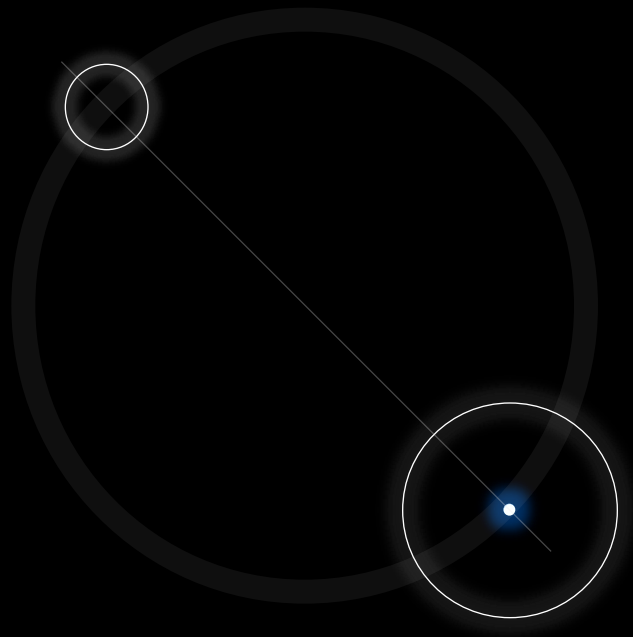
MOVIE RECOMMENDATION SYSTEM

JIAPENG ZHANG

XIAO LIU

MALHAR KHIMSARIA



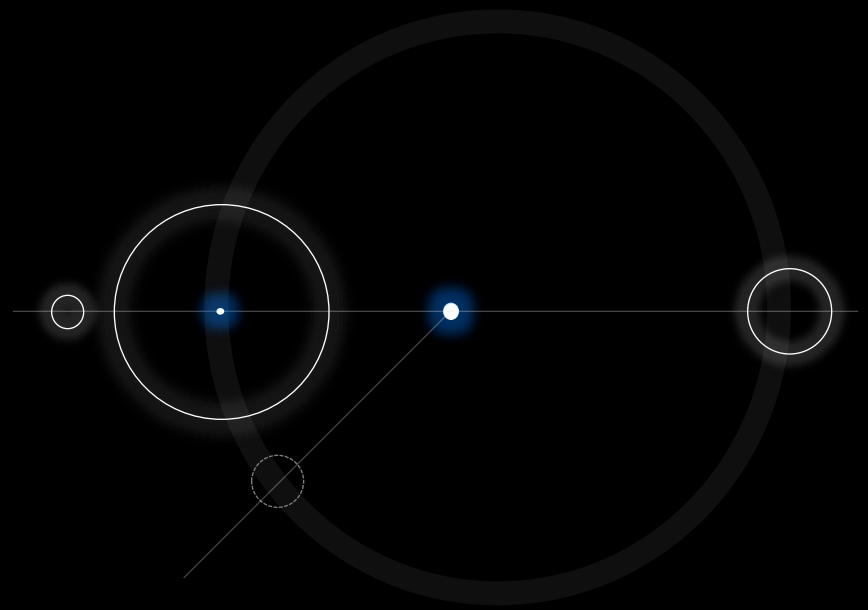
PROJECT GOAL

In the ”information overload” age, make an informed choice is hard. Too many choices often make people at a loss.

We present a movie recommendation system that helps to recommend movies for users based on collaborative filtering approach with KNN and KMeans clustering methods. This system can help users discover movies they might enjoy.

OUTLINE

- Description
- Snapshots
- Project Stumbling Blocks
- Data Collection
- Flow Diagram
- Integrity Constraints
- Future Extensions
- Acknowledgements
- References and Resources used

