

RIPE PUMPKINS

PUMPKINMETER- MOVIE REVIEW AGGREGATION SERVICE

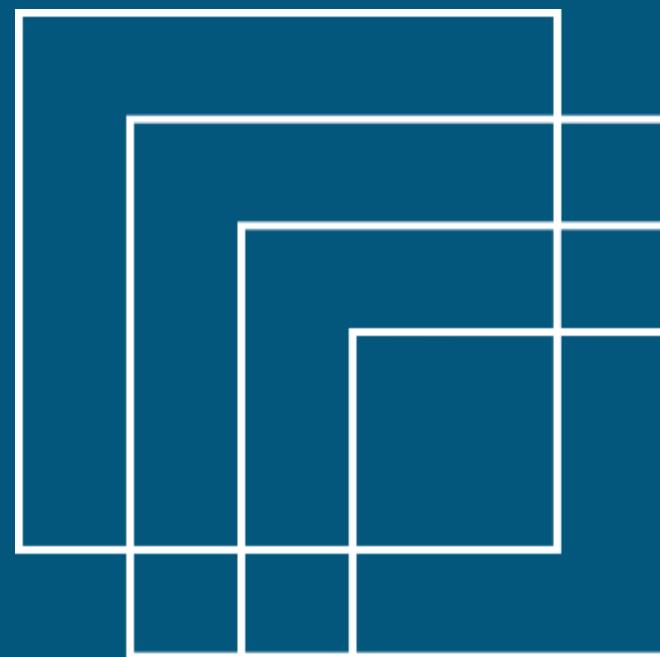
BY- RABIYA FATIMA
COURSE - BUAN 5315 O2 23SQ BIG DATA ANALYSIS



RIPE PUMPKINS
COMPANY

CONTENT

- 01 ABOUT US
- 02 PROBLEM & OBJECTIVE
- 03 DATA OVERVIEW
- 04 COLLABORATIVE FILTERING, SPARK & ALS
- 05 PRE PROCESSING & MODEL BUILDING
- 06 OVERVIEW OF TEST SETUP
- 07 RECOMMENDATION RESULTS & INSIGHTS



ABOUT US

- Ripe Pumpkins is an innovative startup in the movie industry. We have developed a movie review-aggregation service called Ripe Pumpkins.
- Our new initiative, Pumpkinmeter, is a collaborative recommendation model for movie enthusiasts. Pumpkinmeter aims to provide personalized movie recommendations based on user preferences.
- It leverages the success of recommendation models in streaming services to enhance user experience.

Today, we will explore the potential of Pumpkinmeter and present data-driven insights for the board's consideration.



BUSINESS PROBLEM & OBJECTIVE



PROBLEM

Ripe Pumpkins is addressing the business problem of providing personalized recommendations to users in the streaming service industry, aiming to enhance user satisfaction and engagement.



IMPORTANCE

Crucial for the success of streaming services as they enable accurate predictions of user preferences based on the collective behavior and preferences of a large user base, leading to improved content discovery and increased user retention.



OBJECTIVE

To evaluate and quantify the effectiveness of the recommendation system by measuring how well it aligns with user preferences, thereby enabling continuous improvement and optimization of the recommendation algorithms.

DATASET OVERVIEW

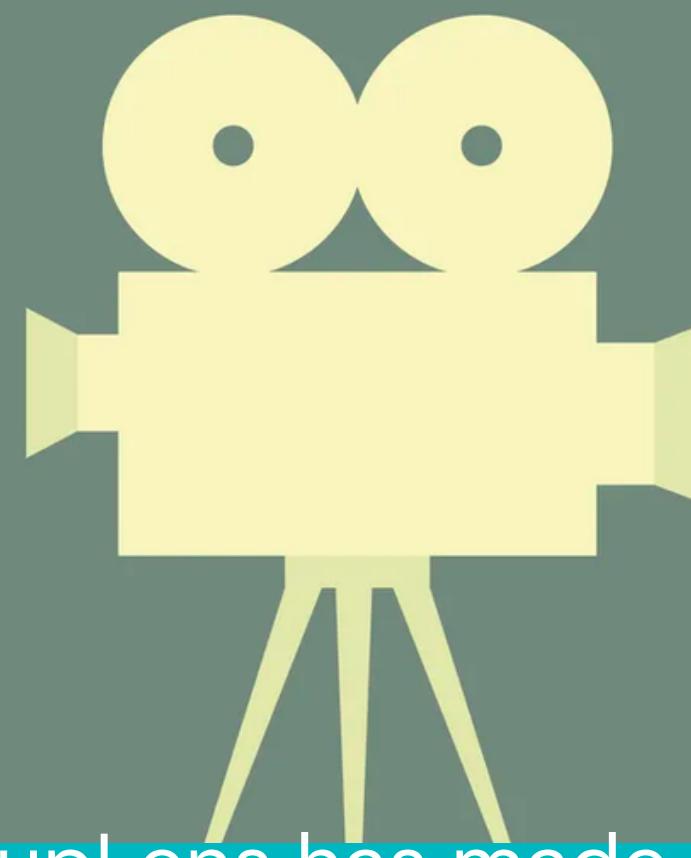
Small Dataset:

