IS 733 DATA MINING

GROUP 11

CASELET 01: MOVIE REVIEW

Team Members and Roles:

Asmita Deshpande (OF68333)
Gayathri Gurram (FR06850)
Ravi Sharma (PK93698)
Tarunsingh Jodha (HG53853)

Roles	Name	Responsibilities
Facilitator	Asmita Deshpande	Ensured that the group is working effectively together and progressing toward their goals. Furthermore, everyone was given the opportunity to speak and contribute.
Spokesperson	Gayathri Gurram	Encouraged effective communication both inside the group and with the rest of the class.
Quality Control	Ravi Sharma	Ensured that the group's work is of good quality and meets the project requirements.
Process Analyst	Tarunsingh Jodha	Ensured that the group analyzed, designed, and optimised the process successfully, while doing the project.

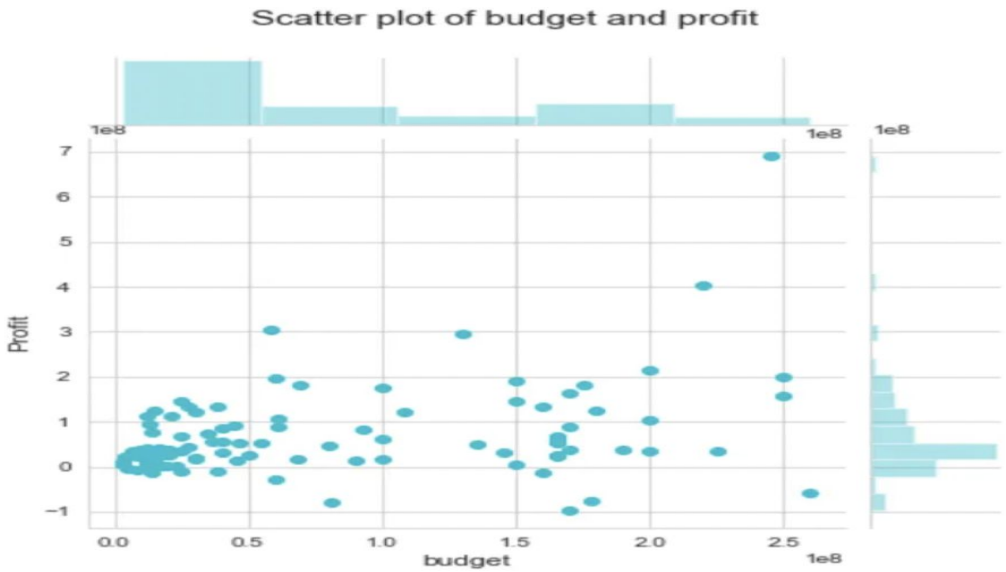
Introduction

- The movie industry is a multibillion-dollar business that heavily relies on audience feedback and reviews to make critical decisions.
- With the advent of social media and online platforms, analyzing large amounts of movie review data has become easier than ever before.
- In this presentation, we put together our observations and results after exploring a dataset of movie reviews to extract insights and understand the sentiment of the audience towards different movies.