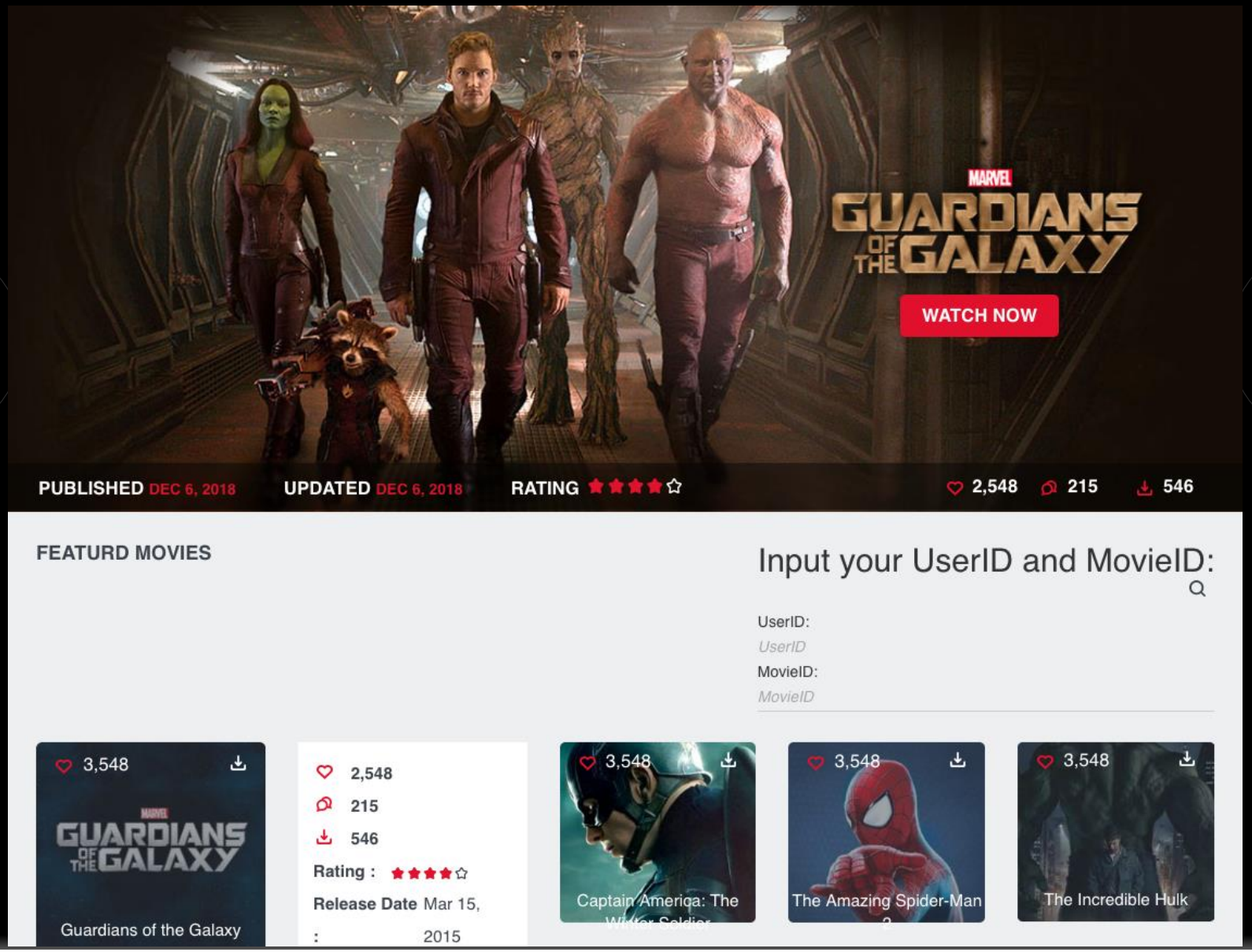


DESCRIPTION

Too many choices of movies often make people confused. To help people to relief this pain, we present a movie recommendation system that helps to recommend movies for users based on collaborative filtering approach with KNN and KMeans clustering methods. With the help of algorithm, users can discovery what might they like.

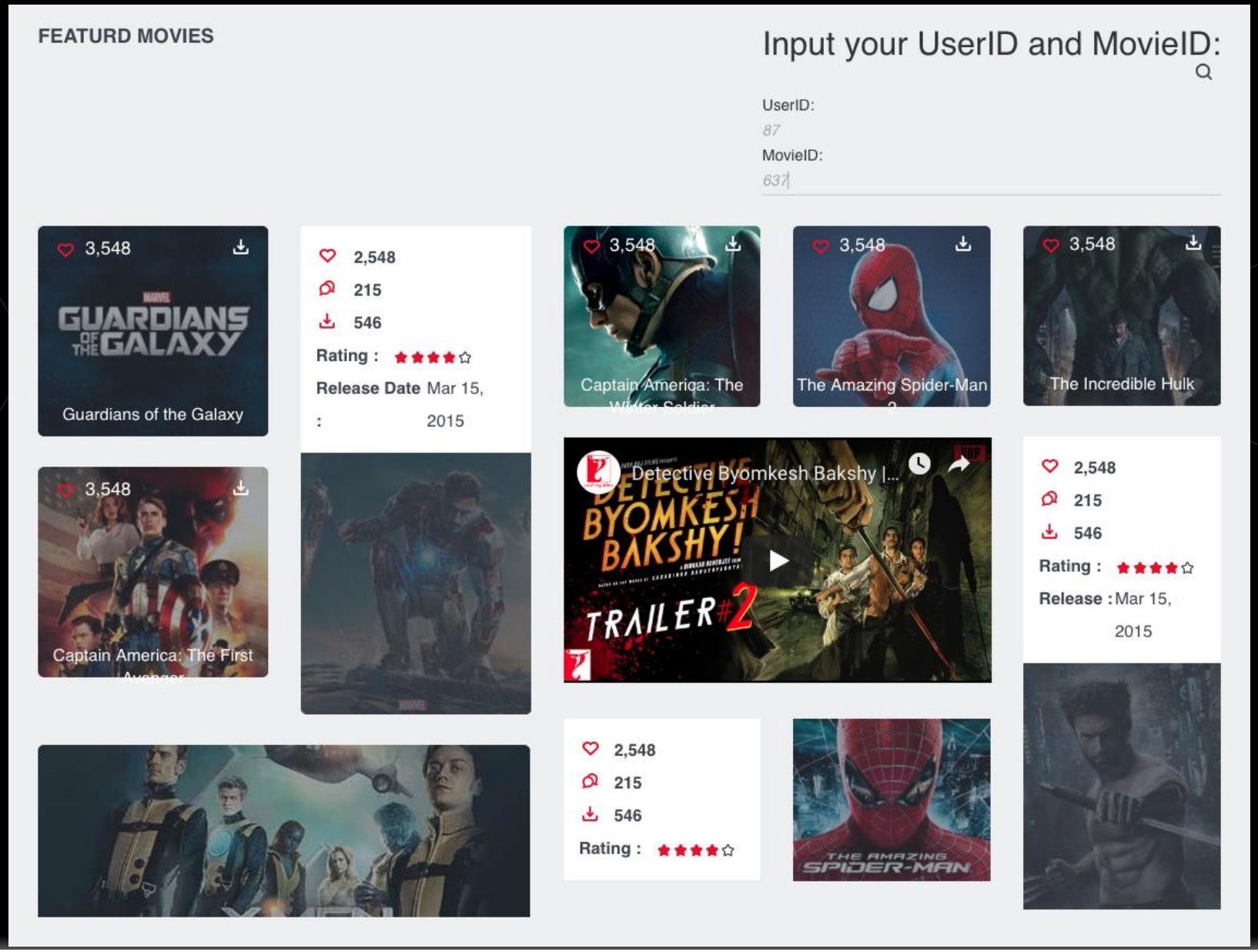
SNAPSHOTS

HOMEPAGE



SNAPSHOTS

Input
UserID:87
MovieID: 637



SNAPSHOTS

Result of UserID:87 MovieID:637

UserID:87

The Movie that you like:637

Movie 637: Return of Martin Guerre, The (Retour de Martin Guerre, Le) (1982)

