

MOVIE RECOMMENDATION SYSTEM

A PROJECT REPORT

Submitted By

Ojasvi Tyagi

(University Roll No- 2000290140081)

Tanu Sharma

(University Roll No-2000290140125)

Rajat Deol

(University Roll No-2000290140096)

Shivani Chauhan

(University Roll No- 2000290140114)

**Submitted in partial fulfilment of the
Requirements for the Degree of**

MASTER OF COMPUTER APPLICATIONS

**Under the Supervision of
Vidushi Mishra
Assistant Professor,
KIET Group of Institutions**



Submitted to

**DEPARTMENT OF COMPUTER APPLICATIONS
KIET Group of Institutions, Ghaziabad
Uttar Pradesh-201206**

CERTIFICATE

Certified that **Ojasvi Tyagi (Enrollment No-20002901401400), Rajat Deol (Enrollment No-200029014005781), Tanu Sharma (Enrollment No-200029014014005810), Shivani Chauhan (Enrollment No-200029014014005799)** have carried out the project work having “**Title of Report – Movie Recommendation**” for Master of Computer Applications from Dr. A.P.J. Abdul Kalam Technical University (AKTU) (formerly UPTU), Technical University, Lucknow under my supervision. The project report embodies original work, and studies are carried out by the student himself / herself and the contents of the project report do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

Date:

Ojasvi Tyagi (2000290140081)

Rajat Deol (2000290140096)

Tanu Sharma (2000290140125)

Shivani Chauhan(2000290140114)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Date:

**Mrs.Vidushi Misra
Assistant Professor
Department of Computer Applications
KIET Group of Institutions, Ghaziabad**

Signature of Internal Examiner

Signature of External Examiner

Dr. Ajay Shrivastava
Head, Department of Computer Applications
KIET Group of Institutions, Ghaziabad

ABSTRACT

This paper discusses about recommendations of the movies. A movie recommendation is important in our social life due to its strength in providing enhanced entertainment. Such a system can suggest a set of movies to users based on their interest, or the popularities of the movies. A recommendation system is used for the purpose of suggesting items to purchase or to see. They direct users towards those items which can meet their needs through cutting down large database of Information. A recommender system, or a recommendation system (sometimes replacing 'system' with a synonym such as platform or engine), is a subclass of information filtering system that seeks to predict the "rating" or "preference" a user would give to an item. They are primarily used in commercial applications. MOVREC also help users to find the movies of their choices based on the movie experience of other users in efficient and effective manner without wasting much time in useless browsing. Keywords: Filtering, Recommendation System, Recommender.

ACKNOWLEDGEMENTS

Success in life is never attained single handedly. My deepest gratitude goes to my thesis supervisor, **Mrs. Vidushi Misra** for her guidance, help and encouragement throughout my research work. Their enlightening ideas, comments, and suggestions.

Words are not enough to express my gratitude to Dr. Ajay Kumar Shrivastava, Professor and Head, Department of Computer Applications, for his insightful comments and administrative help at various occasions.

Fortunately, I have many understanding friends, who have helped me a lot on many critical conditions.

Finally, my sincere thanks go to my family members and all those who have directly and indirectly provided me moral support and other kind of help. Without their support, completion of this work would not have been possible in time. They keep my life filled with enjoyment and happiness.

Ojasvi Tyagi

Rajat Deol

Tanu Sharma

Shivani Chauhan

TABLE OF CONTENTS

Chapter 1 - Introduction	
Project description	
Project Scope	
Hardware / Software	
Chapter 2 Feasibility Study	
Technical feasibility	
Operational feasibility	
Chapter 3 Technologies Used	
Chapter 3 Database Design	
Database Tables	
Flow Chart	
Chapter 4 Form Design	
4.1 Input/Output	
Chapter 6 Testing	
6.1 Test Cases	
Chapter 7 Conclusion	
Chapter 8 Bibliography	

Introduction

Project description

A movie recommendation is important in our social life due to its strength in providing enhanced entertainment. Such a system can suggest a set of movies to users based on their interest, or the popularities of the movies. A recommendation system is used for the purpose of suggesting items to purchase or to see. They direct users towards those items which can meet their needs through cutting down large database of Information. A recommender system, or a recommendation system (sometimes replacing 'system' with a synonym such as platform or engine), is a subclass of information filtering system that seeks to predict the "rating" or "preference" a user would give to an item. They are primarily used in commercial applications. MOVREC also help users to find the movies of their choices based on the movie experience of other users in efficient and effective manner without wasting much time in useless browsing. Keywords: Filtering, Recommendation System, Recommender.

