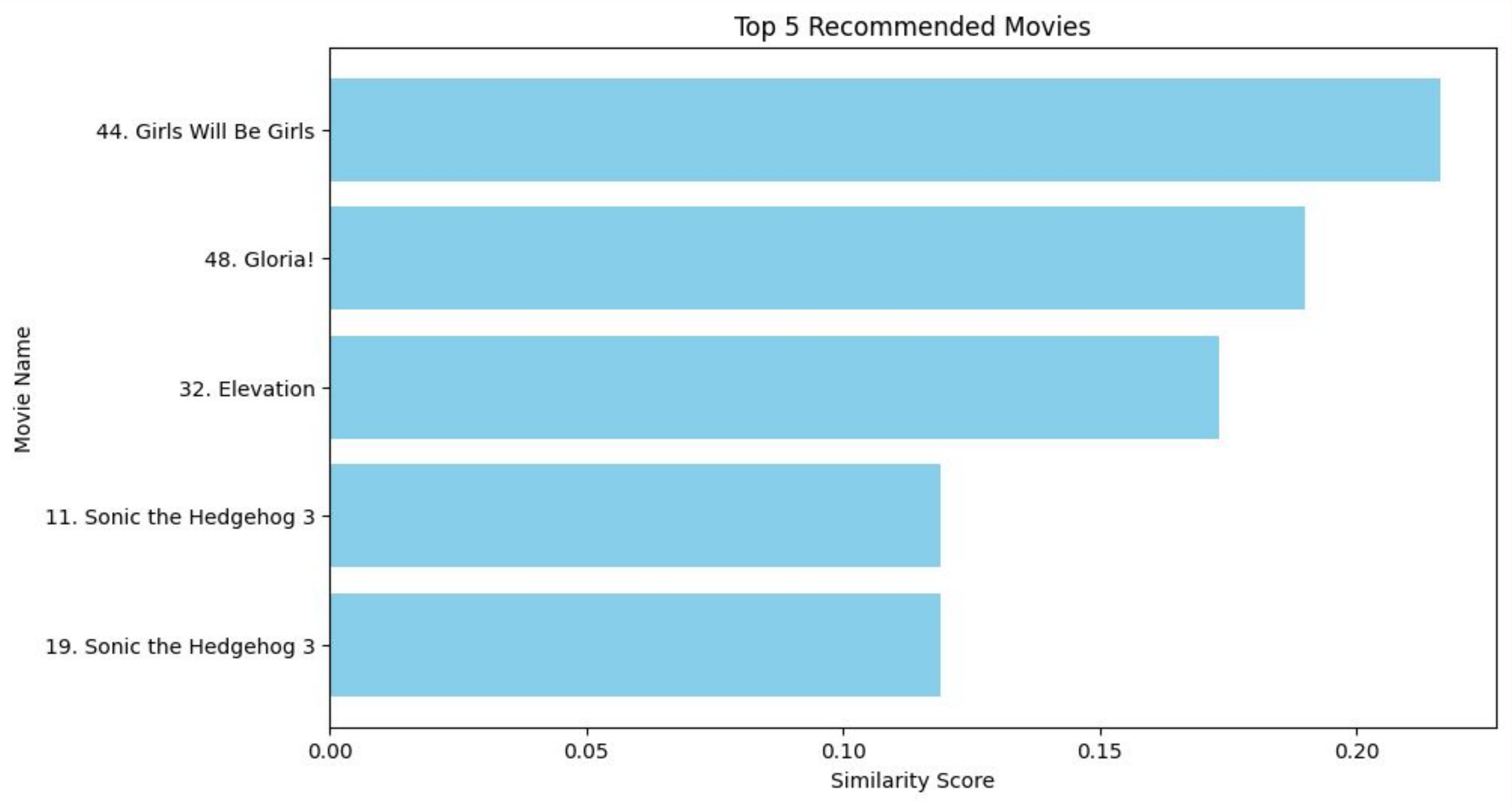
Spermageddon — *Score: 0.109*

Spermageddon — *Score: 0.109*

Doraemon the Movie: Nobita's Earth Symphony — *Score: 0.106*

Top 5 Recommended Movies (Based on Storyline Similarity)

- This bar chart visualizes the **top 5 movie recommendations** based on storyline similarity.
- **X-axis:** Similarity Score (using Cosine Similarity)
- **Y-axis:** Movie Names with their respective indices
- Higher bars indicate **closer storyline matches** to the input movie.
- Duplicate titles (e.g. *Sonic the Hedgehog 3*) may appear due to **duplicate entries** or **similar storyline matches**.


Top 5 Recommended Movies
Movie Name
44. Girls Will Be Girls
48. Gloria!
32. Elevation
11. Sonic the Hedgehog 3
19. Sonic the Hedgehog 3
0.00
0.05
0.10
0.15
0.20
Similarity Score