Your avg score for this class is 3.5

You might also like:

Film id: 6| Film Title: Shanghai Triad (Yao a yao yao dao waipo qiao) (1995)

Film id: 7| Film Title: Twelve Monkeys (1995)

Film id: 9| Film Title: Dead Man Walking (1995)

Film id: 15| Film Title: Mr. Holland's Opus (1995)

Film id: 18| Film Title: White Balloon, The (1995)

Film id: 19| Film Title: Antonia's Line (1995)

Film id: 30| Film Title: Belle de jour (1967)

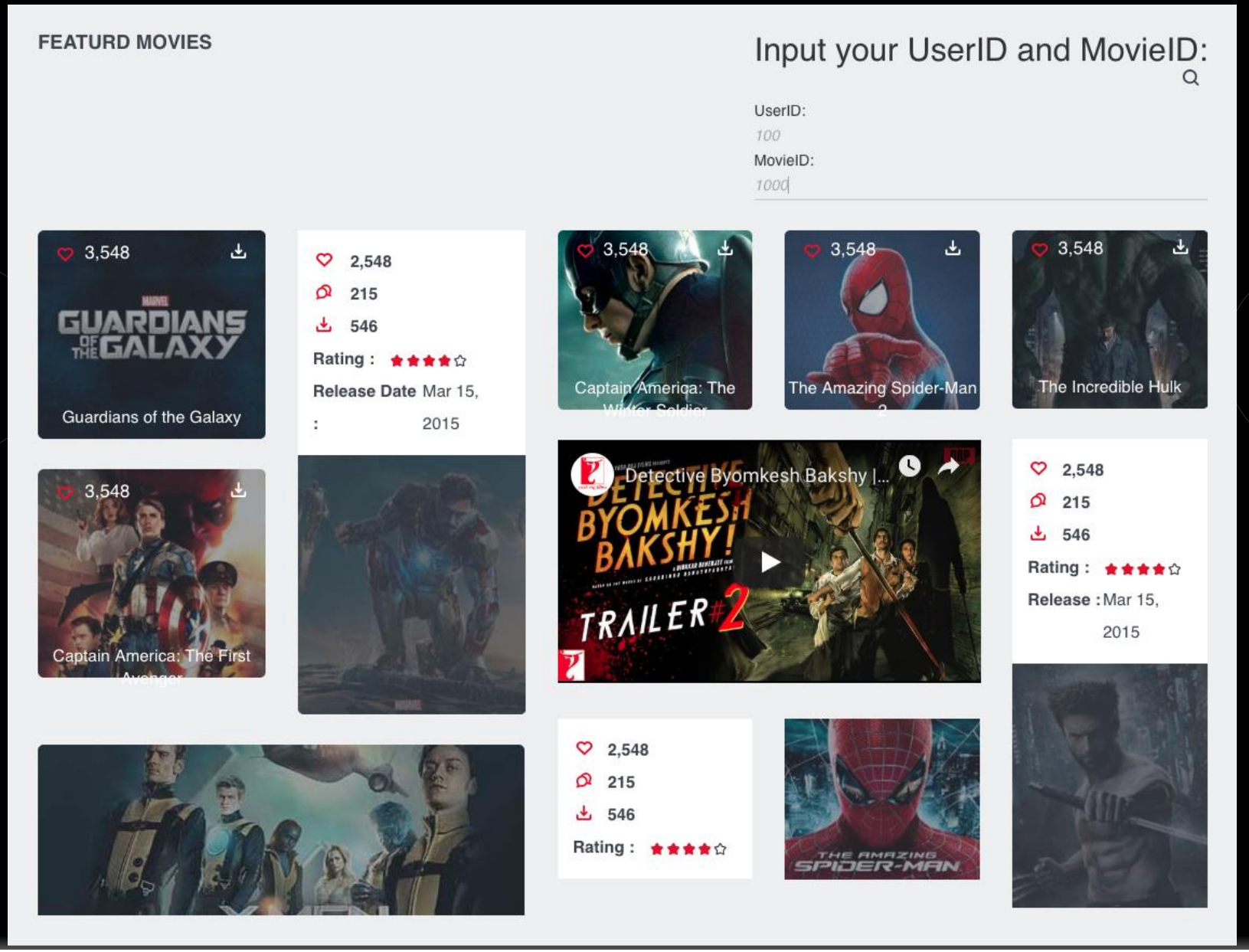
Film id: 37| Film Title: Nadja (1994)

Film id: 46| Film Title: Exotica (1994)

Film id: 52| Film Title: Madness of King George, The (1994)