Project Scope

A recommendation system has become an indispensable component in various e-commerce applications. Recommender systems collect information about the user's preferences of different items (e.g. movies, shopping, tourism, TV, taxi) by two ways, either implicitly or explicitly. An implicit acquisition of user information typically involves observing the user's behavior such as watched movies, purchased products, downloaded applications. On the other hand, a direct procurement of information typically involves collecting the user's previous ratings or history. Collaborative filtering (CF) is the way of filtering or calculating items through the sentiments of other people. It first gathers the movie ratings given by individuals and then recommends movies to the target user based on like-minded people with similar tastes and interests in the past. Additional impression on which some recommender systems are based is clustering. Clustering is a popular unsupervised data mining tool that is used for partitioning a given dataset into homogeneous groups based on some similarity or dissimilarity metric. Collaborative filtering and clustering have been discussed in detail in the next section. Hybrid cluster and optimization approach is applied to improve movie prediction accuracy. Such a hybrid approach has been used to overcome the limitations of typical content-based and collaborative recommender systems. For clustering, k-means algorithm is applied and for optimization, cuckoo search optimization is implemented. K-means algorithm is an enormously greater clustering algorithm when compared to other clustering methods in relations of time, complexity or effectiveness for a particular number of clusters. Clustering algorithm with a bio-inspired algorithm such as cuckoo search delivers optimize results. The cuckoo search has shown best performance when compared with other algorithms such as genetic algorithms and particle swarm optimization. Simulations and comparison of the cuckoo search were greater to these existing algorithms for multimodal objective functions. To find the best results we have to find the most suitable weight among all possible ones.

Hardware / Software used in Project

Hardware Used in Project

- Window 10
- 8 RAM
- i7 processor
- With 512 GB SSD

Software Used in Project

- Hadoop
- Oracle virtualBox
- Centos
- Putty

Feasibility Study

Technical feasibility

The objective of the technical feasibility step is to confirm that the product will perform and to verify that there are no production barriers. Product: The product of this activity is a working model.

Operational Feasibility

The operational feasibility to help users find items that they deem of interest to them. They can be seen as an application of data mining process. In this paper, a new recommender system based on multi-features is introduced. Demographic and psychographic features are used to assess similarities between users.

Technologies Used

Hadoop

Hadoop is an Apache open-source framework written in java that allows distributed processing of large datasets across clusters of computers using simple programming models. The Hadoop framework application works in an environment that provides distributed storage and computation across clusters of computers. Hadoop is designed to scale up from single server to thousands of machines, each offering local computation and storage.

HIVE

Apache Hive is a distributed, fault-tolerant data warehouse system that enables analytics at a massive scale. A data warehouse provides a central store of information that can easily be analyzed to make informed, data driven decisions. Hive allows users to read, write, and manage petabytes of data using SQL.

Hive is built on top of Apache Hadoop, which is an open-source framework used to efficiently store and process large datasets. As a result, Hive is closely integrated with Hadoop, and is designed to work quickly on petabytes of data. What makes Hive unique is the ability to query large datasets, leveraging Apache Tez or MapReduce, with a SQL-like interface.

Map Reduce

MapReduce is a programming paradigm that enables massive scalability across hundreds or thousands of servers in a Hadoop cluster. As the processing component, MapReduce is the heart of Apache Hadoop.

HDFS

HDFS is a distributed file system that handles large data sets running on commodity hardware. It is used to scale a single Apache Hadoop cluster to hundreds (and even thousands) of nodes. HDFS is one of the major components of Apache Hadoop, the others being map reduce and YARN.

Database Design

Dataset

A dataset is a collection of records, similar to a relational database table. Records are similar to table rows, but the columns can contain not only strings or numbers, but also nested data structures such as lists, maps, and other records.

