BUAN 5315 BIG DATA ANALYSIS

A Movie Recommendation Service

Data Translation Challenge

A Movie Recommendation Service

Introduction and Dataset Overview

Technical and Debugging Details

Results

Insights and Conclusion

Introduction

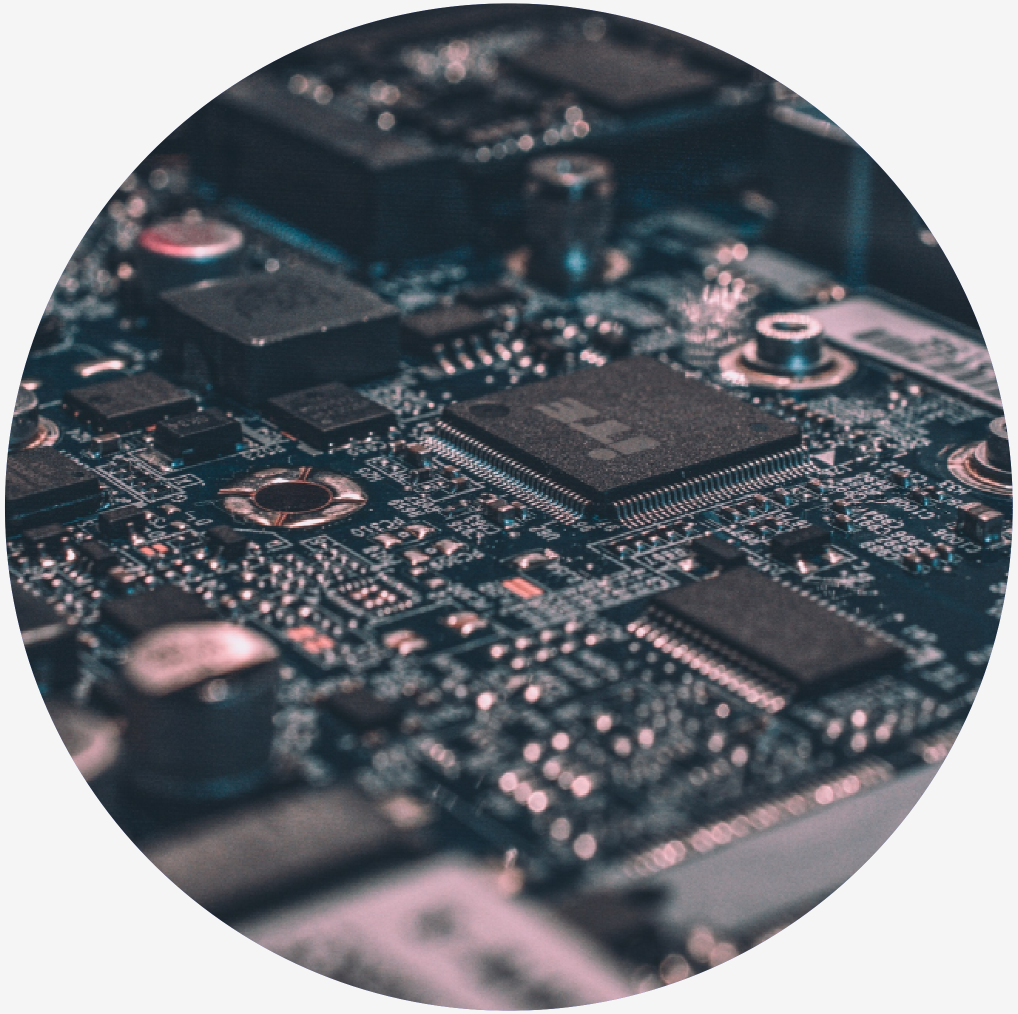
A movie recommendation system is an algorithmic that suggests movies to users based on their preferences

It leverages data analysis and machine learning techniques to provide personalized movie recommendations

Ripe Pumpkins: A new startup business in the movie industry, specializing in movie review-aggregation services