SNAPSHOTS

Input
UserID:100
MovieID: 1000



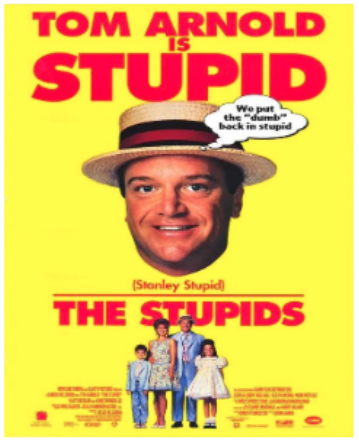
SNAPSHOTS

Result of UserID:100 MovieID:1000

UserID:100

The Movie that you like:1000

Movie 1000: The Stupids (1996)



Your avg score for this class is 2.1538461538461537

You might also like:

Film id: 13| Film Title: Mighty Aphrodite (1995)

Film id: 25| Film Title: Birdcage, The (1996)

Film id: 26| Film Title: Brothers McMullen, The (1995)

Film id: 40| Film Title: To Wong Foo, Thanks for Everything! Julie Newmar (1995)

Film id: 41| Film Title: Billy Madison (1995)

Film id: 42| Film Title: Clerks (1994)

Film id: 63| Film Title: Santa Clause, The (1994)

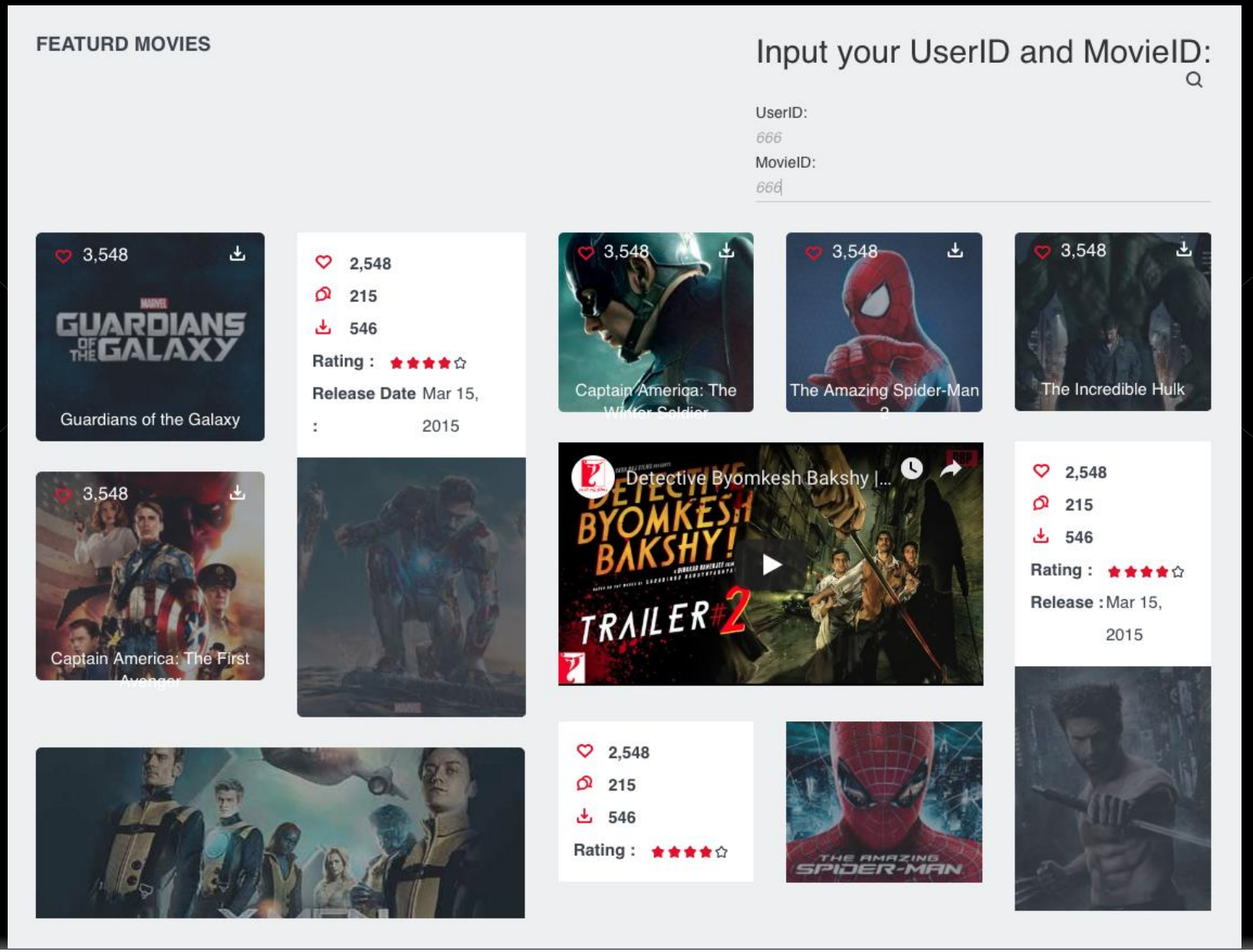
Film id: 67| Film Title: Ace Ventura: Pet Detective (1994)

Film id: 72| Film Title: Mask, The (1994)

Film id: 73| Film Title: Maverick (1994)

SNAPSHOTS

Input
UserID:666
MovieID: 666



SNAPSHOTS

Result of
UserID:666
MovieID:666

UserID:666

The Movie that you like:666

Movie 666: Audrey Rose (1977)

Your avg score for this class is 3.2471364547440293

You might also like:

Film id: 84| Film Title: Robert A. Heinlein's The Puppet Masters (1994)

Film id: 123| Film Title: Frighteners, The (1996)

Film id: 185| Film Title: Psycho (1960)

Film id: 200| Film Title: Shining, The (1980)

Film id: 208| Film Title: Young Frankenstein (1974)

