STATISTICS CASE STUDY AND REPORT

Presented By:-

Group-8

Devarshi Dubey

Sameed Yallur

Karan Naik

Ashish S N

Agenda

- Data Set
- Tools
- Libraries Used
- UNIT - 2
- UNIT - 3
- UNIT - 4
- UNIT - 5
- Conclusion

Data Set

NETFLIX MOVIE RECOMMENDATION SYSTEM

The Netflix movie recommendation system dataset is a treasure trove of insights into the intricate mechanics behind one of the most sophisticated recommendation engines in the entertainment industry.

Comprising a vast collection of anonymized user interactions, it delves into user preferences, ratings, and viewing habits across a diverse array of movies. This dataset embodies the essence of collaborative filtering and machine learning algorithms, showcasing how Netflix harnesses cutting-edge technology to curate personalized recommendations that captivate audiences worldwide.

Its granular details encapsulate the complex interplay between content and user behavior, shedding light on the art and science behind delivering tailored cinematic experiences, making it an invaluable resource for understanding the dynamics of modern-day content recommendation systems.

https://academictorrents.com/details/9b13183dc4d60676b773c9e2cd6de5e5542cee9a

Dataset format

CSV file of the form

Movie id:

userid, rating, date

1:

user19921, 4, 1998-04-05

user 19345, 3, 2001-01-01

2:

user13941, 4, 1998-04-05

user 14362, 3, 2001-01-01

Tools

Libraries Used

CSV

Dataset stored in CSV format, provides accessible and structured insights into user preferences and movie interactions for comprehensive analysis.

math

Revolutionizes data manipulation and analysis with its powerful Data-Frame structure, facilitating effortless handling, cleaning, and exploration of diverse datasets.

pandas

Offers a comprehensive suite of mathematical functions and tools in Python, enabling efficient numerical computations and data manipulation for diverse scientific applications.

numpy

offers efficient array operations and mathematical functions, enabling high-performance numerical computations and data manipulation.

UNIT – II

- Large sample test - 1 sample
- Large sample test - 2 sample
- small sample test - 1 sample
- small sample test - 2 sample
- Paired Sample test
- proportion test - 1 sample
- proportion test - 2 sample

UNIT – III

- The Runs Test
- Sign Test-One Sample
- Paired Sample- Sign Test
- The Wilcoxon Rank Sum Test or Mann-Whitney U test
- The Wilcoxon Signed-rank Test

UNIT – IV

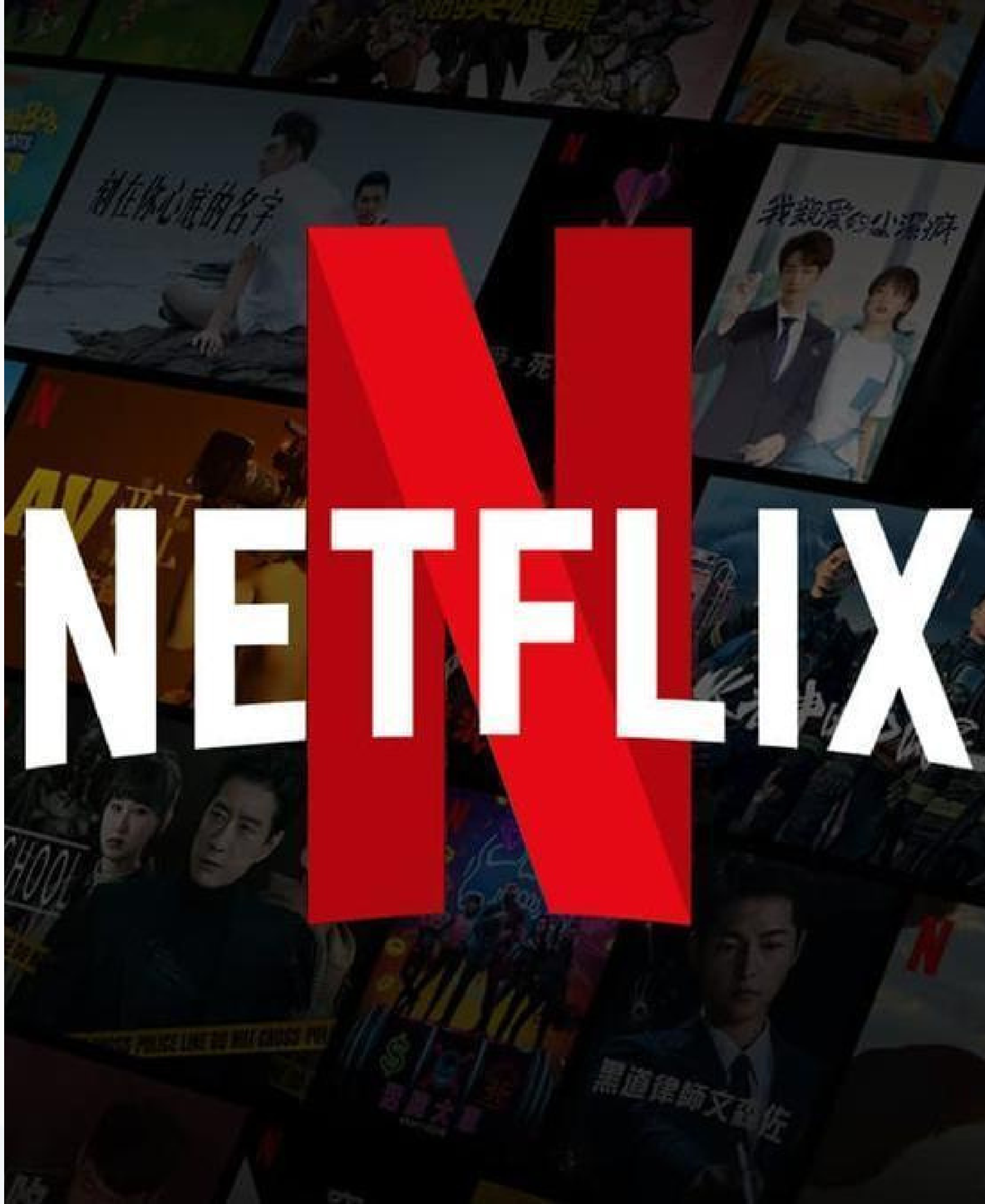
- Chi-Square Test
- Ch-Square Test for Independence
- Test for Homogenity
- 2 X 2 contingency
- 2 X 2 contingency - Yate’s correction

UNIT – V

- sample size estimation - 1 sample
- sample size estimation - 2 sample
- proportion test - 1 sample size estimation
- proportion test - 2 sample size estimation

Conclusion

The Netflix movie recommendation system dataset stands as a testament to the fusion of technology and entertainment, offering a glimpse into the intricate workings of a cutting-edge recommendation engine. Its structured insights into user behavior and movie interactions not only fuel the algorithms behind personalized suggestions but also serve as a goldmine for researchers exploring the depths of collaborative filtering and machine learning. As datasets like these continue to evolve, they underscore the power of data in reshaping how we engage with content, paving the way for more personalized, immersive entertainment experiences while illuminating the incredible potential of data-driven innovation in the entertainment industry.



Thank You