Objective: Implementing Pumpkinmeter, a recommendation engine to enhance the movie-watching experience for users

Dataset Overview



The MovieLens dataset

- 21,000,000 ratings
- 470,000 tag applications
- Applied to 27,000 movies by 230,000 users

Technical Details

- Leverages the powerful Spark's MLlib library to implement collaborative filtering
- Collaborative filtering utilizes the collective preferences of users to make personalized recommendations
- The core algorithm employed in this project is Alternating Least Squares (ALS)
- The iterative nature of ALS, controlled by the iterations parameter, allows the model to refine its predictions
- Apache Spark's machine learning library is used for collaborative filtering

Debugging Details

Compatibility

Compatibility issues and syntax errors arose due to the differences between the two versions, requiring modifications in the codebase.

Processing

Had to handle the data preprocessing and ensure the dataset was persisted for later use

Modification

.Fine-tuning the parameters of the ALS algorithm, such as rank and lambda, required some experimentation to achieve the results

Results

We conducted four test cases of two users and two scenarios to evaluate the recommendation engine's performance Scenario 1: movies with more than 25 reviews | Scenario 2: movies with more than 100 reviews

User 1 - Scenario 1

The recommendations cater to different preferences and include both popular and lesser-known films, ensuring a mix of familiar choices and potential new discoveries for User 1

User 1 - Scenario 2

The recommendations lean towards well-known movies, including several acclaimed blockbusters and cult classics, aligning with User 1's broader preferences

User 2 - Scenario 1

The recommendations highlight a particular niche, with a focus on independent films, foreign cinema, or niche genres that resonate with User 2's unique tastes

User 2 - Scenario 2

These recommendations reflect a broader range of popular and critically acclaimed films, including both mainstream and niche choices

```
TOP 15 recommended movies (with more than 25 reviews):
('Loose Change 9/11: An American Coup (2009)', 2.9481489449099487, 46)
('Jimmy Carr: Comedian (2007)', 2.929652793856018, 30)
('Jim Jefferies: Contraband (2008)', 2.8668544147203816, 31)
('Tom Segura: Completely Normal (2014)', 2.8598621319599564, 33)
('Cosmos: A Spacetime Odissey', 2.858895879752192, 37)
('Jimmy Carr: Being Funny (2011)', 2.8566759705250915, 25)
('Firebase (2017)', 2.850174269700494, 29)
('Louis C.K.: One Night Stand (2005)', 2.839191099963582, 38)
('The Lost Room (2006)', 2.8359560302691724, 280)
('Jim Jefferies: Alcoholocaust (2010)', 2.822037520915163, 53)
('Daniel Tosh: Happy Thoughts (2011)', 2.817290853546212, 27)
('Jimmy Carr: Telling Jokes (2009)', 2.808957177335726, 37)
('Perfectos desconocidos (2017)', 2.8064162374641786, 35)
('DMB (2000)', 2.805681682938827, 29)
('"Story of Luke', 2.8011097627490527, 31)
```

User 1 - Scenario 1

RECOMMENDATIONS

```
TOP 15 recommended movies (with more than 100 reviews):
('The Lost Room (2006)', 2.8359560302691724, 280)
('Black Mirror: White Christmas (2014)', 2.7906121293517714, 1074)
('Sherlock - A Study in Pink (2010)', 2.7589584333082584, 213)
('Band of Brothers (2001)', 2.755669305103818, 984)
('Law Abiding Citizen (2009)', 2.747497605442518, 2570)
('"Matrix', 2.7456716299129305, 84545)
('Black Mirror', 2.7374457472060243, 180)
('"Shawshank Redemption', 2.735032725984513, 97999)
('"Dark Knight', 2.7338810864344367, 44741)
('Saw (2003)', 2.7314120779550466, 674)
('Limitless (2011)', 2.7199078882138386, 9884)
('"Boondock Saints', 2.7151981868757353, 11214)
('Avengers: Infinity War - Part I (2018)', 2.7146535995028804, 2668)
('Gladiator (2000)', 2.707570786796058, 48666)
('Inception (2010)', 2.7033616454658347, 41475)
```