Film id: 217| Film Title: Bram Stoker's Dracula (1992)

Film id: 219| Film Title: Nightmare on Elm Street, A (1984)

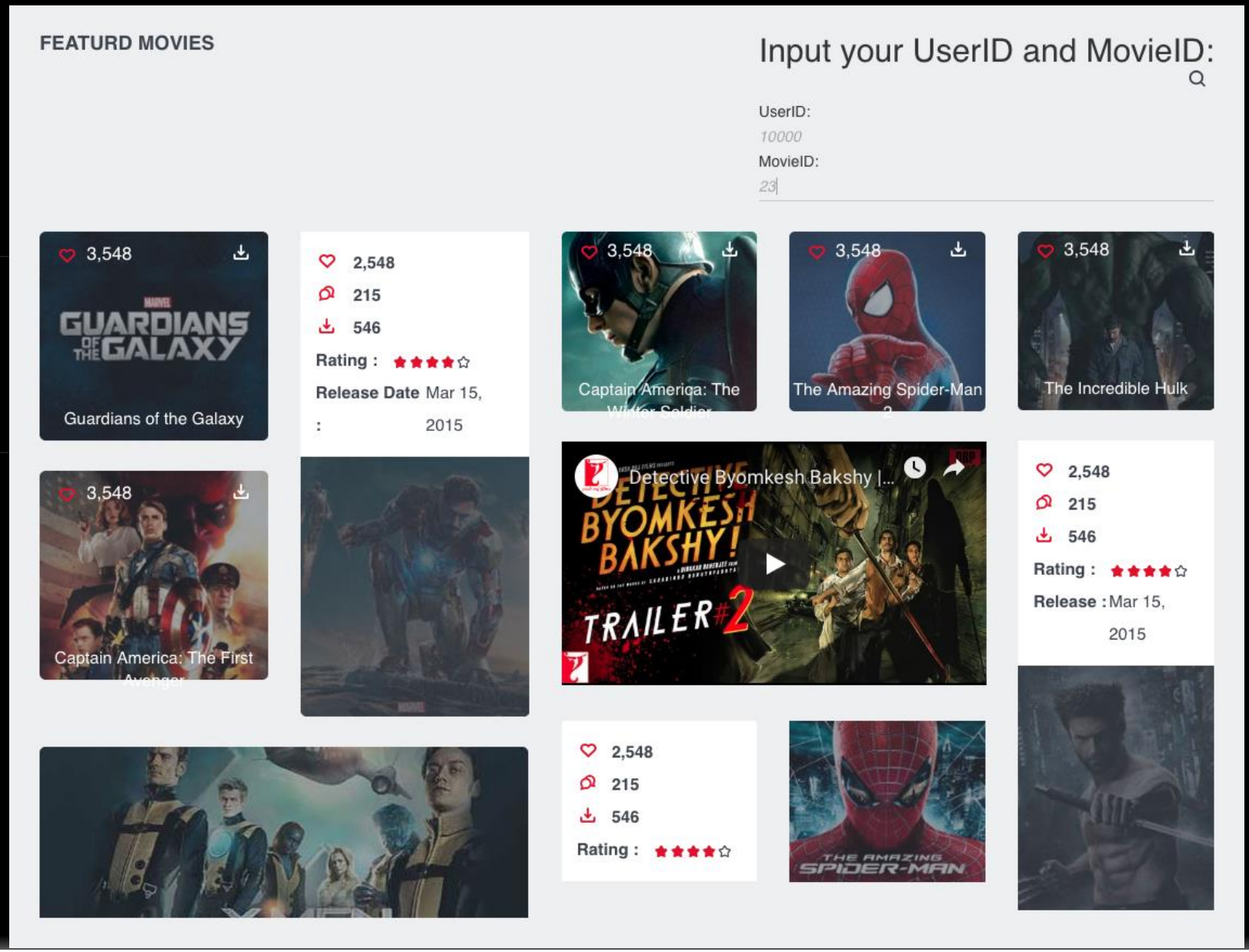
Film id: 234| Film Title: Jaws (1975)

Film id: 288| Film Title: Scream (1996)

Film id: 343| Film Title: Alien: Resurrection (1997)