- Data size: 100,000 ratings and 3,600 tag applications applied to 9,000 movies by 600 users.
- Last updated: September 2018.

Full Dataset:

- Data size: 27,000,000 ratings and 1,100,000 tag applications applied to 58,000 movies by 280,000 users.
- Includes tag genome data with 14 million relevance scores across 1,100 tags.
- Last updated: September 2018.

Dataset is collected by GroupLens Research from the MovieLens website.



GroupLens has made available rating datasets from MovieLens, and we have used the latest datasets.

COLLABORATIVE FILTERING WITH SPARK MLLIB AND ALS

Collaborative Filtering:

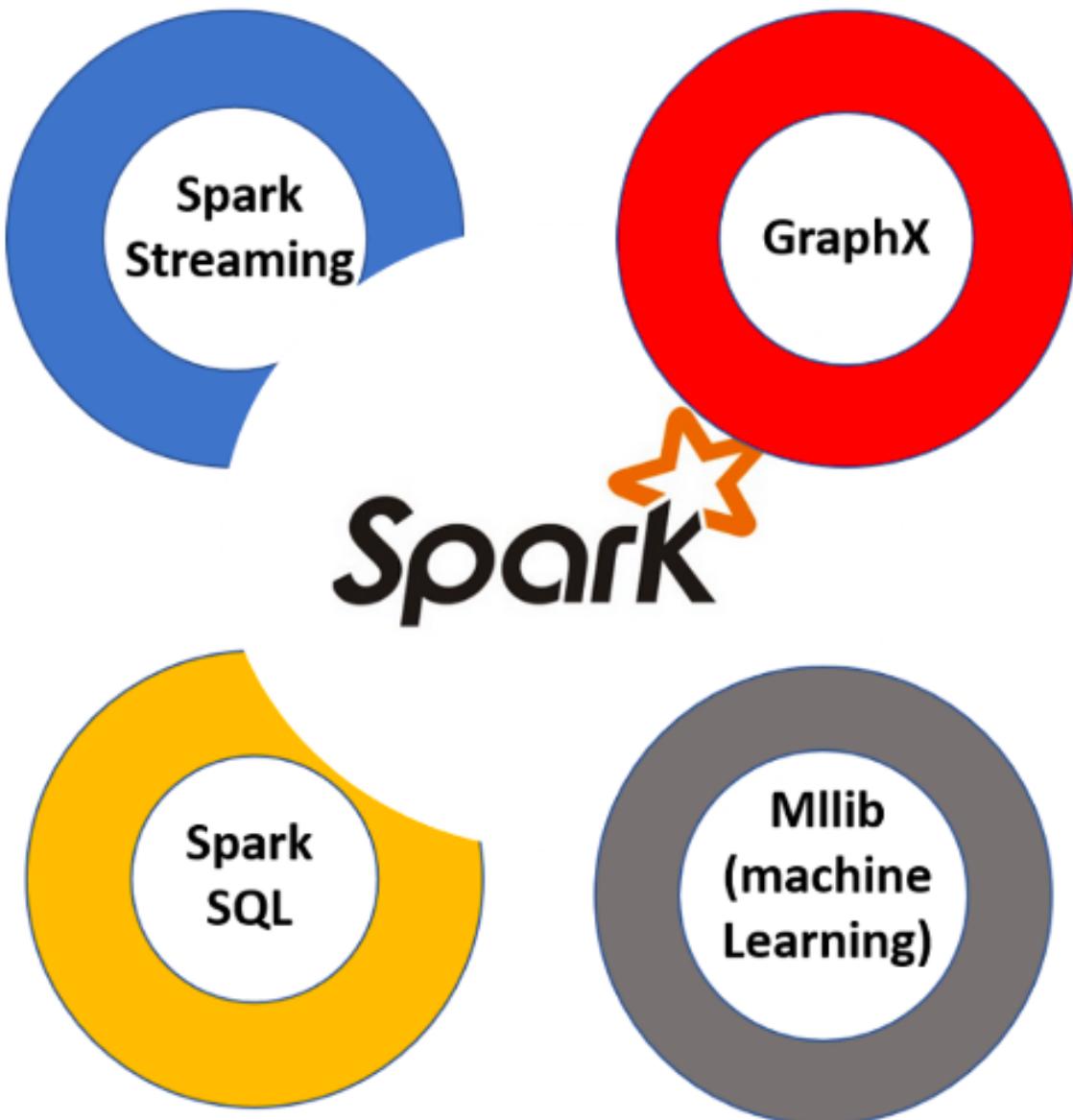
- Technique used in recommender systems to predict user preferences by leveraging the preferences and behaviors of similar users.
- Assumes that users who agreed on preferences in the past are likely to agree in the future.

