iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** XI **Month of publication:** November 2020

DOI: https://doi.org/10.22214/ijraset.2020.32064

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XI Nov 2020- Available at www.ijraset.com

Movie Recommendation System

Adithya B[1], Amrish K N[2], Shreyas D[3], Dr. M. Sujithra[4], Dr. P. Velvadivu[5]
[1, 2, 3]*M.Sc. Data Science, Coimbatore Institute of Technology, Coimbatore.*
[4, 5]*Assistant Professor, Department of Data Science, Coimbatore Institute of Technology, Coimbatore.*

Abstract: Recommendation systems have become an important role in today's digital world. ***During the last few decades, with the rise of YouTube, Amazon, Netflix and many other such web services, recommender systems have taken more and more place in our lives. From e-commerce (suggest to buyers' articles that could interest them) to online advertisement (suggest to users the right contents, matching their preferences), recommender systems are today unavoidable in our daily online journeys. Users get more data related to them using these systems. Here, we are developing a recommendation system to suggest movies to users.***
Keywords: Bigdata, Pyspark, Machine Learning, Recommend

I. DATASET DESCRIPTION

The dataset has 10 million ratings and 100,000 tag applications applied to 10,000 movies by 72,000 users. It was released in 1/2009.
Dataset link: https://grouplens.org/datasets/movielens/
Attributes

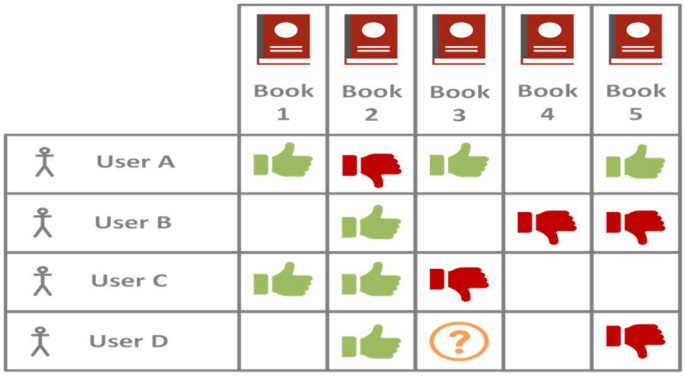
A. Movie id
B. Movie name
C. Rating
D. User id
E. No of ratings

A. PYSPARK

Apache spark is an open source clustering computing framework. This has been written in Scala language. Pyspark is a collaboration of Apache spark. It can perform both stream processing and batch processing. It is widely used for machine learning and real time streaming analytics. When you use PySpark, you will likely to get high data processing speed of about 10x faster on the disk and 100x faster in memory. By reducing the number of read-write to disk, this would be possible.

II. COLLABORATIVE FILTERING

Collaborative Filtering is the most common technique used when it comes to building intelligent recommender systems that can learn to give better recommendations as more information about users is collected. Collaborative Filtering is the most common technique used when it comes to building intelligent recommender systems that can learn to give better recommendations as more information about users is collected. Collaborative filtering is a technique that can filter out items that a user might like on the basis of reactions by similar users. It works by searching a large group of people and finding a smaller set of users with tastes similar to a particular user. It looks at the items they like and combines them to create a ranked list of suggestions.