SNAPSHOTS

Input(Invlid)
UserID:10000
MovieID: 23

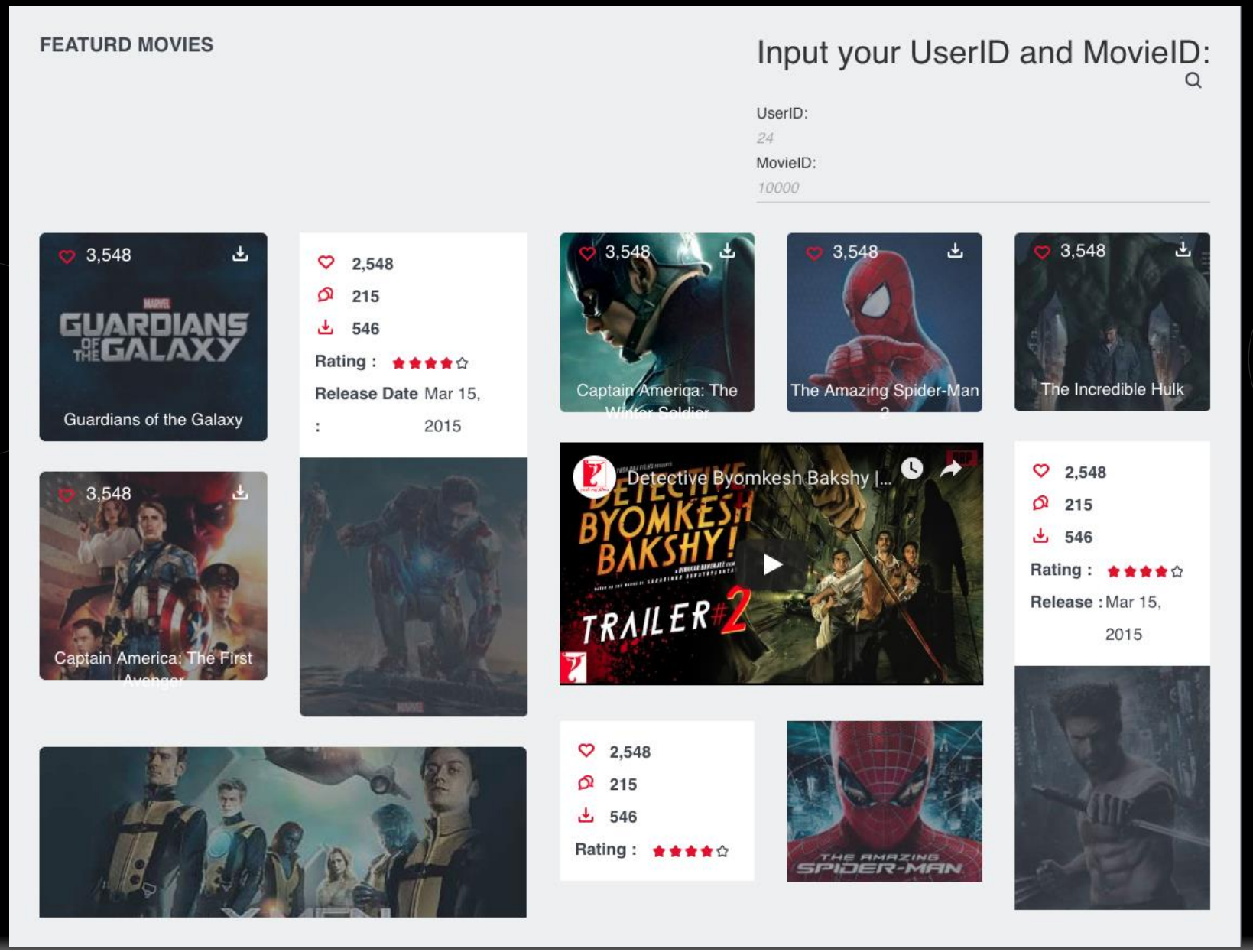


SNAPSHOTS

Result of
UserID:10000
MovieID:23

SNAPSHOTS

Input(Invalid)
UserID:24
MovieID: 10000

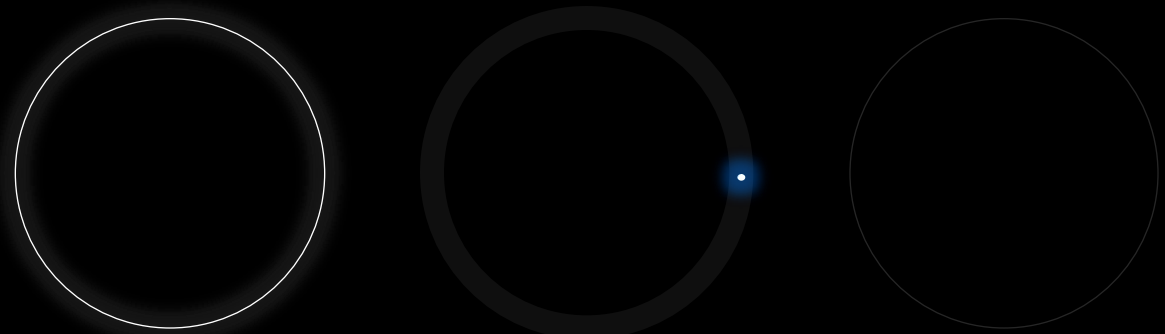


SNAPSHOTS

Result of
UserID:24
MovieID:10000

PROJECT STUMBLING BLOCKS

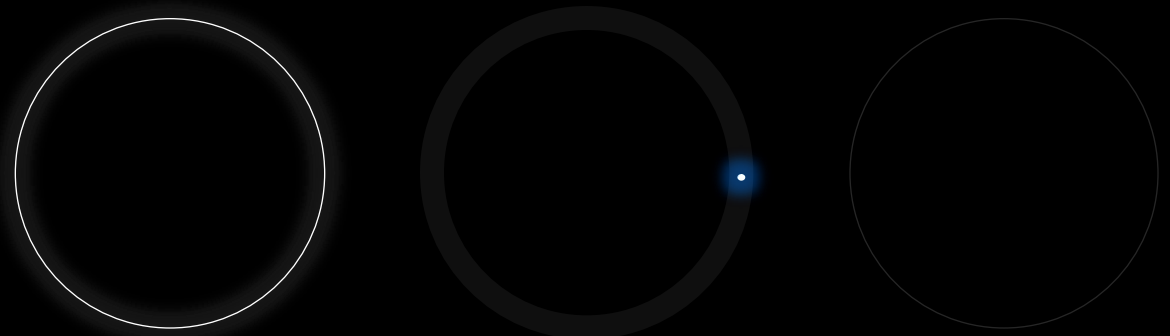
Because of the sparse feature of the massive date and cold-start problem of some new user, the recommended results can be inaccurate sometime.