Spark MLlib:

- Machine learning library provided by Apache Spark for scalable and distributed machine learning.
- Simplifies development of collaborative filtering and other machine learning algorithms.

ALS (Alternating Least Squares):

- Algorithm provided by Spark MLlib for collaborative filtering.
- Decomposes user-item interaction matrix into lower-rank user and item matrices.
- Minimizes prediction error through iterative optimization of matrices.



PREPROCESSING AND MODEL BUILDING

Preprocessing crucial to optimize the performance and accuracy of the recommender model. It involves tasks such as data cleaning, handling missing values, and transforming data into the required format for model training.

Precomputed Tasks:

- Dataset Loading: Load the dataset into Spark RDD (Resilient Distributed Dataset), which allows for distributed processing.
- Dataset Parsing: Parse the dataset to extract relevant information, such as user-item interactions and ratings.

Persisting RDDs and Recommender Model:

- Persisting RDDs: RDDs can be persisted in memory or disk to avoid recomputation and improve efficiency during iterative processes.



Overview of the test setup:

Two new users 1& 2 were selected for the test, each user provided ratings for 10 movies.

Two scenarios were considered for generating recommendations:

Scenario 1 (Full dataset)

filtered out movies with less than 25 ratings

Scenario 2 (Full dataset)

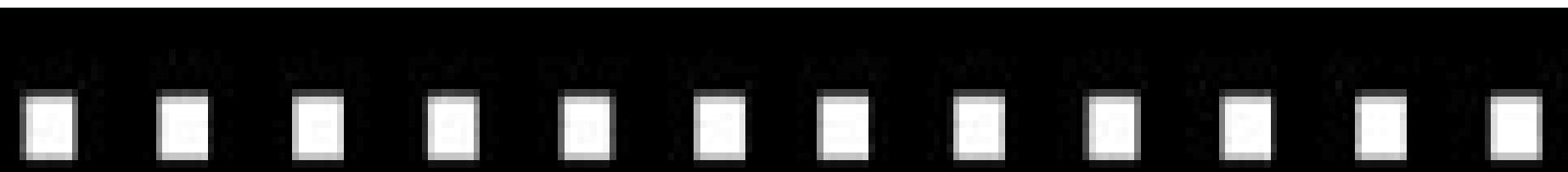
filtered out movies with less than 100 ratings.

RECOMMENDATION RESULTS OF USER 1

Scenario 1

TOP recommended movies (with more than 25 reviews):

- ('Planet Earth II (2016)', 4.78239333803813, 853)
- ('Planet Earth (2006)', 4.782130643517068, 1384)
- ('Connections (1978)', 4.683413122489011, 49)
- ('Cosmos', 4.6805558922740715, 157)
- ('Life (2009)', 4.669316616143593, 166)
- ('Blue Planet II (2017)', 4.666810415400615, 349)
- ('Band of Brothers (2001)', 4.656276806142959, 984)
- ('Music for One Apartment and Six Drummers (2001)', 4.651242438966417, 31)
- ("Hollow Crown", 4.646724260064673, 36)
- ('The Godfather Trilogy: 1972-1990 (1992)', 4.638730790928868, 421)



The recommendations include popular nature documentaries like "Planet Earth II," "Planet Earth," and "Blue Planet II." These documentaries are highly rated by other users and have received a significant number of reviews, indicating their widespread appeal.