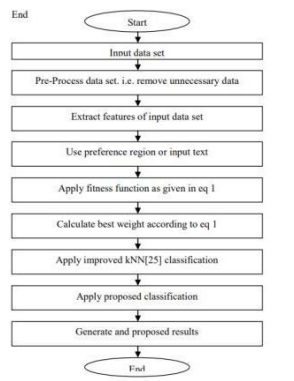
Basically, we have stored our Dataset with the name Dataset with csv file format. The Dataset contains many movies with their releasing year (which movie has been released in which year) as well as Genre such as Comedy, Action, Drama, Family, Horror, Science Thriller, Romance, Biography.

Database Tables

For the database portion, we use a table of dataset which include a large amount of details of the movies. Screenshot of such Dataset is:

S.NO	MOVIE NAME	GENRE
1	Aag Ka Dai	1990
2	Aashiqui	1990 Romance
3	Awaargi	1990 Thriller
4	Baaghi: A	1990 Action, Romance
5	Dil	1990 Comedy, Romance
6	Jawani Zin	1990 Romance
7	Deewana I	1990 Romance
8	Ghayal	1990 Thriller
9	Izzatdaar	1990 Drama, Thriller
10	Maha-Sanj	1990 Drama
11	Swarg	1990 Drama
12	Zakhami Zai	1990 Drama
13	Agneepath	1990 Action, Crime, Drama
14	100 DAYS	1991 Thriller
15	Alpoba	1991 Fantasy adventure
16	Dil Hai Ke	1991 Romance
17	Henna	1991 Romance, Drama
18	Hum	1991 Action, Drama
19	Kurbaan	1991 Romance
20	Lamhe	1991 Romance, Drama
21	Love	1991 Romance
22	Patthar Ke	1991 Romance, Thriller
23	Phool Aur	1991 Thriller
24	Prahaar: T	1991 Action
25	Saajan	1991 Romance
26	Sadak	1991 Romance
27	Sanam Bew	1991 Romance
28	Saudagar	1991 Romance, Drama
29	Adharm	1992 Thriller
30	Beta	1992 Drama
31	Badrinath	1993 Romance

Flow Chart



Form Design

Input / Output Form (Screenshot)

• Input Screenshot

Table creation

```
root@localhost:~/hadoop-2.7.3/sbin
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
hive> drop table Dataset;
OK
Time taken: 8.29 seconds
hive> CREATE EXTERNAL TABLE IF NOT EXISTS Dataset
  > (SNO INT,
  > MOVIE NAME STRING,
  > YEAR INT,
  > GENRE STRING
  > )
  > ROW FORMAT DELIMITED
  > FIELDS TERMINATED BY ','
  > STORED AS TEXTFILE
  > LOCATION '/dataset';
OK
Time taken: 0.403 seconds
hive> select * from Dataset;
OK
NULL    MOVIE NAME      NULL    GENRE
1       Aag Ka Dariya  1990
2       Aashiqui       1990    Romance
3       Awaargi 1990   Thriller
4       Baaghi: A Rebel for Love      1990    "Action
5       Dil           1990    "Comedy
```

Query for fetching the data:

```
Time taken: 0.403 seconds
hive> select * from Dataset;
OK
NULL    MOVIE NAME      NULL    GENRE
1       Aag Ka Dariya  1990
2       Aashiqui       1990    Romance
3       Awaargi 1990   Thriller
4       Baaghi: A Rebel for Love      1990    "Action
5       Dil           1990    "Comedy
```