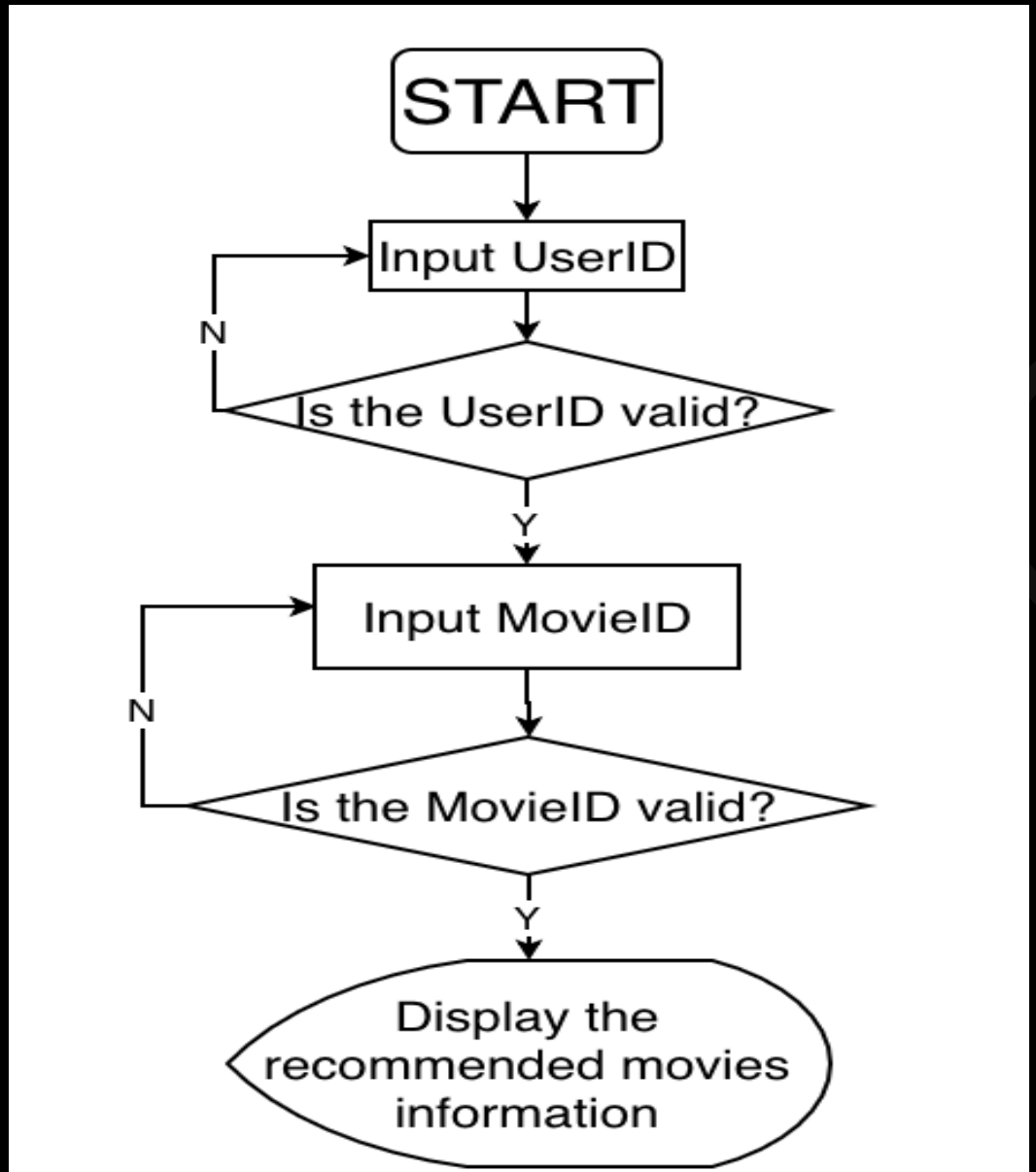
DATA COLLECTION

GroupLens Research has collected and made available rating data sets from the MovieLens web site (http://movielens.org). The data sets were collected over various periods of time, depending on the size of the set.

In this project we used the Stable benchmark dataset, which included 1 million ratings from 6000 users on 4000 movies. Released 2/2003.