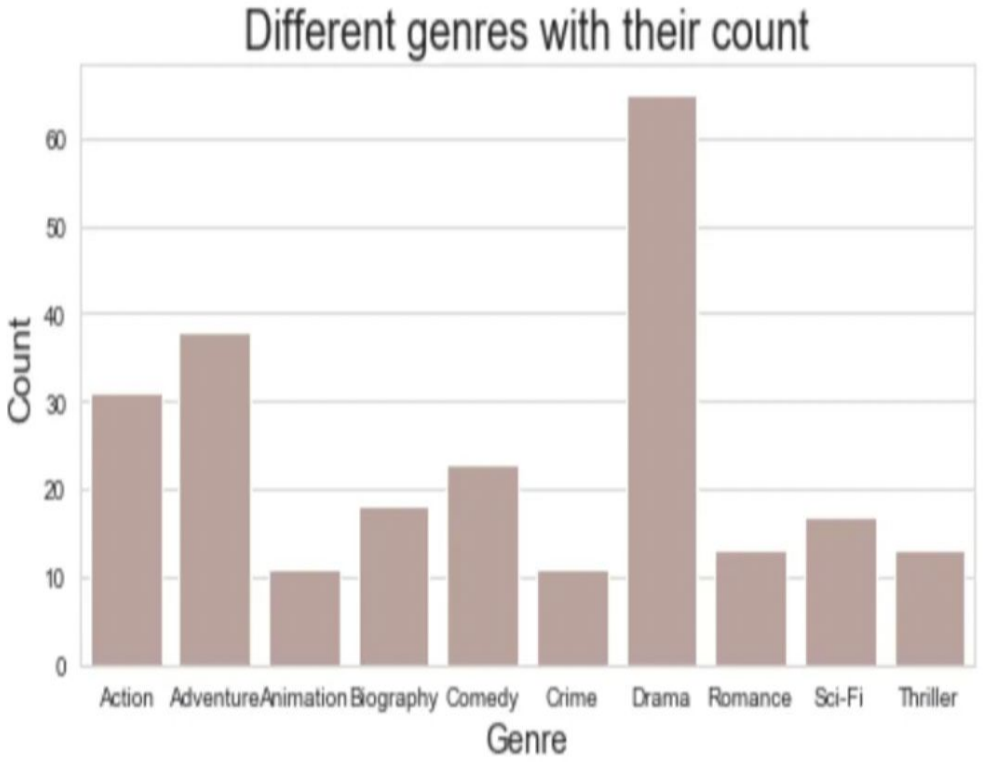
Background / Motivation:

- Rise of online streaming services has led to overwhelming content choices.
- Movie recommendation systems analyze viewing history and provide personalized suggestions.
- Motivation: Improve user experience, increase satisfaction, and drive customer loyalty.
- TMDb dataset: Valuable resource with comprehensive movie information for accurate recommendations.
- Key Takeaways: Recommendation systems crucial for user satisfaction and competitiveness. TMDb dataset enables personalized suggestions.
- Effective movie recommendation systems can drive user engagement and retention in the highly competitive streaming industry.
- By leveraging the TMDb dataset and advanced recommendation algorithms, streaming services can deliver tailored movie suggestions that align with users' preferences, leading to a more personalized and enjoyable viewing experience.

EDA



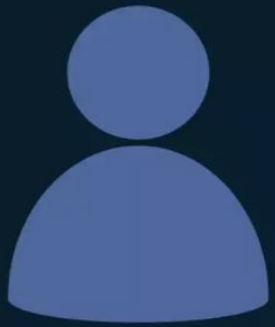
- Slightly **skewed** to right
- A couple of movies that have very high budgets but still having negative profits
- Slightly positive trend between ‘budget’ and ‘profit’



The data set of the top 100 movies has **Drama** genre the most.