©IJRASET: All Rights are Reserved

III. OUTPUT AND INFERENCE

```
Most rated movies:
(average rating, movie name, number of reviews)
(4.457238321660348, 'Shawshank Redemption, The (1994)', 31126)
(4.415085293227011, 'Godfather, The (1972)', 19814)
(4.367142322253193, 'Usual Suspects, The (1995)', 24037)
(4.363482949916592, "Schindler's List (1993)", 25777)
(4.321966205837174, 'Sunset Blvd. (a.k.a. Sunset Boulevard) (1950)', 3255)
(4.319740945070761, 'Casablanca (1942)', 12507)
(4.316543909348442, 'Rear Window (1954)', 8825)
(4.315439034540158, 'Double Indemnity (1944)', 2403)
(4.313629402756509, 'Third Man, The (1949)', 3265)
(4.314119283602851, 'Seven Samurai (Shichinin no samurai) (1954)', 5751)
(4.306805399325085, 'Paths of Glory (1957)', 1778)
(4.303215119343423, 'Godfather: Part II, The (1974)', 13281)
(4.298072023101749, 'Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb (1964)', 11774)
(4.297154471544715, 'Lives of Others, The (Das Leben der Anderen) (2006)', 1230)
(4.294842186297152, 'Dark Knight, The (2008)', 2598)
(4.292379632836855, "One Flew Over the Cuckoo's Nest (1975)", 14435)
(4.277613703484938, 'Yojimbo (1961)', 1693)
(4.275907715582451, 'Wallace & Gromit: The Wrong Trousers (1993)', 7932)
(4.274952621604548, 'Wallace & Gromit: A Close Shave (1995)', 6332)
(4.267953020134228, 'Big Sleep, The (1946)', 2980)
(4.2662155856276245, 'M (1931)', 2143)
(4.261899313501144, 'City of God (Cidade de Deus) (2002)', 4370)
(4.261317249919736, 'Raiders of the Lost Ark (Indiana Jones and the Raiders of the Lost Ark) (1981)', 21803)
(4.261366341347299, 'North by Northwest (1959)', 8402)
(4.257643884892087, 'General, The (1927)', 1112)
(4.248271527341295, 'Rashomon (Rashômon) (1950)', 1591)
(4.246875322797232, "Amelie (Fabuleux destin d'Amélie Poulain, Le) (2001)", 9681)
```

The above output is a basic recommendation for the users. It contains the list of movies with highest average ratings and number of ratings above 500.

```
For rank 4 the RMSE is 0.8298404088886056
For rank 8 the RMSE is 0.8304823068644719
For rank 12 the RMSE is 0.8347057117552411
The best model was trained with rank 4
```

We checked error for 3 ranks of matrix, and chose rank 2 as it has the least error among them.

```
My movie ratings: [(0, 1, 3), (0, 2, 2), (0, 6, 4), (0, 10, 3), (0, 11, 3), (0, 22, 3), (0, 31, 3), (0, 32, 4),
(0, 47, 5), (0, 50, 5)]
```

This is given as input as the ratings given by user for 10 movies. With these, we are going to predict top movies which this user will like.

```
My highest rated movies as predicted (for movies with more than 75 reviews):
(4.445590345459509, u'American Beauty (1999)', 1775)
(4.394451696808654, u'Fight Club (1999)', 693)
(4.29927344173635, u'Pulp Fiction (1994)', 1039)
(4.277656880495439, u'Paths of Glory (1957)', 105)
(4.2654477477781505, u'Raising Arizona (1987)', 658)
(4.258061810500499, u'Reservoir Dogs (1992)', 603)
(4.251964788817356, u'Clerks (1994)', 711)
(4.244293576533277, u'Shawshank Redemption, The (1994)', 1088)
(4.23562818710623, u'Matrix, The (1999)', 1250)
(4.2307444008380175, u'Monty Python and the Holy Grail (1974)', 759)
(4.226396344970352, u'American History X (1998)', 341)
(4.224260354182621, u'Life Is Beautiful (La Vita \ufffd bella) (1997)', 587)
(4.22006116171962, u'Close Shave, A (1995)', 318)
(4.219422220834151, u'Almost Famous (2000)', 744)
(4.181639787867589, u'Wrong Trousers, The (1993)', 425)
(4.1761144228049885, u'Braveheart (1995)', 1300)
(4.174496017804163, u'Godfather, The (1972)', 1047)
(4.172596219198133, u'GoodFellas (1990)', 811)
(4.163993624303176, u'Apocalypse Now (1979)', 539)
(4.163579871619603, u'Requiem for a Dream (2000)', 214)
```

So, we obtained the top-rated movies similar to the likes of the user.

©IJRASET: All Rights are Reserved

IV. CONCLUSION

This model can be used to recommend movies to customers based on their interest. This model has a good accuracy and less error.

