INTELLIGENT MOVIE RECOMMENDATIONS WITH NEURAL NETWORKS AND NLP

Sub-title: Enhancing Personalized Movie Suggestions

PRESENTATION BY :DSF-PT-07:GROUP 12

Problem Statement

The Challenge:

- Overwhelming number of movies leads to choice paralysis.
- Traditional recommendation systems fall short in personalization.

Project Goal:

- Develop a recommendation system that delivers highly personalized movie suggestions using advanced machine learning techniques.

Overview

Objective:

- Leverage neural networks for collaborative filtering and Natural Language Processing (NLP) for content analysis.

- Address user dissatisfaction with generic movie recommendations by providing more tailored suggestions.

Value to Stakeholders:

Product: Flick Pickle Engine

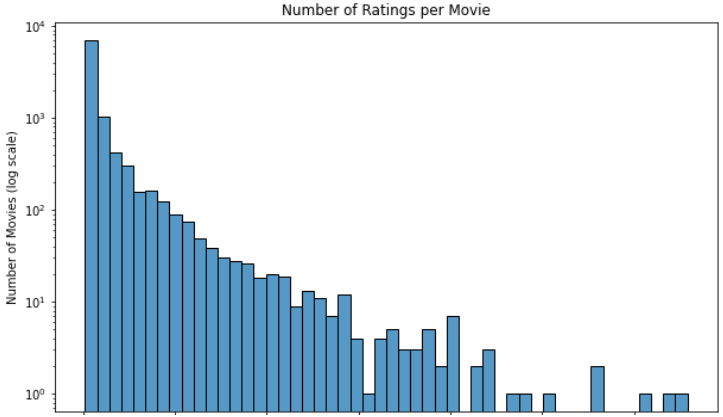
- Improved user experience and engagement through personalized recommendations.

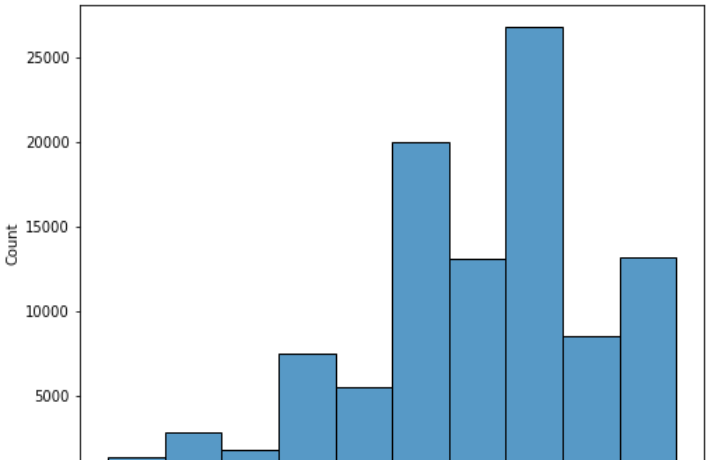
Data Understanding

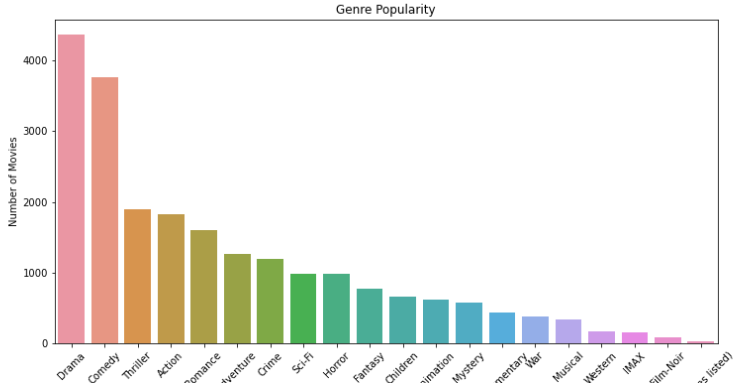
- Dataset Used: MovieLens 20M Dataset
  - 25 million ratings, 20 million user-generated tags.
  - Rich movie metadata (title, genre, tags).