CONTENT-BASED FILTERING EXAMPLE

Watched by user



Similar movies based on the “genre,” “cast,” and “director” features

Crime Movies

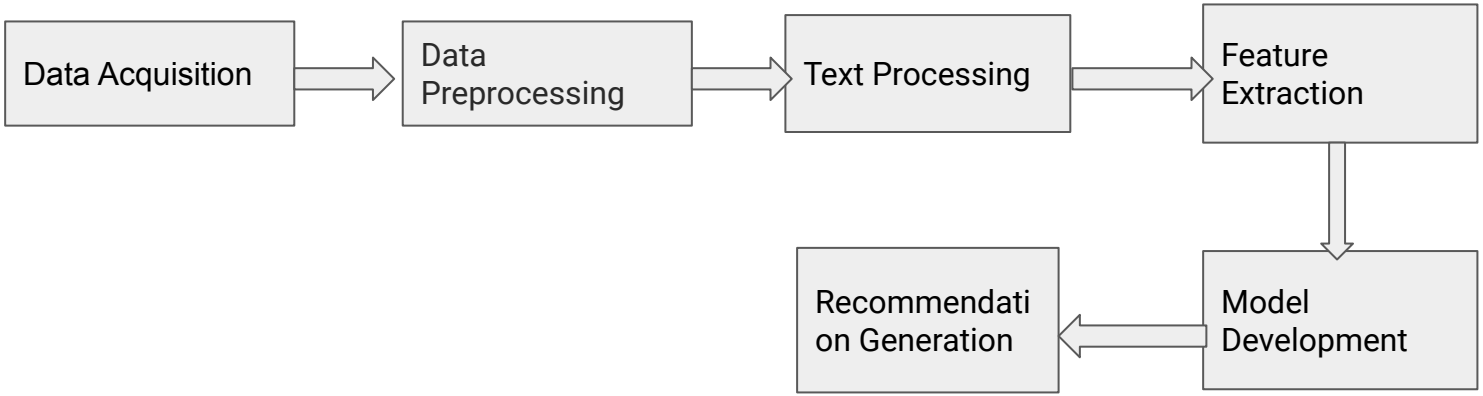
Al Pacino Movies

Robert De Niro Movies

Joe Pesci Movies

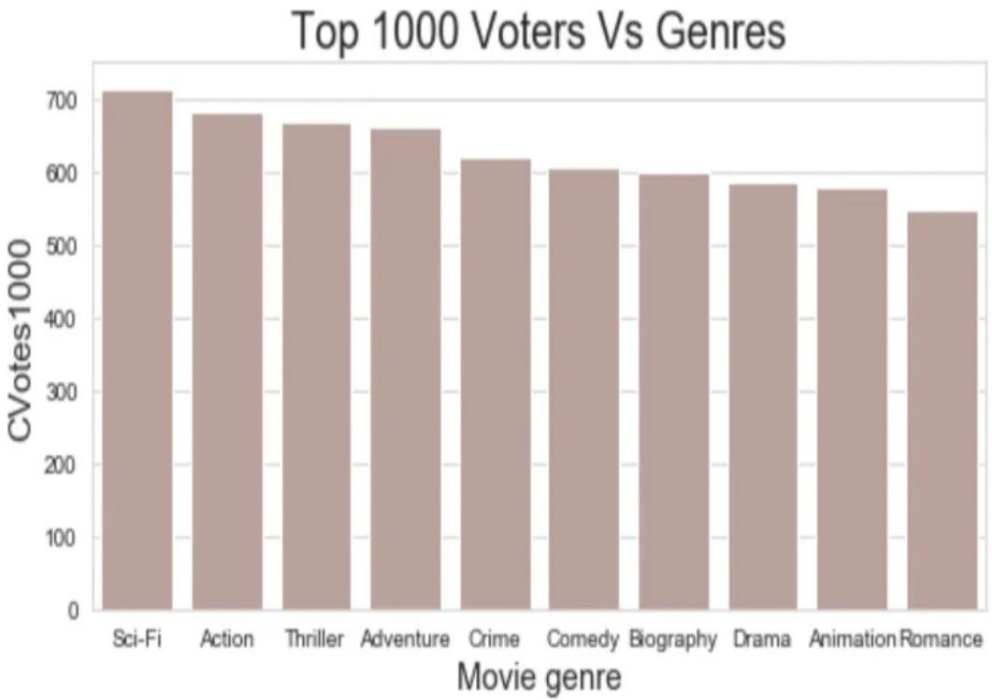
Martin Scorsese Movies

Recommended to user



Model Development Pipeline

Results and Insights



- Sci-Fi is the most popular amongst the top 1000 voters
- The romance genre has been voted the least from the top 1000 voters
- It is observed that the top 1000 IMDb voters prefer Sci-Fi over Drama and also the dataset contains movies from Drama more compared to other genres

```
[ ] recommend('John Carter')
```

```
Star Trek: Insurrection
Mission to Mars
Captain America: The First Avenger
Escape from Planet Earth
Ghosts of Mars
```

```
recommend('Spider')
```

```
Hit & Run
Donnie Brasco
From Paris with Love
Fled
The Perfect Match
```

- It’s a metadata based recommender system
- Successful in capturing more information due to more metadata provided and result in better recommendations

Conclusion/ Future Works:

- TMDb dataset + personalized suggestions = enhanced user experience.
- Benefits: Increased engagement, satisfaction, content exploration.
- Techniques Explored: Collaborative, content-based, and hybrid approaches for accurate recommendations
- Stay competitive: Accurate recommendations attract and retain users, drive revenue
- Improve accuracy: Explore matrix factorization, deep learning
- Enhance performance: Incorporate user feedback
- Leverage advanced algorithms, rich datasets for better recommendations.

Thank you.