Insights: User 1 seems to have a preference for nature and documentary genres. The top recommendations align with their interests, suggesting that the model has successfully captured their taste. The model identifies patterns in the ratings and leverages the collective preferences of other users to make accurate recommendations. User 1 may find these recommendations valuable and enjoyable, considering their interest in nature-related content.

Foresight: Based on User 1's current preferences, it is likely that they will continue exploring and enjoying documentaries, particularly those related to nature and the environment. The model can further refine its recommendations by suggesting similar documentaries or expanding the user's viewing horizon by suggesting related genres or subjects.



RECOMMENDATION RESULTS OF USER 1

Scenario 2

TOP recommended movies (with more than 100 reviews):

- ('Planet Earth II (2016)', 4.782383753879586, 853)
- ('Planet Earth (2006)', 4.782124899602405, 1384)
- ('Cosmos', 4.68055178739926, 157)
- ('Life (2009)', 4.669304881294838, 166)
- ('Blue Planet II (2017)', 4.666779538587825, 349)
- ('Band of Brothers (2001)', 4.6562688722225545, 984)
- ('The Godfather Trilogy: 1972-1990 (1992)', 4.638728814312792, 421)
- ('Death on the Staircase (Soupçons) (2004)', 4.63538619149481, 130)
- ('Alone in the Wilderness (2004)', 4.626788065592677, 343)
- ('"Godfather', 4.618566081451554, 60904)



Like Scenario 1, the top recommendations include popular nature documentaries such as "Planet Earth II," "Planet Earth," and "Blue Planet II." These documentaries consistently received more than 100 ratings from users.

Insights: User 1's movie preferences remain consistent with their interest in nature and documentaries, as observed in Scenario 1. The model's recommendations are aligned with their established preferences, indicating that the model has effectively captured User 1's taste even with the increased number of ratings. The larger number of ratings allows the model to gain a deeper understanding of User 1's preferences and deliver personalized recommendations.

Foresight: With User 1's continued engagement and extensive rating history, it is expected that the model's recommendations will become increasingly accurate and personalized over time. As User 1 explores more movies and provides ratings, the model will continue to refine its understanding of their preferences.

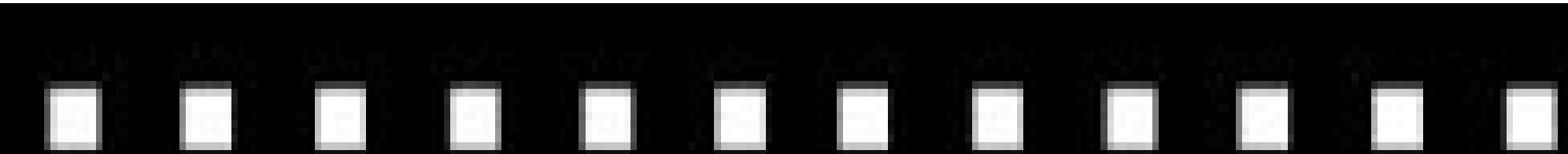


RECOMMENDATION RESULTS OF USER 2

Scenario 1

TOP recommended movies (with more than 25 reviews):

- ('Connections (1978)', 4.688590221985768, 49)
- ('Runaway Brain (1995)', 4.436740196681846, 30)
- ('Winter in Prostokvashino (1984)', 4.4353161979966, 67)
- ("Hollow Crown", 4.432863124313808, 36)
- ('Cosmos: A Spacetime Odissey', 4.428878044907876, 37)
- ('Vacations in Prostokvashino (1980)', 4.418828466434498, 96)
- ("Dylan Moran: Like", 4.409149770419394, 76)
- ('Between the Folds (2008)', 4.395310608962136, 61)
- ("Jim Henson's The Storyteller (1989)", 4.392088347302028, 36)
- ('Drishyam (2013)', 4.356462406102509, 37)



User 2's top recommendations include movies like "Connections," "Runaway Brain," and "Winter in Prostokvashino," which are lesser-known titles in comparison to the nature documentaries recommended to User 1.

Insights: User 2's movie preferences seem to be more eclectic, as indicated by the varied recommendations. The model recognizes their interest in less mainstream movies and suggests titles that align with their preferences. The recommendations include niche movies and lesser-known gems that User 2 may find appealing based on their extensive rating history.