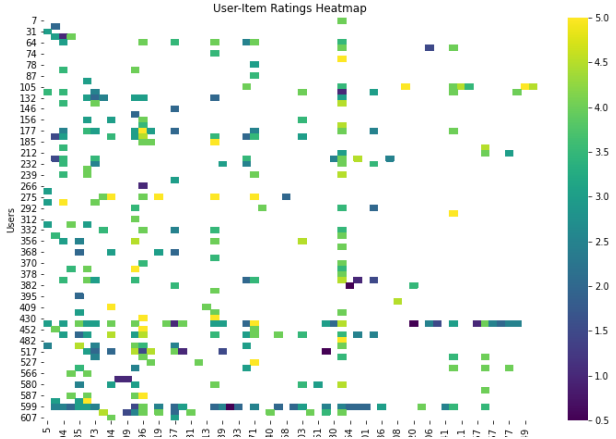
Key Data Components

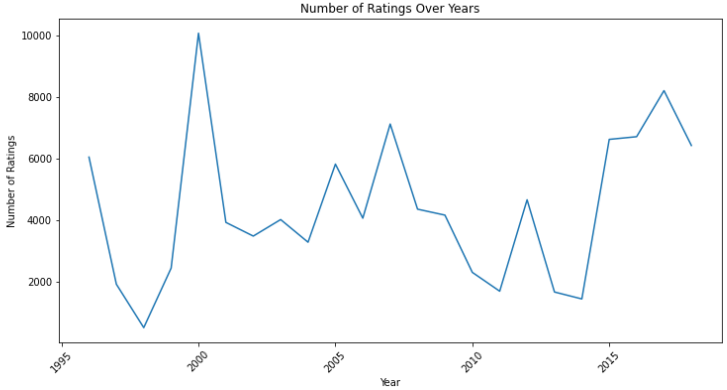
- Ratings: User ratings on movies (1 to 5 scale).
- Movies: Metadata such as title, genre, and tags.
- Users: User profiles based on past ratings.

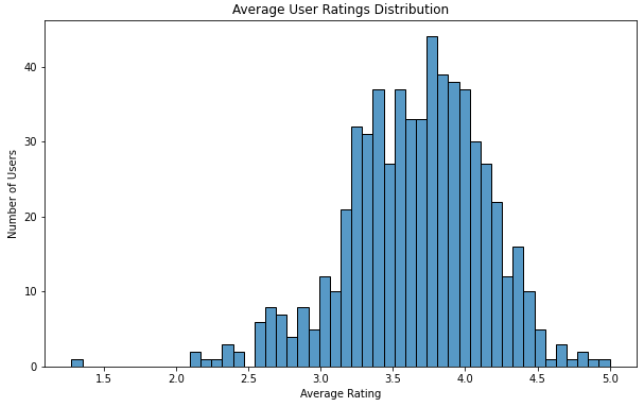

Number of Ratings per Movie
Number of Movies (log scale)
10^4
10^3
10^2
10^1
10^0