User 1 - Scenario 2

RECOMMENDATIONS

```
TOP 15 recommended movies (with more than 25 reviews):
('"Very Potter Sequel', 5.402075474560762, 35)
('Sense & Sensibility (2008)', 5.3776839046314455, 69)
('Anne of Green Gables: The Sequel (a.k.a. Anne of Avonlea) (1987)', 5.335650927546585, 342)
('Cranford (2007)', 5.323038303364424, 35)
('North & South (2004)', 5.310135356892424, 389)
('Drishyam (2013)', 5.255280979471657, 37)
('Anne of Green Gables (1985)', 5.25009056091133, 706)
('Pride and Prejudice (1995)', 5.199110857850529, 2919)
('Murder on the Orient Express (2010)', 5.182371057817914, 29)
('I Can Only Imagine (2018)', 5.180623692610432, 30)
('Winter in Prostokvashino (1984)', 5.1467791284786335, 67)
('Boys (2014)', 5.141713053618693, 96)
('Little Dorrit (2008)', 5.132472734383583, 55)
("Won't You Be My Neighbor? (2018)", 5.0994357608513745, 83)
('The Case for Christ (2017)', 5.09381497368954, 33)
```

User 2 - Scenario 1

RECOMMENDATIONS

```
TOP 15 recommended movies (with more than 100 reviews):
('Anne of Green Gables: The Sequel (a.k.a. Anne of Avonlea) (1987)', 5.335650927546585, 342)
('North & South (2004)', 5.310135356892424, 389)
('Anne of Green Gables (1985)', 5.25009056091133, 706)
('Pride and Prejudice (1995)', 5.199110857850529, 2919)
('"Sound of Music', 5.022057410686416, 17154)
('"Civil War', 4.979776328784126, 431)
('Wild China (2008)', 4.960534393596472, 105)
('Emma (2009)', 4.955623211434325, 385)
("Schindler's List (1993)", 4.953946402634649, 71516)
('Sense and Sensibility (1995)', 4.949351577356179, 24552)
('Jane Eyre (2006)', 4.930973659948204, 327)
("It's a Wonderful Life (1946)", 4.929851651350551, 17770)
('My Fair Lady (1964)', 4.90786584177767, 12089)
('Persuasion (2007)', 4.90078438594999, 349)
('Hidden Figures (2016)', 4.896389604611755, 2647)
```

User 2 - Scenario 2

RECOMMENDATIONS

Insights and Business Implications

- The analysis provided valuable insights into individual customer preferences, showcasing the ability of the Pumpkinmeter recommendation engine to deliver personalized movie suggestions
- The results highlight the potential of the recommendation engine to enhance customer satisfaction by offering tailored recommendations that align with users' tastes
- These insights have important business implications, indicating that a well-implemented recommendation system can increase user engagement, loyalty, and overall customer retention

Conclusion

The Pumpkinmeter recommendation engine demonstrates its potential to revolutionize the movie-watching experience

The successful implementation of collaborative filtering using Spark's MLlib library showcases the power of advanced machine learning techniques in delivering personalized recommendations

Moving forward, continued analysis and improvements can further enhance the accuracy and effectiveness of the recommendation engine, paving the way for a competitive advantage in the movie review-aggregation market

The insights derived from the project highlight the significant impact of personalized recommendations on customer satisfaction, engagement, and ultimately, customer retention

References

[1] https://www.codementor.io/@jadianes/building-a-recommender-with-apache-spark-python-example-app-part1-du1083qbw

[2] https://projectsbasedlearning.com/apache-spark-machine-learning/machine-learning-project-creating-movies-recommendation-engine-using-apache-spark/

[3] https://medium.com/edureka/spark-mllib-e87546ac268