Foresight: User 2's preference for lesser-known movies suggests that they enjoy exploring unique and unconventional content. As the model continues to learn from their ratings, it can provide more tailored recommendations within this niche. The model may suggest movies from similar genres or introduce User 2 to movies with a similar style or artistic approach. This can further enhance their movie-watching experience and lead them to discover hidden cinematic treasures.

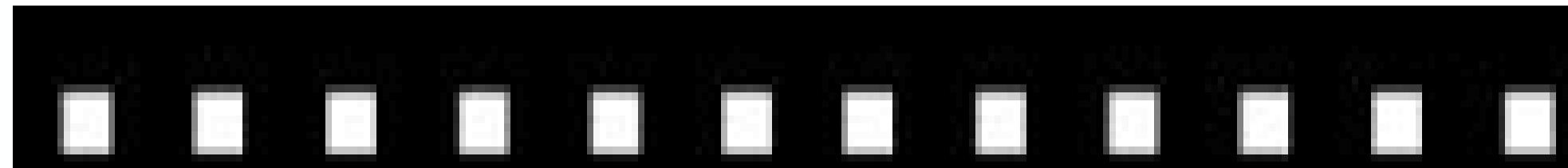


RECOMMENDATION RESULTS OF USER 2

Scenario 2

TOP recommended movies (with more than 100 reviews)

- ('Planet Earth (2006)', 4.305927347385432, 1384)
- ("Can't Change the Meeting Place (1979)", 4.2825077050239475, 119)
- ("Shawshank Redemption', 4.274695412820099, 97999)
- ('Planet Earth II (2016)', 4.261703820317978, 853)
- ("Operation 'Y' & Other Shurik's Adventures (1965)", 4.250130051658622, 416)
- ('The Adventures of Sherlock Holmes and Doctor Watson: King of Blackmailers (1980)', 4.249744305376558, 115)
- ('The Adventures of Sherlock Holmes and Dr. Watson: The Hound of the Baskervilles (1981)', 4.248151685555941, 193)
- ('Sherlock Holmes and Dr. Watson: Acquaintance (1979)', 4.245380796026798, 136)
- ('The Adventures of Sherlock Holmes and Dr. Watson: Bloody Signature (1979)', 4.244919757628342, 141)
- ('There Once Was a Dog (1982)', 4.236640882245879, 256)



The recommendations provided reflect a diverse range of genres and styles, catering to User 2's varied taste. Some of the recommended movies include "Shawshank Redemption," "Can't Change the Meeting Place (1979)," and "There Once Was a Dog (1982)"

Insights: User 2's movie preferences exhibit a wide-ranging interest in different genres and styles, as indicated by the diverse recommendations. The model recognizes this and suggests popular and critically acclaimed movies from various categories. User 2's rating history reflects an openness to exploring different cinematic experiences, and the model successfully captures their preferences.

Foresight: With User 2's inclination towards diverse movie genres, the model can further refine its recommendations by suggesting movies that align with their specific preferences within each genre. For example, if User 2 has shown a fondness for crime dramas, the model can recommend iconic films in that genre like "The Godfather" or "Goodfellas."



CONCLUSION

Our recommender system utilized collaborative filtering techniques to generate personalized movie recommendations for users. By leveraging the ratings and preferences of other users, we were able to provide relevant and tailored movie suggestions to each individual.

Key Findings:

We discovered that users tend to prefer movies with higher average ratings and higher review counts.

Additionally, we found that collaborative recommendations can effectively identify movies that align with users' tastes and preferences.

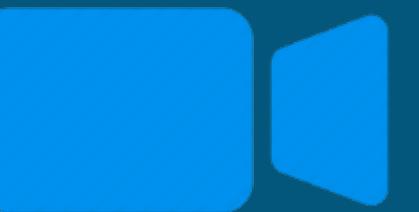
Value of Pumpkin meter

Pumpkin meter recommender system bring significant value to movie enthusiasts. Users can discover new movies based on the ratings and reviews of like-minded individuals, enhancing their movie-watching experiences

Future Possibilities

We can incorporate additional data sources, such as movie metadata and user demographics, to improve the accuracy and personalization of recommendations

THANK'S FOR
WATCHING



RIPE PUMPKINS
COMPANY