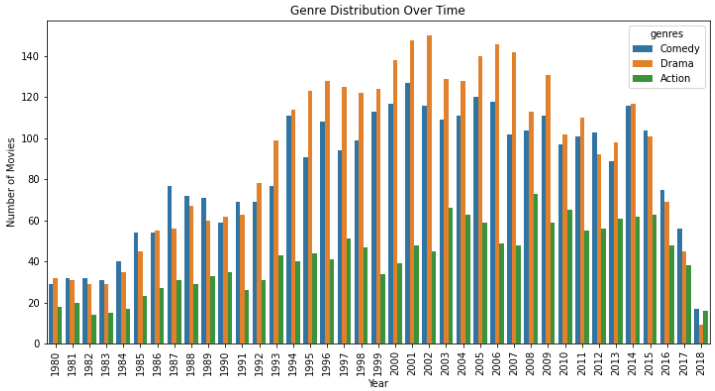

Count
25000
20000
15000
10000
5000

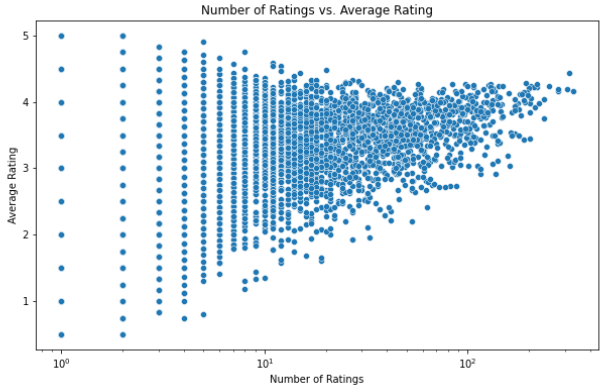

Genre Popularity
Number of Movies
4000
3000
2000
1000
0
Drama
Comedy
Thriller
Action
Romance
Adventure
Crime
Sci-Fi
Horror
Fantasy
Children
Animation
Mystery
Documentary
War
Musical
Western
IMAX
Film-Noir


User-Item Ratings Heatmap
Users
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5


Number of Ratings Over Years
Number of Ratings
Year
10000
8000
6000
4000
2000
1995
2000
2005
2010
2015


Average User Ratings Distribution
Number of Users
Average Rating
40
30
20
10
0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0


Genre Distribution Over Time
Number of Movies
Year
genres
Comedy
Drama
Action


Number of Ratings vs. Average Rating
Average Rating
Number of Ratings
5
4
3
2
1
10^0
10^1
10^2

WORD CLOUD OF MOVIES

Data Preparation

- Data Cleaning: Removing duplicates and handling missing values.
- Feature Engineering:
  - Clustering user-generated tags using NLP to reduce complexity.
  - Transforming user and movie data into matrices for collaborative filtering.
- Preprocessing:
  - Genres and tags transformed into numerical features for modeling.

Modeling

- Collaborative Filtering:
  - User-based filtering: Recommendations based on similar user preferences.
  - Item-based filtering: Recommendations based on similar movies.
- Hybrid Model:
  - Incorporates both collaborative and content-based features.
  - Tackles the cold-start problem for new users or movies.
- NLP for Tag Clustering:
  - Clusters related tags to simplify content-based filtering.

Evaluation

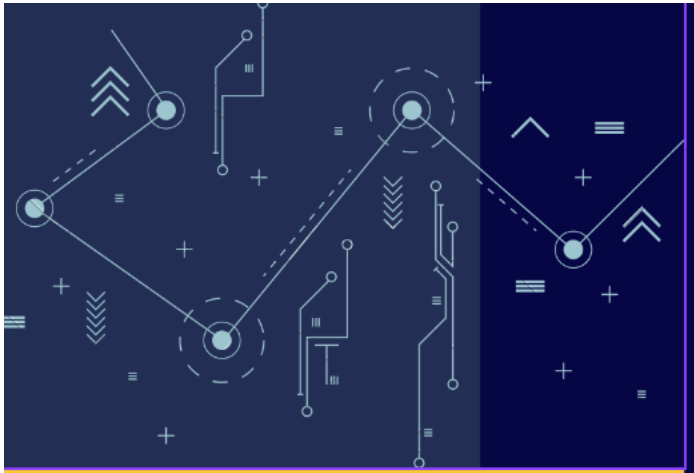
- **Evaluation Metrics:**
  - RMSE (Root Mean Square Error): Measures accuracy in predicting user ratings.
  - MAE (Mean Absolute Error): Provides another accuracy measure.
- **Additional Considerations**
  - Future inclusion of NDCG and F1-score for ranking quality.

Results & Value

- **Key Results**
  - Successfully personalized recommendations.
  - Hybrid model improved performance by integrating user and content-based data.
- **Value to Stakeholders**
  - Users spend less time searching and more time enjoying relevant content.
  - Higher engagement and satisfaction with the platform.

Future Work

- Implement reinforcement learning for adaptive recommendations.
- Introduce user controls for more personalized movie suggestions.
- Explore A/B testing to validate different recommendation strategies.



THANK YOU

Q&A

DSFPT07 Phase 4 Group 12

Branely Ope
Brian Kipngetich
Cynthia Atieno
Geoffrey Mwangi
Linet Maz'susa
Maureen Wanjeri
Mercy Silali