Output Screenshot

Output of the queries

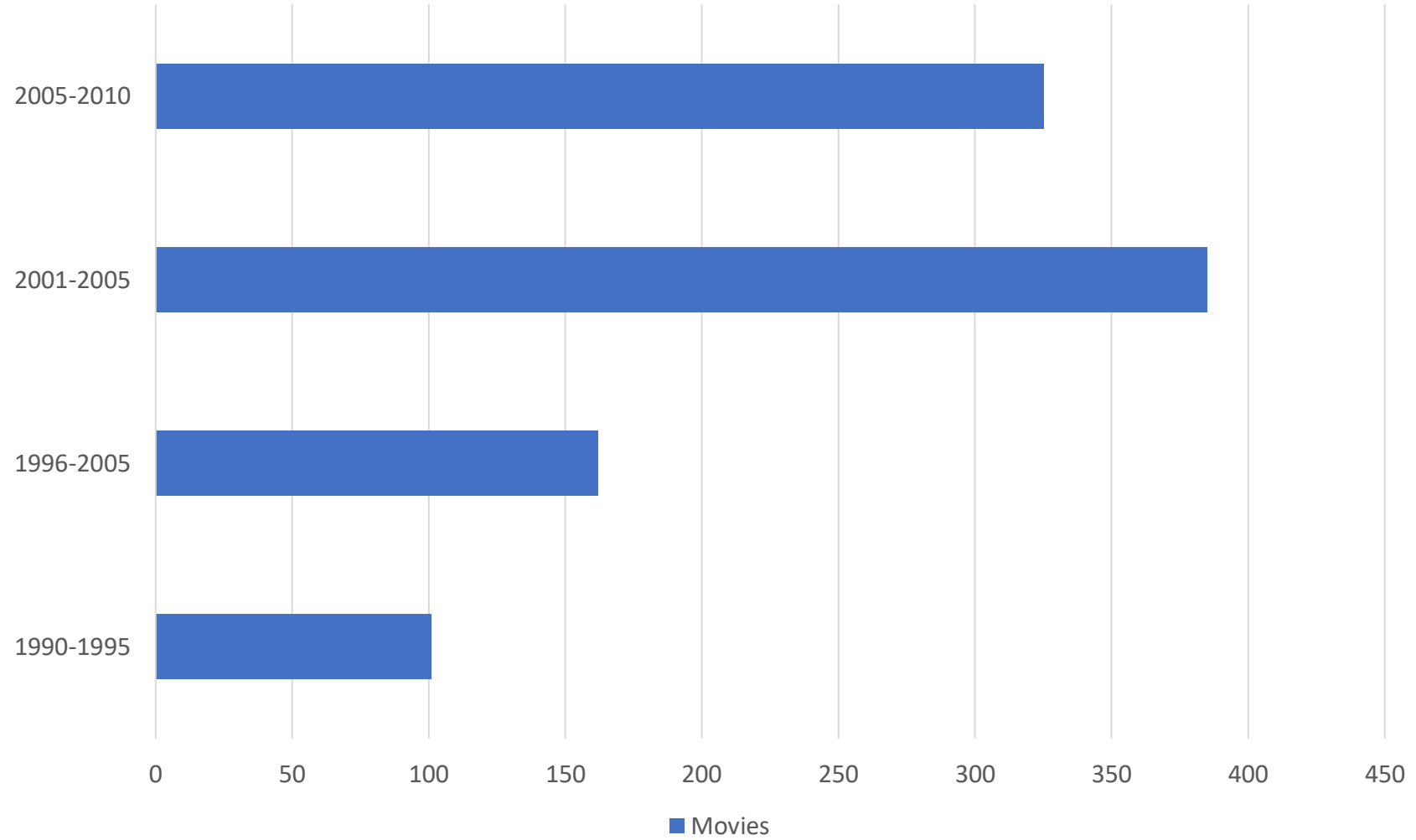
```
OK
Time taken: 0.403 seconds
hive> select * from Dataset;
OK
NULL      MOVIE NAME      NULL      GENRE
1          Aag Ka Dariya  1990
2          Aashiqui      1990      Romance
3          Awaargi 1990   Thriller
4          Baaghi: A Rebel for Love      1990      "Action
5          Dil          1990      "Comedy
```

Testing

- Check whether the machine working properly on oracle virtual box
- Check is there any issue in the implementation of virtual machine
- Check the entering valid credentitals like Password of local host should be correct
- Check the dataset we are using should contain more than 500 enteries (Minimum)

Conclusion

Movies



BIBLIOGRAPHY

I have done this project with the help of my supervisor Mrs. Vidushi Misra & alumni mentor & taking references from the following:

www.edurekha.com

www.javatpoint.com

I used:

- Oracle
- Virtual box
- CentOS-7
- Putty-64bit-0.7installer
- WinSCP
- Internet Explorer
- Chrome