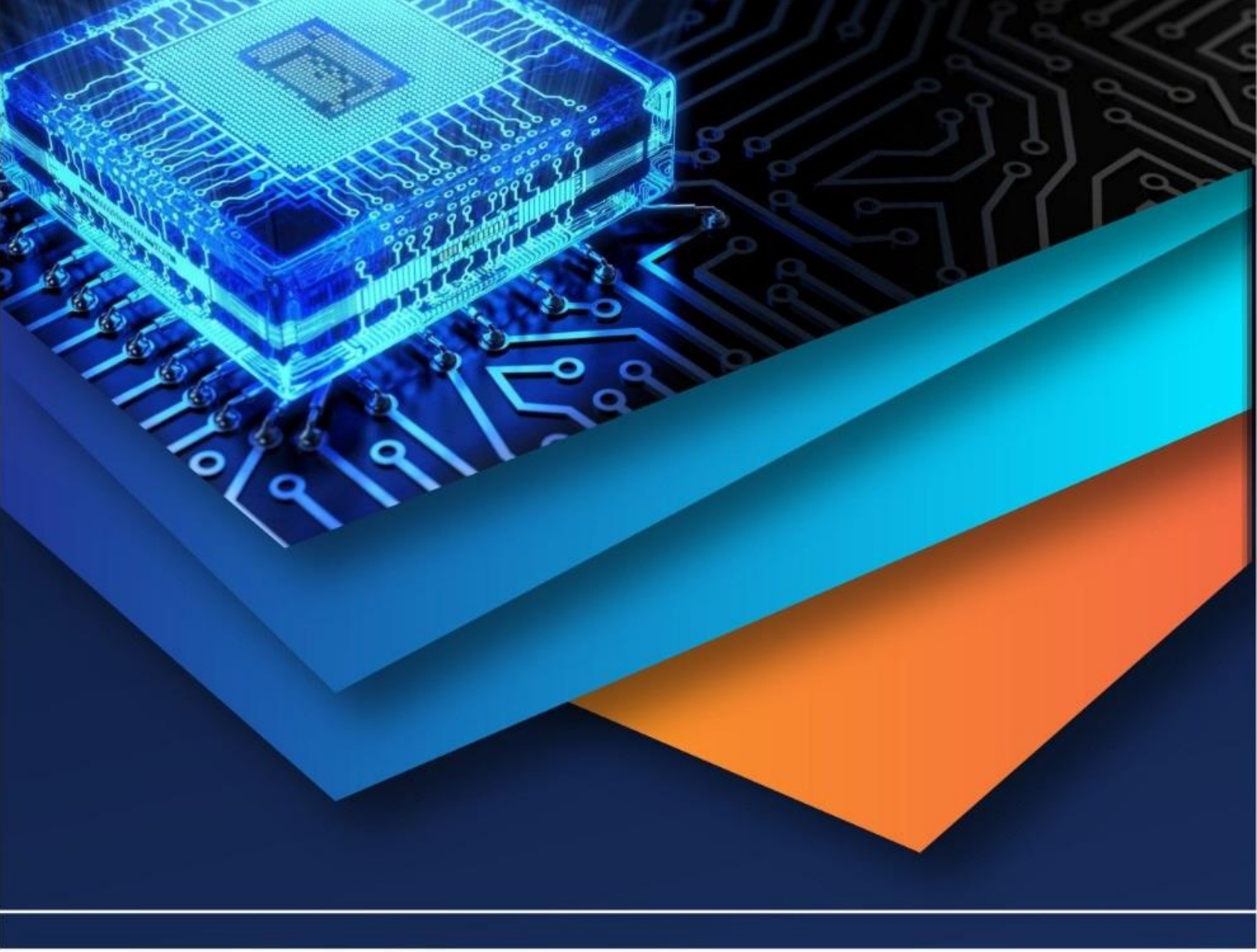
V. FUTURE SCOPE

This recommendation system can be used in real time recommendations like in Netflix, Amazon Prime, YouTube. It can also be used in product recommendations like in flipkart, amazon, etc.,

REFERENCES

[1] https://ieeexplore.ieee.org/document/8663822
[2] https://www.researchgate.net/publication/319487277_Design_and_Implementation_of_Movie_Recommendation_System_Based_on_Knn_Collaborative_Filtering_Algorithm
[3] https://link.springer.com/chapter/10.1007/978-3-642-21793-7_63
[4] https://www.semanticscholar.org/paper/A-Movie-Recommender-System%3A-MOVREC-Kumar-Yadav/621c3acadd3de1401d2bbf039ba4afb43f6e4dd2

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)