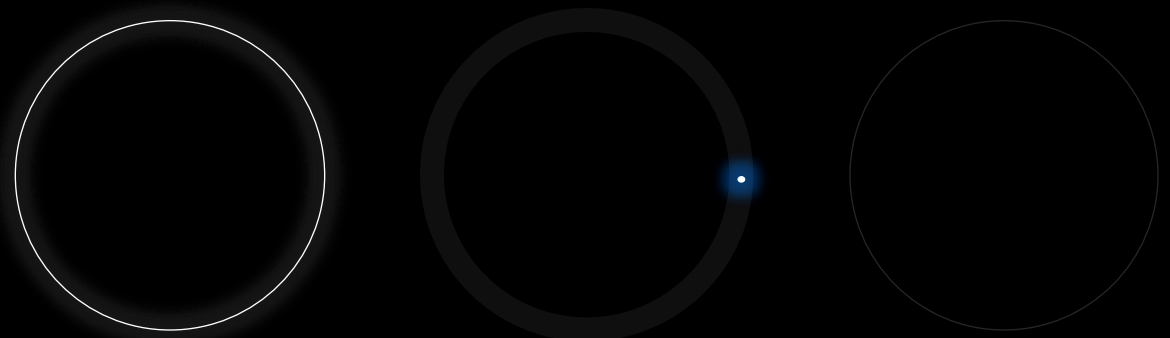


FLOW DIAGRAM



INTEGRITY CONSTRAINTS

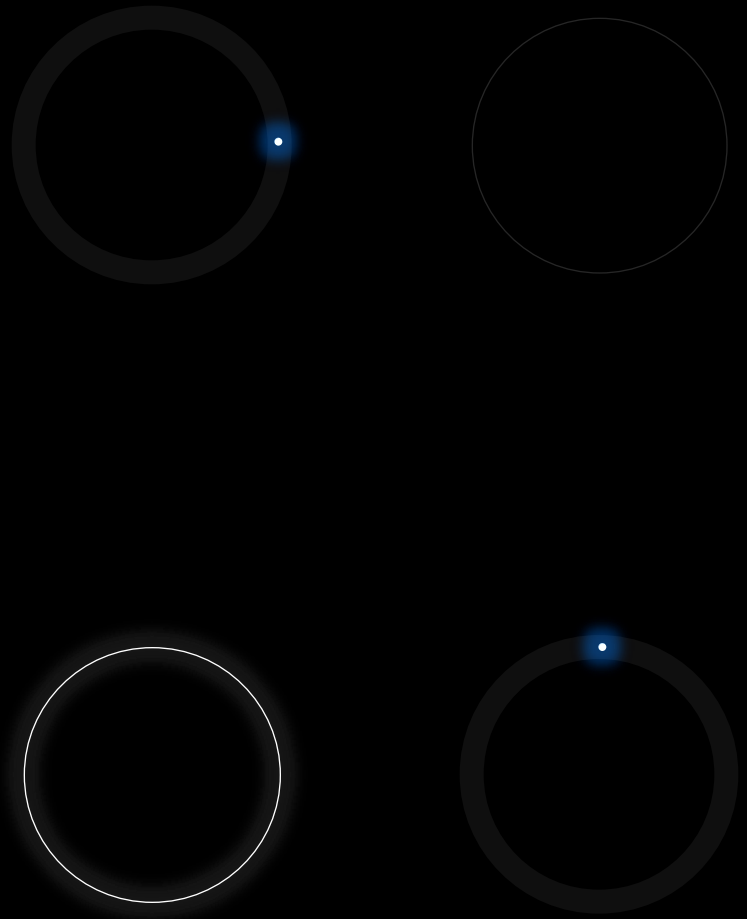
- Input Integrity Constraint: Title(string)
- Output Integrity Constraint 1: MovieID (integer, range between 0 and 3952)
- Output Integrity Constraint 2: Title(string)
- Output Integrity Constraint 3: Ratings(integer, range between 0 and 5)



EXTENTION FEATURES

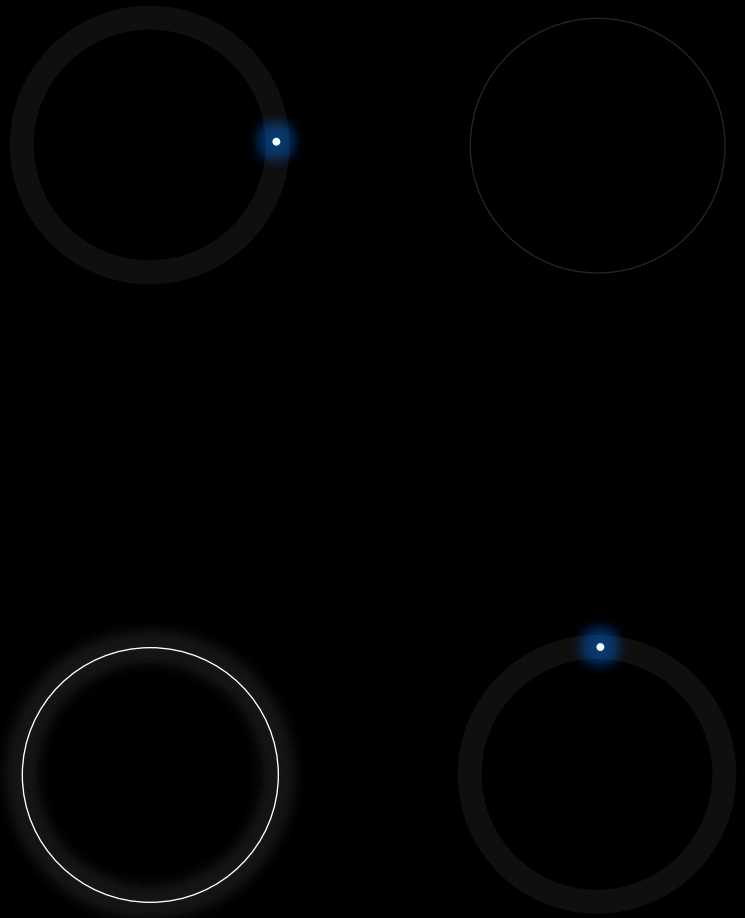
Our model is entirely based on off-the-shelf data, i.e. MovieLens data. If we can modify our model to support online machine training, that would be much more powerful and suitable for real-life applications.

Still, there's much room for improvement for our Web UI. For example,
it's possible to add drop down menu to select movies, adding more convenience.



ACKNOWLEDGEMENTS

Thanks for the hard work of GroupLens Research, without whose datasets we cannot finish our project.
And also thanks for our TA, Harsha's advice on how to do our project.



REFERENCES

- **Dataset**: https://grouplens.org/datasets/movielens/1m/
- **Collaborative Filtering**: https://en.wikipedia.org/wiki/Collaborative_filtering
- **sklearn.cluster.Kmeans**: https://scikitlearn.org/stable/modules/generated/sklearn.cluster.KMeans.html#sklearn.cluster.KMeans

- **NumPy**: https://docs.scipy.org/doc/numpy-1.13.0/reference/